



**UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE AGRONOMÍA  
COMISIÓN DE ESTUDIOS DE POSTGRADO  
POSTGRADO EN CIENCIA DEL SUELO**



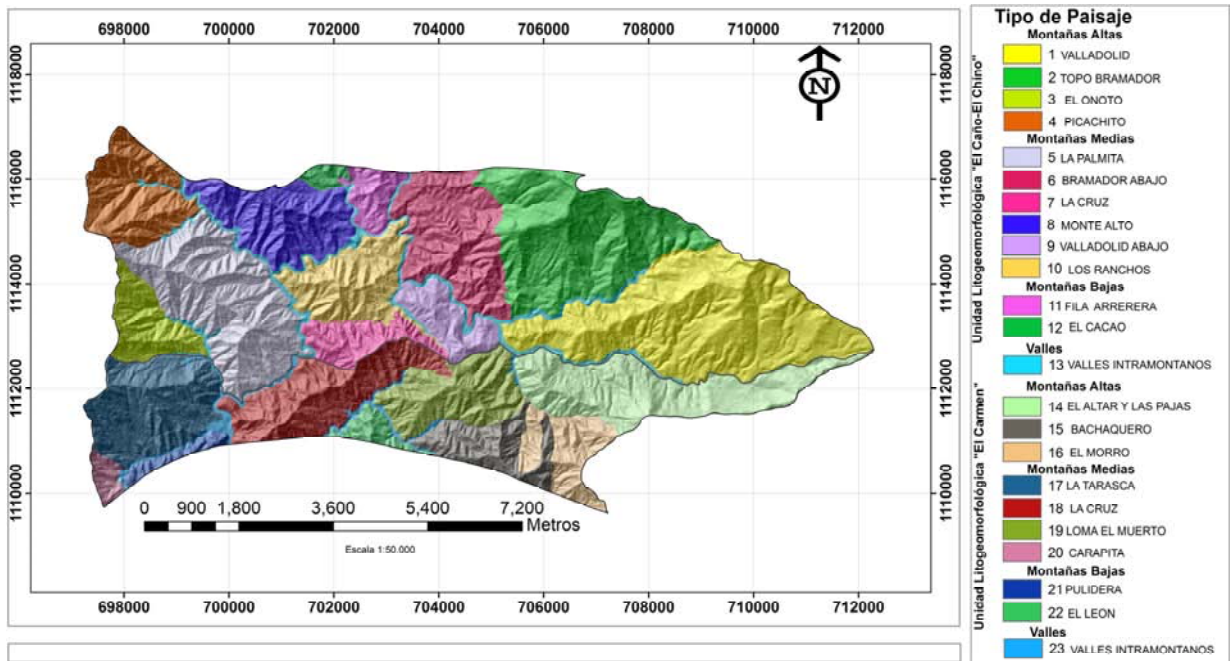
**INVENTARIO DE SUELOS Y PAISAJES CON APOYO DE TÉCNICAS DE  
CARTOGRAFÍA DIGITAL EN ÁREAS MONTAÑOSAS. CASO CUENCA DEL  
RÍO CARAMACATE, ESTADO ARAGUA.**

**Ángel Rafael Valera Valera**

**MARACAY, 2014**

## ABSTRACT

Sustainable development of the country urgently needs soil information at appropriate scales. However, conventional methods of soil survey are slow and expensive, and limit the production of this information with the requisite promptness, especially in dynamic, complex and difficult as mountain environments. The evolution of geospatial technologies such as geographic information systems (*GIS*) and remote sensing, the development of geomorphometric techniques, and the progress of the multivariate statistical methods and artificial intelligence technologies, such as artificial neural networks and fuzzy logic provide new opportunities to overcome this limitation and produce soil information more efficiently, with adequate precision and accuracy in the shortest time possible. In order to propose a digital mapping approach, an alternative to conventional soil survey for the production of reliable soil of semidetached level, in upper watersheds and landscapes, we selected a reference area in the Caramacate river basin (Guárico river tributary) on the mountains of North-Central from Aragua state, Venezuela. To obtain the landscape units underwent quantitative unsupervised classification based on fuzzy Kohonen clustering network (*FKCN*). MDE evaluations were conducted in different spatial resolutions (8, 10, 15 and 20 m) was selected the most representative model and obtained a total of 10 digital classes of land surface, meaning geomorphological was through the interpretation of the spatial distribution of descriptions of the class centers and the similarity values generated in the output layer of the network. For the study of soils sampled five nested hierarchical levels, to analyze the structure of spatial variation in soil attributes at various scales of variation: lithological units, landscape type, relief type or digital classes of land surface, a window 3x3 and in a cell (15 m) spatial resolution, which generated a total of 64 observations ( $2 \times 2 \times 2 \times 4 \times 2$ ). It was determined that the level that contributed most to the total variance of the soil is the relief type (mountain slopes), indicating that digital classes of land surface are considered an appropriate level for subsequent sampling, interpolation and generalization of attributes, with emphasis on digital soil mapping at scales between 1:20,000 and 1:100,000. Techniques were applied digital soil mapping considering digital model classes or digital classes of land surface, geomorphology and structure of soil spatial variability. To do this, four approaches to inference for mapping of soil properties based on sampling sites were used. The first is a model of soil-landscape spatial inference combined with theories of fuzzy sets (*FCM*), the second is based on a multiple linear regression (*MLR*), the third is an approach that combines regression analysis multiple linear with interpolation techniques (*RK*, Regression Kriging), and the fourth combining means neuro-fuzzy techniques *FKCN* with residuals interpolation (*RKF*). Assessing the reliability of prediction models allowed us to determine that the accuracy is related to the structure of variation in soil properties as it is high when the greatest variation occurs at the type of terrain and landscape, and is low when a short range high variation exists. The maps of each soil property were integrated using two spatial inference techniques: the first based on the algorithm of neuro-fuzzy clustering *FKCN* using eleven sets of artificial neurons as input layers to the system to obtain a final digital model nine classes for the *FCM-FKCN* approach and ten soil class for *RK-FKCN* and *RKF-FKCN* approaches. The second technique is based on the application of a soil-landscape model combined with theories of fuzzy sets for a group of pre-established taxonomic categories. Finally, the results obtained are synthesized as a proposal for digital mapping information generation soils and landscapes in mountainous areas.



**Figura 3-16.** Principales tipos de paisaje identificados en las unidades litogeomorfológicas de la zona de estudio.

### 3.3.7.3. Tipo de relieve

Para la categoría de tipos de relieve, se realizó un proceso de integración, reclasificación, vectorización y reorganización de las clases DST dentro de cada unidad litogeomorfológica y entre cada tipo de paisaje. La distribución geoespacial de los valores de las funciones de pertenencia y el concepto central de las clases permitieron la reorganización de las mismas en diez (10) unidades digitales de superficie de terreno distribuidas en cada unidad litogeomorfológica. De esta manera, se obtuvieron 20 geoformas o entidades espaciales representativas de los tipos de relieve presentes en el área de estudio.

Para la descripción de los tipos de relieve se utilizaron varios atributos de carácter local, tales como: orientación de la pendiente, NDVI, posición relativa y la configuración del terreno. A este nivel categórico, el sector estudiado fue dividido en 105 delineaciones o unidades de relieve (Figura 3-17) con variaciones graduales distribuidas en las 23 unidades de paisaje, dentro de las principales unidades litogeomorfológicas, cuya descripción detallada se presenta en los Cuadros 3-10 y 3-11.

En cada uno de los tipos de relieve discriminados por unidades litogeomorfológicas, es evidente la congruencia entre la orientación de la pendiente del terreno y la complejidad de la forma de terreno. La combinación entre curvaturas (vertical y horizontal) define con mayor claridad la alta convexidad de las crestas y la alta concavidad de las vegas de los valles intramontanos.

## TABLA DE CONTENIDO

	pág.
DEDICATORIA .....	V
AGRADECIMIENTO .....	VI
RESUMEN .....	VIII
ABSTRACT.....	X
LISTA DE CUADROS.....	XVII
LISTA DE FIGURAS.....	XX
<b>INTRODUCCIÓN</b> .....	1
<b>CAPÍTULO I</b> .....	5
EL PROBLEMA DE INVESTIGACIÓN.....	5
1.1. PLANTEAMIENTO DEL PROBLEMA.....	5
1.2. HIPÓTESIS.....	9
1.3. OBJETIVOS DE LA INVESTIGACIÓN .....	9
1.3.1. Objetivo general.....	9
1.3.2. Objetivos específicos .....	9
<b>CAPÍTULO II</b> .....	10
MARCO CONCEPTUAL .....	10
2.1. ESTUDIOS DE LAS RELACIONES SUELO-PAISAJE.....	10
2.1.1. Base conceptual de los modelos de relaciones suelo-paisaje .....	11
2.1.2. Estudios de relaciones suelo-paisaje en Venezuela .....	13
2.2. MODELOS DE VARIACIÓN ESPACIAL .....	14
2.3. CARTOGRAFÍA DIGITAL DEL PAISAJE .....	15
2.3.1. Geomorfometría.....	15
2.3.2. Geomorfometría, MDE y atributos superficiales del terreno .....	16
2.3.3. Métodos de evaluación de la cartografía digital del terreno.....	17
2.4. MODELO DIGITAL DE ELEVACIÓN Y RELACIONES SUELO-PAISAJE .....	18
2.4.1. Modelos digitales de elevación y variables topográficas .....	18
2.4.2. Modelos digitales de elevación y atributos del suelo .....	20
2.5. TÉCNICAS DE INTELIGENCIA ARTIFICIAL .....	21
2.5.1. Redes neuronales artificiales .....	21
2.5.1.1. Funcionamiento de una red neuronal artificial .....	22

2.5.1.2. Arquitectura o topología de las RNA .....	24
2.5.1.3. Los mapas auto-organizados de Kohonen .....	24
2.5.2. Teoría de los conjuntos borrosos o difusos.....	25
2.5.3. Sistemas neuro-borrosos: integración de RNA y LB.....	28
2.5.3.1. Red de agrupamiento borroso de Kohonen (FKCN) .....	29
2.6. APLICACIONES DE TÉCNICAS IA EN LA CIENCIA DEL SUELO .....	30
2.6.1. Las redes neuronales artificiales en la Ciencia del Suelo .....	30
2.6.2. La lógica borrosa en la Ciencia del Suelo.....	32
2.6.3. Los Sistemas neuro-borrosos en la Ciencia del Suelo .....	33
2.7. CARTOGRAFÍA DIGITAL DE SUELOS .....	33
2.7.1. Aplicaciones de las técnicas de cartografía digital de suelos .....	34
2.7.2. Cartografía digital de suelos y densidad de muestreo.....	36
2.7.3. Sistemas de inferencia espacial de propiedades del suelo .....	37
2.7.4. Modelos de predicción espacial.....	39
2.7.4.1. Modelos geoestadísticos .....	40
2.7.4.2. Correlación ambiental.....	40
2.7.4.3. Modelos híbridos (deterministas-estocásticos).....	42
2.7.4.3.1. Kriging de regresión .....	42
<b>CAPÍTULO III.....</b>	<b>45</b>
<b>CLASIFICACIÓN DIGITAL DEL TERRENO CON REDES NEUROBORROSAS EN LA</b>	
<b>CUENCA DEL RÍO CARAMACATE, ESTADO ARAGUA.....</b>	<b>45</b>
3.1. INTRODUCCIÓN .....	45
3.2. MATERIALES Y MÉTODOS.....	46
3.2.1. Área de estudio .....	46
3.2.1.1. Localización geográfica.....	46
3.2.1.2. Características físico-naturales .....	46
3.2.1.3. Área piloto .....	48
3.2.2. Procesamiento de la información cartográfica.....	49
3.2.3. Parámetros descriptivos de la superficie de terreno.....	50
3.2.4. Sistema clasificador del terreno: Red de agrupamiento borroso de Kohonen	
(FKCN) .....	54
3.2.5. Número óptimo de clases digitales de terreno .....	56
3.2.6. Selección del modelo representativo de las clases digitales.....	57

3.2.7. Significado geomorfológico de las clases digitales de superficie de terreno .....	58
3.2.8. Correlación entre clases digitales de superficie de terreno y el sistema jerárquico de clasificación de paisajes .....	58
3.2.9. Evaluación del modelo digital de superficie de terreno.....	58
3.3. RESULTADOS Y DISCUSIÓN .....	61
3.3.1. Características de los modelos digitales de elevación .....	61
3.3.2. Número óptimo de clases digitales de superficie de terreno .....	61
3.3.3. Selección del modelo representativo de las clases digitales .....	63
3.3.4. Influencia de la litología como variable discreta en la clasificación digital de la superficie de terreno.....	66
3.3.5. Evaluación del número de variables ambientales en la definición de unidades geomorfométricas .....	68
3.3.6. Significado geomorfológico de las clases digitales de superficie de terreno .....	72
3.3.7. Correlación entre las clases digitales de superficie de terreno y el sistema jerárquico de clasificación de paisajes.....	75
3.3.7.1. Unidades Litogeomorfológicas.....	76
3.3.7.2. Tipo de paisaje .....	77
3.3.7.3. Tipo de relieve .....	78
3.3.8. Evaluación del modelo de clases digitales de superficie de terreno.....	82
3.3.9. Tiempo de procesamiento en la clasificación digital del paisaje.....	85
3.4. CONCLUSIONES .....	88
<b>CAPÍTULO IV</b> .....	90
<b>ESTRUCTURA DE VARIACIÓN ESPACIAL DE PROPIEDADES DEL SUELO EN LA CUENCA DEL RÍO CARAMACATE</b> .....	90
4.1. INTRODUCCIÓN.....	90
4.2. MATERIALES Y MÉTODOS .....	92
4.2.1. Estructura del muestreo anidado .....	92
4.2.2. Muestreo de suelos.....	94
4.2.3. Análisis estadístico .....	97
4.2.4. Análisis de varianza anidado .....	97
4.3. RESULTADOS Y DISCUSIÓN.....	99
4.3.1. Análisis estadístico univariado .....	99
4.3.2. Distribución de los datos: Supuestos de normalidad .....	101
4.3.3. Análisis de varianza anidado .....	102

4.3.4. Estructura de variación y aporte a la variación total de los suelos en la cuenca del río Caramacate .....	105
4.3.4.1. Contribución de cada uno de los niveles del muestreo jerárquico a la variabilidad total del suelo .....	105
4.3.4.2. Patrón de variación espacial de los suelos en los diferentes niveles jerárquicos.....	106
4.4. CONCLUSIONES .....	109
<b>CAPÍTULO V</b> .....	110
<b>MODELOS DE PREDICCIÓN ESPACIAL DE PROPIEDADES DEL SUELO EN LA CUENCA DEL RÍO CARAMACATE</b> .....	110
5.1. INTRODUCCIÓN .....	110
5.2. MATERIALES Y MÉTODOS .....	111
5.2.1. Datos de suelos .....	111
5.2.2. Variables auxiliares.....	112
5.2.2.3. Factor Material Parental.....	114
5.2.2.2. Factor Relieve .....	114
5.2.2.3. Factor Clima.....	114
5.2.2.4. Factor Biota.....	115
5.2.2.5. Otros factores: Susceptibilidad a los movimientos en masa.....	115
5.2.3. Métodos de estimación de propiedades del suelo.....	115
5.2.3.1. Modelo de inferencia espacial bajo lógica borrosa.....	116
5.2.3.1.1. Modelo de similitud de suelos .....	116
5.2.3.1.2. Técnicas de inferencia para derivar los valores de similitud.....	117
5.2.3.1.3. Inferencia de propiedades del suelo basada en puntos de muestreo .....	117
5.2.3.1.4. Generación de mapas de propiedades de suelos con FCM.....	118
5.2.3.2. Modelo de regresión lineal múltiple .....	119
5.2.3.3. Modelo kriging de regresión.....	121
5.2.3.3.1. Generación de mapas de propiedades de suelos con técnicas RK.....	124
5.2.3.4. Modelo de redes neuro-borrosas y kriging de residuales .....	125
5.2.4. Evaluación de la bondad de ajuste de los modelos de predicción .....	126
5.2.5. Evaluación de la confiabilidad de los modelos de predicción .....	127
5.3. RESULTADOS Y DISCUSIÓN.....	130
5.3.1. Estadísticos descriptivos .....	130
5.3.2. Modelos de predicción de propiedades del suelo .....	131

5.3.2.1. Predicción de propiedades del suelo con técnicas borrosas.....	132
5.3.2.1.1. Evaluación de la confiabilidad de los modelos de predicción .....	133
5.3.2.1.2. Descripción de los modelos de inferencia borrosa .....	137
5.3.2.2. Predicción de propiedades del suelo con la aplicación de técnicas RLM .....	141
5.3.2.2.1. Análisis de regresión lineal múltiple .....	141
5.3.2.2.2. Modelos de regresión lineal múltiple.....	143
5.3.2.2.3. Evaluación de la confiabilidad de los modelos de predicción RLM .....	145
5.3.2.3. Predicción de propiedades del suelo con la aplicación de técnicas RK .....	146
5.3.2.3.1. Análisis de la tendencia espacial general.....	146
5.3.2.3.2. Análisis geoestadístico (tendencia local).....	146
5.3.2.3.3. Evaluación de la confiabilidad de los modelos de predicción RK .....	151
5.3.2.3.4. Descripción de los modelos RK.....	153
5.3.2.4. Predicción de propiedades del suelo con FKR .....	158
5.3.2.5. Descripción de modelos FKR .....	162
5.3.2.5.1. Evaluación de la confiabilidad de los modelos de predicción FKR .....	164
5.3.3. Comparación entre métodos de predicción y variables estimadas .....	168
5.4. CONCLUSIONES .....	174
<b>CAPÍTULO VI</b> .....	176
<b>CARTOGRAFÍA DIGITAL DE CLASES DE SUELO EN UN SECTOR DE LA CUENCA DEL RÍO CARAMACATE</b> .....	176
6.1. INTRODUCCIÓN .....	176
6.2. MATERIALES Y MÉTODOS .....	177
6.2.1. Propiedades del suelo y categorías taxonómicas .....	177
6.2.2. Métodos de predicción de clases de suelo .....	179
6.2.2.1. Métodos de integración de propiedades del suelo .....	179
6.2.2.1.1. Estructura de la red de agrupamiento de clases locales de suelo.....	179
6.2.2.1.2. Número óptimo de clases locales de suelo .....	180
6.2.2.2. Modelo de inferencia de clases basada en casos típicos.....	180
6.2.3. Evaluación de la confiabilidad de los mapas de clases de suelo .....	183
6.3. RESULTADOS Y DISCUSIÓN.....	184
6.3.1. Número de clases de suelo.....	184
6.3.2. Clases locales de suelos bajo el enfoque FCM-FKCN .....	184
6.3.3. Clases locales de suelos bajo el enfoque RK-FKCN .....	190



6.3.4. Clases locales de suelos bajo el enfoque FKR-FKCN.....	195
6.3.5. Clases locales de suelo bajo el enfoque de casos típicos.....	198
6.3.6. Evaluación de la confiabilidad de los mapas de suelo.....	203
6.4. CONCLUSIONES .....	206
<b>CAPÍTULO VII .....</b>	<b>208</b>
<b>PROPUESTA.....</b>	<b>208</b>
<b>PROPUESTA DE APLICACIÓN DE TÉCNICAS DE CARTOGRAFÍA DIGITAL AL INVENTARIO DE SUELOS Y PAISAJES EN ÁREAS MONTAÑOSAS.....</b>	<b>208</b>
7.1. INTRODUCCIÓN .....	208
7.2. PROPUESTA METODOLÓGICA.....	209
7.2.1. Desarrollo de una base de datos ambientales (suelo-paisaje).....	209
7.2.2. Identificación de patrones ambientales mediante clases digitales de superficie de terreno .....	210
7.2.2.1. Definición de unidades geomorfológicas .....	211
7.2.3. Muestreo de suelos.....	212
7.2.3.1. Densidad de observaciones .....	213
7.2.3.2. Geoformas y densidad de observaciones .....	214
7.2.3.3. Distribución de los sitios de muestreo en el terreno .....	215
7.2.3.4. Propiedades de suelo a determinar en cada sitio de muestreo .....	216
7.2.4. Productos de la Cartografía Digital de Suelos .....	216
<b>RESUMEN DE CONCLUSIONES .....</b>	<b>219</b>
<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>221</b>

## LISTA DE CUADROS

	pág.
<b>Cuadro 3-1.</b> Principales unidades litogemomorfológicas presentes en la cuenca del río Caramacate.....	49
<b>Cuadro 3-2.</b> Atributos topográficos utilizados en la clasificación digital del terreno de la cuenca del río Caramacate. ....	52
<b>Cuadro 3-3.</b> Características de los MDE generados para el área espacial atributiva en la zona de estudio.....	61
<b>Cuadro 3-4.</b> Parámetros estadísticos descriptivos del MDE de la cuenca del río Caramacate.....	62
<b>Cuadro 3-5.</b> Variación de la proporción (%) de las clases del modelo digital de superficie de terreno con la resolución del MDE, en el área de estudio.....	64
<b>Cuadro 3-6.</b> Variación de la proporción de las clases digitales con el número de variables ambientales en el área de estudio. ....	71
<b>Cuadro 3-7.</b> Valores del concepto central de cada clase digital de superficie de terreno identificada en la zona de estudio de la cuenca del río Caramacate.....	72
<b>Cuadro 3-8.</b> Unidades litogemomorfológicas dominantes en el sector de estudio de la cuenca del río Caramacate. ....	76
<b>Cuadro 3-9.</b> Unidades de paisaje presentes en el sector de estudio de la cuenca del río Caramacate.....	77
<b>Cuadro 3-10.</b> Tipos de relieve desarrollados sobre materiales geológicos de la Formación Metatobas de El Caño-El Chino en la cuenca del río Caramacate.....	80
<b>Cuadro 3-11.</b> Tipos de relieve desarrollados sobre materiales geológicos de la Formación Metalavas de El Carmen en la cuenca del río Caramacate.....	81
<b>Cuadro 3-12.</b> Proporción de límites coincidentes entre el modelo de unidades de relieve y las clases digitales de superficie de terreno. ....	84
<b>Cuadro 3-13.</b> Matriz de confusión obtenida de la clasificación digital FKCN y el mapa de tipos de relieve.....	84
<b>Cuadro 3-14.</b> Evaluación temporal en la producción de clases digitales de superficie de terreno. ....	86
<b>Cuadro 4-1.</b> Niveles empleados en el análisis de varianza anidado y sus principales parámetros.....	93
<b>Cuadro 4-2.</b> Resumen de la distribución de observaciones en el diseño de muestreo anidado en suelos de la cuenca del río Caramacate.....	95
<b>Cuadro 4-3.</b> Propiedades del suelo analizadas en cada sitio de muestreo.....	98
<b>Cuadro 4-4.</b> Estadísticos descriptivos de las propiedades morfológicas de los suelos de un sector de la cuenca del río Caramacate. ....	100
<b>Cuadro 4-5.</b> Estadísticos descriptivos de las propiedades físicas de los suelos de un	

sector de la cuenca del río Caramacate.....	100
<b>Cuadro 4-6.</b> Estadísticos descriptivos de las propiedades químicas de los suelos de un sector de la cuenca del río Caramacate.....	101
<b>Cuadro 4-7.</b> Distribución del porcentaje de varianza aportado por las variables morfológicas de los suelos de la cuenca del río Caramacate.....	102
<b>Cuadro 4-8.</b> Distribución del porcentaje de varianza aportado por las variables físicas de los suelos de la cuenca del río Caramacate.....	104
<b>Cuadro 4-9.</b> Distribución del porcentaje de varianza aportado por las variables químicas de los suelos de la cuenca del río Caramacate. ....	105
<b>Cuadro 5-1.</b> Datos digitales utilizados para representar las variables auxiliares <i>scorpan</i> en el área de estudio.....	113
<b>Cuadro 5-2.</b> Estadísticos descriptivos de las variables edáficas empleadas en la generación y validación de modelos de predicción. ....	130
<b>Cuadro 5-3.</b> Coeficientes de correlación lineal entre las propiedades edáficas y las variables auxiliares. ....	131
<b>Cuadro 5-4.</b> Indicadores de la bondad de ajuste de los modelos de predicción de propiedades del suelo con técnicas de inferencia borrosa. ....	132
<b>Cuadro 5-5.</b> Evaluación de la confiabilidad de los modelos de predicción de propiedades del suelo con técnicas de inferencia borrosa. ....	134
<b>Cuadro 5-6.</b> Coeficientes de concordancia de las propiedades del suelo agrupadas por unidades litogeomorfológicas y susceptibilidad a movimientos en masa. ....	136
<b>Cuadro 5-7.</b> Prueba de normalidad de los residuos obtenidos con la aplicación de las técnicas RLM.....	142
<b>Cuadro 5-8.</b> Modelos de predicción de propiedades del suelo basados en regresión lineal múltiple. ....	143
<b>Cuadro 5-9.</b> Indicadores de la bondad de ajuste de los modelos de predicción de propiedades del suelo con RLM. ....	144
<b>Cuadro 5-10.</b> Evaluación de la exactitud de los modelos de predicción de propiedades del suelo mediante RLM en el área estudiada. ....	145
<b>Cuadro 5-11.</b> Parámetros geoestadísticos obtenidos de los variogramas de los residuales de las variables edáficas en el área estudiada. ....	146
<b>Cuadro 5-12.</b> Indicadores de la bondad de ajuste de los modelos de predicción de propiedades del suelo con RK. ....	150
<b>Cuadro 5-13.</b> Evaluación de la exactitud de los modelos de predicción de propiedades del suelo mediante RK.....	151
<b>Cuadro 5-14.</b> Coeficientes de concordancia de las propiedades del suelo agrupadas por unidades litogeomorfológicas y susceptibilidad a movimientos en masa. ....	152
<b>Cuadro 5-15.</b> Valor promedio de las propiedades del suelo en cada clase digital de	

superficie de terreno de la unidad litogeomorfológica Metatobas de El Caño-El Chino. ....	158
<b>Cuadro 5-16.</b> Valor promedio de las propiedades del suelo en cada clase digital de superficie de terreno de la unidad litogeomorfológica Metalavas de El Carmen. ....	159
<b>Cuadro 5-17.</b> Prueba de normalidad de los residuos obtenidos con la aplicación de las técnicas FKR .....	160
<b>Cuadro 5-18.</b> Parámetros geoestadísticos obtenidos de los variogramas de los residuales de las propiedades del suelo en el área estudiada. ....	160
<b>Cuadro 5-19.</b> Indicadores de la bondad de ajuste de los modelos de predicción de propiedades del suelo con FKR en el área estudiada. ....	161
<b>Cuadro 5-20.</b> Evaluación de la exactitud de los modelos de predicción de propiedades del suelo con técnicas FKR en el área estudiada. ....	165
<b>Cuadro 5-21.</b> Coeficientes de concordancia de las propiedades del suelo agrupadas por unidades litogeomorfológicas y susceptibilidad a movimientos en masa. ....	166
<b>Cuadro 5-22.</b> Evaluación de la exactitud de los modelos de predicción de propiedades del suelo en el área estudiada. ....	171
<b>Cuadro 6-1.</b> Identificación de las categorías taxonómicas representativas de los suelos de la cuenca del río Caramacate. ....	181
<b>Cuadro 6-2.</b> Centroides de las clases locales de suelo borrosas representativas de la cuenca del río Caramacate, integradas con FKCN. ....	186
<b>Cuadro 6-3.</b> Correlación entre las clases de suelo borrosas con las familias taxonómicas y con el tipo de relieve en el área estudiada. ....	188
<b>Cuadro 6-4.</b> Centroides de las clases locales de suelo representativas de la cuenca del río Caramacate, integradas con FKCN. ....	191
<b>Cuadro 6-5.</b> Correlación entre las clases de suelo con las familias taxonómicas y con el tipo de relieve. ....	193
<b>Cuadro 6-6.</b> Centroides de las clases de suelo neuro-borrosas representativas de la cuenca del río Caramacate, integradas con FKCN. ....	195
<b>Cuadro 6-7.</b> Correlación entre las clases de suelos identificadas con las familias taxonómicas y con el tipo de relieve en el área estudiada. ....	198
<b>Cuadro 6-8.</b> Clasificación natural de los suelos en el sector de la cuenca del río Caramacate estudiado. ....	201
<b>Cuadro 6-9.</b> Discriminación de los suelos por unidad litogeomorfológica el sector de la cuenca del río Caramacate estudiado. ....	202
<b>Cuadro 6-10.</b> Resultados de la calibración de modelos de clases de suelo en un sector de la cuenca del río Caramacate. ....	203
<b>Cuadro 6-11.</b> Resultados de la validación de modelos de clases de suelo en un sector de la cuenca del río Caramacate. ....	203

**TESIS DE GRADO PRESENTADA COMO REQUISITO PARCIAL PARA  
OPTAR AL TÍTULO DE  
“DOCTOR EN CIENCIA DEL SUELO”**

**COMITÉ ASESOR**

---

**Dr. Jesús Viloría Rendón**  
(Tutor)

---

**Dra. Clara Rondón de Rodríguez**

## LISTA DE FIGURAS

	pág.
<b>Figura 2-1.</b> Esquema del funcionamiento general de una neurona artificial. ....	23
<b>Figura 2-2.</b> Límites difusos o de transición gradual y límites discretos. ....	27
<b>Figura 2-3.</b> Estructura de la red neuronal borrosa de Kohonen (FKCN). ....	30
<b>Figura 3-1.</b> Ubicación relativa del área de estudio dentro de la cuenca del río Caramacate, estado Aragua. ....	47
<b>Figura 3-2.</b> Unidades litogeomorfológicas presentes en el área de referencia espacial de la cuenca del río Caramacate. ....	49
<b>Figura 3-3.</b> Hojas cartográficas del área de referencia espacial-ARE (a), ortofotomapas (b), imagen satelital SPOT (c), mapas geológicos (d) y las coberturas de curvas de nivel (e) e hidrografía (f) del área de estudio. ....	51
<b>Figura 3-4.</b> Modelo digital de elevación del área de referencia espacial de la cuenca del río Caramacate. ....	51
<b>Figura 3-5.</b> Variables auxiliares derivadas del MDE (atributos e índices topográficos) y de la imagen satelital multiespectral en el área de la cuenca del río Caramacate estudiada. ....	53
<b>Figura 3-6.</b> Estructura de la RNA utilizada en el análisis de atributos morfométricos en el área de la cuenca del río Caramacate estudiada. ....	54
<b>Figura 3-7.</b> Diseño de la aplicación de transectas ( $t_i$ ) para la determinación de la proporción de límites de coincidencia entre modelos en el área de la cuenca del río Caramacate estudiada. ....	59
<b>Figura 3-8.</b> Variación del índice de rendimiento borroso (FPI) en función del número de clases para once variables ambientales: a) MDE 10 m, b) MDE 15 m; c) MDE 20 m y d) MDE 30 m. ....	62
<b>Figura 3-9.</b> Variación de las diferencias entre la proporción relativa de píxeles y la media de las clases digitales de superficie de terreno a distintas resoluciones espaciales en el área estudiada. ....	64
<b>Figura 3-10.</b> Modelos digitales de superficie de terreno a diferente resolución espacial: a) MDE 10 m, b) MDE 15 m, c) MDE 20 m, y d) MDE 30 m en la zona estudiada. ....	65
<b>Figura 3-11.</b> Modelo de clases digitales de superficie de terreno incluyendo la variable litología en el proceso de clasificación en el área estudiada. ....	67
<b>Figura 3-12.</b> Valores de los pesos de la capa de distancia de cada uno de los atributos en las clases DST generadas por FKCN en el área estudiada. ....	68
<b>Figura 3-13.</b> Modelos resultantes de las pruebas de sensibilidad al número de variables utilizadas como entrada al sistema FKCN en el área estudiada. ....	70
<b>Figura 3-14.</b> Modelos de la distribución de los valores de las funciones de	

pertenencia a las diferentes clases digitales de superficie de terreno. ....	73
<b>Figura 3-15.</b> Principales unidades litogeomorfológicas discriminadas en la zona de estudio. ....	76
<b>Figura 3-16.</b> Principales tipos de paisaje identificados en las unidades litogeomorfológicas de la zona de estudio. ....	78
<b>Figura 3-17.</b> Clases digitales de superficie de terreno correlacionada con los principales tipos de relieve identificados en la zona de estudio. ....	79
<b>Figura 3-18.</b> Representación de clases digitales de superficie de terreno y los tipos de relieve en el área estudiada. ....	83
<b>Figura 3-19.</b> Comparación del tiempo de procesamiento para once variables de entrada con diferentes resoluciones espaciales del MDE. ....	87
<b>Figura 4-1.</b> Diseño del muestreo balanceado con cinco niveles de jerarquía en la cuenca del río Caramacate. ....	94
<b>Figura 4-2.</b> Esquema del muestreo realizado en cada clase digital de superficie del terreno representativa del sector estudiado. ....	96
<b>Figura 4-3.</b> Variogramas representativos de los atributos en la cuenca del río Caramacate: (a) corto alcance en N5, (b) corto alcance en N4, (c) transicional N3-N2, (d) transicional N5-N4 y N3-N2, y (e) transicional N3-N1. ....	107
<b>Figura 5-1.</b> Distribución de los sitios de muestreo en el sector de estudio de la cuenca del río Caramacate. ....	112
<b>Figura 5-2.</b> Modelos digitales representativos de los factores formadores del suelo en la cuenca del río Caramacate. ....	113
<b>Figura 5-3.</b> Ejemplo esquemático del concepto del modelo RK en una muestra representativa. ....	123
<b>Figura 5-4.</b> Mapas de propiedades morfológicas y físicas obtenidos con las técnicas de inferencia bajo lógica borrosa en el área estudiada. ....	138
<b>Figura 5-5.</b> Mapas de propiedades químicas obtenidos con las técnicas de inferencia bajo lógica borrosa en el área estudiada. ....	139
<b>Figura 5-6.</b> Mapa de incertidumbre del pH de los suelos del sector evaluado en la cuenca del río Caramacate. ....	141
<b>Figura 5-7.</b> Gráfico de residuos en función de los valores estimados indicando el patrón típico de homogeneidad de varianzas. ....	142
<b>Figura 5-8.</b> Semivariogramas de los residuales de las propiedades de los suelos en el área estudiada. ....	147
<b>Figura 5-9.</b> Representación del procedimiento <i>kriging</i> de regresión en la estimación del pH de los suelos en el área estudiada. ....	149
<b>Figura 5-10.</b> Mapas de propiedades morfológicas y físicas estimadas con la aplicación de las técnicas de inferencia RK en el área estudiada. ....	155

<b>Figura 5-11.</b> Mapas de propiedades químicas del suelo estimadas con la aplicación de las técnicas RK en el área estudiada. ....	156
<b>Figura 5-12.</b> Mapas del error estándar de predicción del espesor del horizonte A (a), el contenido de carbono orgánico (b) y el pH de los suelos (c) en el área estudiada.....	157
<b>Figura 5-13.</b> Mapas de propiedades morfológicas y físicas estimadas con la aplicación de las técnicas de inferencia FKR en el área estudiada. ....	163
<b>Figura 5-14.</b> Mapas de propiedades químicas del suelo estimadas con la aplicación de técnicas de inferencia FKR en el área estudiada. ....	164
<b>Figura 5-15.</b> Variación del coeficiente de determinación ajustado de los modelos de predicción generados con diferentes técnicas de inferencia en el área estudiada.....	168
<b>Figura 5-16.</b> Variación del valor promedio de la suma de cuadrado de los errores de los modelos de predicción generados con diferentes técnicas de inferencia en el área estudiada. ....	169
<b>Figura 5-17.</b> Variación de la raíz del error cuadrático medio de los modelos de predicción generados con diferentes técnicas de inferencia en el área estudiada.....	170
<b>Figura 5-18.</b> Variación del error relativo de predicción de las propiedades del suelo estimadas con diversas técnicas de inferencia. ....	172
<b>Figura 6-1.</b> Estructura topológica de la red neuro-borrosa utilizada en la generación de las clases de suelo. ....	180
<b>Figura 6-2.</b> Clasificación de pixeles de suelos basado en los valores de similitud. ....	183
<b>Figura 6-3.</b> Variación del índice de rendimiento borroso (FPI) en función del número de clases para el agrupamiento de propiedades del suelo en el área estudiada. ....	185
<b>Figura 6-4.</b> Modelo digital de clases de suelo (FCM-FKCN) en un sector de la cuenca del río Caramacate. ....	189
<b>Figura 6-5.</b> Modelo digital de clases de suelo (RK-FKCN) en un sector de la cuenca del río Caramacate. ....	194
<b>Figura 6-6.</b> Modelo digital de clases de suelo neuro-borrosas (FKR-FKCN) en un sector de la cuenca del río Caramacate.....	197
<b>Figura 6-7.</b> Mapa de suelos a nivel de familia de clases de tamaño de partícula en la cuenca del río Caramacate. ....	199
<b>Figura 7-1.</b> Esquema del proceso de aplicación de cartografía digital de suelos y paisaje en áreas de montaña de la cuenca del río Caramacate.....	218



## INTRODUCCIÓN

En Venezuela existe deficiencia de información de suelos, especialmente a escalas semi-detalladas y detalladas. Según García (1995), los mapas de suelos a escala 1:50.000 ó mayor abarcan solamente 3% del territorio nacional. Además, la planificación del uso de la tierra en las áreas montañosas es con frecuencia limitada por la falta de información confiable sobre el recurso suelo. En estas áreas se han introducido usos no sostenibles de la tierra en lugares no aptos, lo que ha dado lugar a diversos grados de degradación de los recursos naturales.

La generación de la información faltante por medio de técnicas tradicionales de levantamiento de suelos es un proceso oneroso, ya que se requiere un gran número de observaciones de campo. En adición a esto, estas técnicas demandan un trabajo intensivo de fotointerpretación, realizado por un experto. Con el desarrollo de equipos de computación y las tecnologías de sistemas de información geográfica (SIG), teledetección y geoposicionamiento satelital (GPS), se ha incrementado la disponibilidad de datos ambientales. Esto ha permitido el avance de los enfoques cuantitativos que pueden complementar las técnicas de levantamiento tradicional cualitativo.

En la actualidad existe un enfoque alternativo denominado Cartografía Digital de Suelos (CDS o DSM por *Digital Soil Mapping*), el cual es empleado para la evaluación de las relaciones cuantitativas entre el recurso suelo y su ambiente. Este enfoque puede ser muy eficaz si se dispone de información auxiliar de los factores de formación del suelo, es capaz de representar y explicar la variación espacial de las propiedades relevantes del suelo.

La CDS hace un uso extenso de los avances tecnológicos a través de la geomática, como herramienta para la captura de datos espaciales de sensores remotos y métodos de geoposicionamiento, y los SIG para el procesamiento, modelización y análisis de datos. Se fundamenta en la geomorfometría, para el análisis de la superficie del terreno a través de modelos digitales de elevación (MDE) y en la edafometría, para el modelaje cuantitativo de las relaciones entre el suelo y el paisaje.

En esta investigación se propone la aplicación de las técnicas señaladas para un estudio a nivel semi-detallado, a fin de predecir características edáficas en un sector de la cuenca del río Caramacate (tributario del río Guárico). En tal sentido, se considera la definición de unidades de paisaje, la evaluación de la variabilidad de los suelos, la predicción de

propiedades edáficas, la evaluación de la exactitud y la incertidumbre de las predicciones, y la generación de clases de suelo como modelos de integración de las relaciones suelo-paisaje. Esta cuenca es considerada como área crítica y de atención prioritaria, ya que está sometida a intensos procesos de degradación ambiental y carece de suficiente información para apoyar la implantación de planes de manejo.

Generalmente, el proceso de levantamiento de suelos delimita unidades cartográficas apoyado en criterios geomorfológicos, subdividiendo así el paisaje en unidades relativamente uniformes con relación a la influencia de los factores formadores de los suelos. Este enfoque requiere del conocimiento de los principales elementos que actúan de manera simultánea en la complejidad de las relaciones suelo-paisaje, como la alta heterogeneidad edáfica, la dependencia espacial de las propiedades de los suelos y la complejidad de las relaciones entre los suelos y el paisaje

En este sentido, el enfoque utilizado en esta investigación se inició con la aplicación de técnicas neuro-borrosas para la definición de unidades morfométricas del terreno, a partir de parámetros derivados de modelos digitales de elevación (MDE) e imágenes de satélite multiespectrales. La lógica borrosa (LB) y las redes neuronales artificiales (RNA), simulan dos de las características más importantes del cerebro humano, como la capacidad de aprendizaje y la capacidad de procesar información incompleta, ambigua, compleja e incierta.

Las unidades geomorfométricas, interpretadas con criterios de geomorfología y génesis de suelos, condujeron a crear una imagen entendible del paisaje, y sirvieron como marco de referencia para el muestreo de suelos. Así, el muestreo permitió la cuantificación de la variabilidad del suelo a múltiples escalas, como base para el modelado y la predicción. Posteriormente, se aplicaron cuatro enfoques para la inferencia espacial de propiedades de los suelos, a partir de variables ambientales derivadas de MDE e imágenes satelitales, y finalmente se obtuvieron modelos de clases locales de suelo con el apoyo de técnicas de inteligencia artificial.

La tesis está estructurada en siete capítulos. Los cinco últimos se desarrollaron en forma de artículos extensos, cuya organización responde a los objetivos planteados. Estos capítulos conducen a un conjunto de conclusiones parciales, que finalmente son sintetizados como un cuerpo único de conclusiones.

En el *primer capítulo* se presenta el planteamiento del problema de investigación, las interrogantes que surgen del planteamiento, la justificación, la hipótesis y los objetivos de la investigación.

En el *segundo capítulo* se exponen los conceptos e ideas que sirven de marco a la línea de investigación basada en el estudio de las relaciones suelo-paisaje con técnicas de cartografía digital de suelos. Se integran aspectos de modelos de relaciones suelo-paisaje con métodos avanzados de procesamiento digital de datos espaciales y la aplicación de tecnologías de inteligencia artificial (lógica borrosa y redes neuronales artificiales) a la cartografía de suelos.

En el *tercer capítulo* se genera una clasificación digital del terreno dentro del área de estudio a partir de variables ambientales derivadas de un MDE y de imágenes satelitales, por medio de algoritmos basados en lógica borrosa y redes neuronales artificiales (FKCN, *Fuzzy Kohonen Clustering Network*). Se establece el significado geomorfológico de las clases digitales de terreno con base en su distribución espacial, las descripciones de los centros de clases y matrices de membresías y la información sobre geología, clima y vegetación del área de estudio.

En el *cuarto capítulo* se evalúa la estructura de variación espacial de las propiedades del suelo, con la finalidad de determinar el aporte de diferentes niveles de detalle (píxel, grupos de píxeles, clases morfométricas, unidades litogeomorfológicas) a la variación total del suelo en el área de estudio, permitiendo la definición de la escala de representación de dichas propiedades.

En el *quinto capítulo* se aplican sistemas de inferencia espacial en la cuenca del río Caramacate para la obtención de modelos de predicción de propiedades del suelo con cuatro enfoques: a) modelo de inferencia espacial suelo-paisaje combinado con teorías de conjuntos borrosos, b) análisis de regresión lineal múltiple, c) combinación del análisis de regresión lineal múltiple con técnicas de interpolación, y d) combinación de técnicas neuro-borrosas con interpolación de residuos. Además, se evalúa la confiabilidad de los modelos de inferencia espacial aplicados a la predicción de propiedades de los suelos, con un conjunto de datos independientes.

En el *sexto capítulo* se generan diversos modelos digitales de clases de suelo, a partir de la integración de modelos de variación espacial de propiedades del suelo con redes neuro-borrosas, y con técnicas de inferencia espacial suelo-paisaje apoyada en la teoría de los conjuntos borrosos.

El *séptimo capítulo* constituye una propuesta de la aplicación del enfoque de cartografía digital para la generación de información de suelos y paisajes en áreas montañosas. En la misma se conjugan varios componentes de la cartografía digital de suelos como base fundamental del enfoque de relaciones suelo-paisaje, para la producción de mapas de

propiedades y clases de suelo a nivel semi-detallado en un sector de la cuenca del río Caramacate. El enfoque aplicado y sus resultados suministran un marco conceptual y metodológico alternativo al levantamiento de suelos convencional, para la generación de información de suelos y paisajes en cuencas altas, que podría ser aplicado a otros sectores del país donde la información edáfica sea escasa o inexistente.

## CAPÍTULO I

### EL PROBLEMA DE INVESTIGACIÓN

#### 1.1. PLANTEAMIENTO DEL PROBLEMA

La información acerca de los recursos naturales de un país, región o localidad, es fundamental para la planificación y toma de decisiones con la mayor objetividad posible, en cuanto al uso de la tierra, el manejo de los suelos y la evaluación de problemas ambientales. Dentro de este entorno, el recurso edáfico es un componente cuya génesis, morfología, características y distribución geográfica se evalúa mediante los inventarios de suelos.

Sin embargo, en Venezuela el inventario de suelos se encuentra en un estado estacionario desde hace aproximadamente tres décadas, lo cual ha generado un vacío en la información de este recurso, especialmente a escalas semi-detalladas y detalladas. Hace aproximadamente treinta años que el Ministerio del Ambiente, organismo rector de la información edafológica en el país, no realiza levantamientos de suelos y sumado a ello, la formación de personal especializado en el área agrológica, por parte de las instituciones universitarias ha disminuido. Esto ha contribuido a restringir la producción de información básica agrológica, creando un vacío de información que afecta la toma de decisiones en la planificación del uso de la tierra. Actualmente, la única institución oficial que realiza estudios de suelos es el Instituto Nacional de Tierras (INTI) y únicamente se han evaluado algunos de los fundos afectados por la aplicación de la Ley de Tierras.

En tal sentido, se hace necesario retomar los estudios agrológicos, con la finalidad de cubrir los vacíos de información y responder a las exigencias actuales de la sociedad venezolana, considerando los progresos tecnológicos que conduzcan a la obtención de información confiable, precisa y rápida, acorde con los avances en esta materia a nivel mundial.

Para la evaluación del estado del inventario de suelos a nivel mundial es oportuno referir el análisis realizado por Zinck (1994) basado en las respuestas de representantes de 49 países de América Central, América del Sur, África y el Medio y Lejano Oriente (Asia). En este estudio, las conclusiones generales acerca del estado de la cobertura de los mapas de suelos fueron las siguientes: a) muchos países poseen clases de mapas generales a escalas menores a 1:250.000, b) la cobertura cartográfica para la planificación regional a escalas comprendidas entre 1:100.000 y 1:250.000 está extensamente incompleta, dificultando la identificación de áreas de alto potencial, o áreas con problemas críticos y sus prioridades para inventarios más detallados, c) los mapas de suelos apropiados para la planificación de proyectos a escalas alrededor de 1:25.000 cubre un porcentaje muy pequeño de los países, y

d) los mapas de suelos convenientes para la planificación de operaciones, generalmente a escalas mayores de 1:25.000, raras veces son mencionados.

Asimismo, Zhu *et al.* (2001) analizan la situación en los Estados Unidos de Norteamérica como resultado de las limitaciones de la manera convencional de conducir los levantamientos de suelos, ya que consumen demasiado tiempo. Estos autores destacan que esa nación posee alrededor de 900 millones de ha y la tasa de levantamiento de suelos es aproximadamente 4 millones ha/año. A esta tasa de producción, se requerirían 220 años para actualizar todo el inventario de suelos en ese país, y aún duplicando el esfuerzo entrenando más personal, llevaría más de un siglo completar esta tarea.

En Venezuela, según García y Salazar (1994), García (1995) y Comerma (2007) la superficie cubierta por estudios de suelos a gran visión (escala 1:250.000 ó 1: 500.000) incluye un poco más de 87,3 millones de hectáreas, equivalente al 95% del territorio nacional. Los mapas de suelos a escala 1:100.000 cubren 10% del país (aproximadamente 8.000.000 ha) y los mapas semi-detallados de suelos a escala 1:50.000 ó mayor (1:25.000) abarcan solamente un 3% del territorio nacional, equivalente a unas 2.800.000 ha. Añadiendo a esta información las evaluaciones ecosistémicas realizadas por PDVSA (2006), que incluyen estudios de suelos en una superficie de 2.171 km<sup>2</sup> a escala 1:100.000 en la Faja Petrolífera del Orinoco, se incrementó la superficie en la escala considerada, en menos del 0,3%.

De acuerdo con este panorama, existe una gran escasez de estudios de suelos tanto a nivel mundial como nacional, que resulta en una demanda de información edáfica a medianas y grandes escalas. Por ello es necesario intensificar las evaluaciones del recurso suelo, tanto en los sectores de mayor tradición agrícola en el país, como en aquellas zonas donde no existe ningún tipo de información, como las áreas montañosas.

Adicionalmente, los levantamientos convencionales de suelos se basan en la aplicación de métodos cualitativos y esquemas jerárquicos tradicionales, que involucran grandes inversiones de tiempo, dinero, mano de obra y logística (Viloria, 1991; Jácome, 1998; Rossiter, 2004). Consecuentemente, es necesario considerar otras alternativas que permitan predecir las características del suelo a partir de atributos del relieve y del paisaje.

Para superar las limitaciones relacionadas con la falta de información del recurso suelo en el país, existen técnicas que pueden contribuir a mejorar la eficiencia y la efectividad de los inventarios de suelos, tales como la edafometría, la geomorfometría y la geomática (Viloria, 2006). La edafometría se refiere a la aplicación de métodos estadísticos y matemáticos (lineales y no lineales) para el modelaje cuantitativo de los suelos, con el propósito de analizar su distribución, propiedades y comportamiento (McBratney, 1986;

Webster, 1994). La geomorfometría es la ciencia del análisis y descripción cuantitativa de la superficie del terreno a diversas escalas (Pike *et al.*, 2009). Por último, la geomática hace referencia a un conjunto de técnicas que integran los medios para la captura, tratamiento, análisis, interpretación, difusión y almacenamiento de información georeferenciada, con la finalidad de mejorar la toma de decisiones. Sus componentes más importantes son los sistemas de teledetección, los sistemas de información geográfica (SIG) y los sistemas satelitales de geoposicionamiento (GPS) (Jácome, 2002).

Este amplio abanico de opciones para apoyar la cartografía digital de suelos (CDS), producto del desarrollo de la geomática y de la aplicación de técnicas edafométricas y geomorfométricas, debe combinarse con información de geomorfología y génesis de suelos, a fin de contribuir a la creación de una imagen entendible del paisaje y su relación con el suelo (Viloria, 2006). En este sentido, durante las últimas décadas los avances tecnológicos han creado un gran potencial para desarrollar nuevos enfoques de cartografía de suelos, estimulados por los avances realizados en el campo de la tecnología de la información (Zinck, 2005).

Los métodos estadísticos multivariados y las técnicas geoestadísticas demuestran de manera cuantitativa, que el análisis del terreno puede emplearse para predecir los atributos del suelo en áreas relativamente pequeñas, con material parental de alto grado de homogeneidad. Además, la mayoría de las predicciones de las características del suelo derivadas de modelos digitales de elevación (MDE) se basan en enfoques estadísticos que describen las relaciones entre el suelo y los atributos topográficos del paisaje (Moore *et al.*, 1993; Gessler *et al.*, 1995; Thompson *et al.*, 2006; Willson y Gallant, 2000). Al respecto, son numerosos los autores que han considerado la conveniencia de derivar atributos del terreno de los MDE para la CDS (McSweeney *et al.*, 1994; Zhu, 1997; McKenzie y Ryan, 1999; McKenzie *et al.*, 2000; Scull *et al.*, 2003; Klingseisen, 2004; Dobos *et al.*, 2006; Burrough *et al.*, 2006).

Dentro de los modelos de predicción espacial destacan las redes neuronales artificiales (RNA) y la lógica difusa o borrosa (LB). Estas herramientas se basan en tecnologías de inteligencia artificial e intentan simular dos de las características más importantes del cerebro humano, como la capacidad de aprendizaje y la capacidad de procesar información incompleta o que no es precisa. Mediante la combinación de estas técnicas de inteligencia artificial con los métodos estadísticos multivariados es posible alcanzar un desarrollo tecnológico más amplio para la predicción de atributos del suelo (Minasny y McBratney, 2002) y la clasificación digital del paisaje (Bishop *et al.*, 2012).

Otro de los desarrollos metodológicos significativos para el estudio de las relaciones suelo-paisaje, son los métodos predictivos que combinan la regresión lineal múltiple de los datos

## ACTA 1



de suelo y la información ambiental, con la interpolación de los residuales (*kriging* de regresión) (Hengl *et al.*, 2007; Zhu y Lin., 2010; Sun *et al.*, 2012). Este método de análisis, aunado al desarrollo de los SIG, apoyados con información auxiliar de adecuada resolución espacial, ofrece nuevas oportunidades para superar las limitaciones de los levantamientos convencionales, para producir información edáfica de manera más eficiente, con adecuada precisión, exactitud y rapidez.

Las técnicas mencionadas aunadas al desarrollo de la plataforma tecnológica actual, permiten combinar la informática y los sensores remotos como herramientas eficientes en el inventario cartográfico, en la actualización de la información y en el continuo monitoreo de zonas de alta prioridad agrícola para el desarrollo sustentable. Sin embargo, es importante desarrollar propuestas de modelos de predicción robustos capaces de dar respuestas que satisfagan a múltiples usuarios, interesados en algunas propiedades y clases de suelo o en información básica para la alimentación de modelos interpretativos.

El planteamiento anterior deja abierta una serie de interrogantes cuyas respuestas se buscan con el desarrollo de esta investigación:

Los MDE y las imágenes satelitales son las fuentes de datos más comúnmente usadas en la CDS. Estos modelan el relieve y la superficie del terreno como una serie continua de celdas; pero no proveen información sobre el material parental del suelo. Esta última información podría ser obtenida de mapas geológicos; pero estos modelan las variaciones litológicas del área de interés como un conjunto de unidades discretas con límites abruptos. ¿Se puede utilizar la formación geológica como una variable de entrada o es más conveniente dividir el paisaje en diferentes unidades litológicas y generar una clasificación dentro de cada unidad por separado?

¿Cuál es el significado geomorfológico de las clases digitales de paisaje, generadas por medio de criterios morfométricos?

¿Cuál es la magnitud de la variabilidad espacial de propiedades del suelo dentro de un píxel, en una ventana de 2x2 ó 3x3 píxeles, entre píxeles de una misma clase digital de terreno y entre píxeles de clases diferentes? Si la variabilidad del suelo en cortas distancias es significativa, este procedimiento podría limitar seriamente la calidad de las predicciones.

¿Cuál es la manera más eficiente de distribuir los puntos de muestreo de suelo en el área de estudio, tomando en consideración la estructura de variación espacial del suelo y las dificultades de acceso en áreas montañosas?

¿Cuál es el grado de precisión y exactitud alcanzable en la predicción de atributos del suelo por medio del modelo digital de clases de paisaje generado para el área de estudio?

En síntesis ¿De qué manera el procesamiento de información digital podría contribuir al inventario de suelos como alternativa al levantamiento de suelos convencional, en cuencas altas de zonas tropicales?

## **1.2. HIPÓTESIS**

En cuencas altas de zonas tropicales con escasa información edáfica, la aplicación de técnicas de cartografía digital de suelos es una alternativa eficaz que permite lograr un alto grado de certeza en la predicción de propiedades edáficas y clases locales de suelo.

## **1.3. OBJETIVOS DE LA INVESTIGACIÓN**

### **1.3.1. Objetivo general**

Proponer un enfoque alternativo al levantamiento de suelos convencional para la generación de información de suelos y paisajes en cuencas altas, con apoyo de técnicas de cartografía digital.

### **1.3.2. Objetivos específicos**

- Generar un modelo digital de superficie de terreno en un sector de la cuenca del río Caramacate, a partir de parámetros derivados de un MDE y de imágenes satelitales, por medio de algoritmos basados en lógica borrosa y redes neuronales artificiales.
- Determinar el aporte de diferentes niveles de detalle (píxel, grupos de píxeles, clases digitales del terreno, tipos de paisaje, unidades litogeomorfológicas) a la variación total del suelo en el sector estudiado.
- Generar modelos de predicción de propiedades del suelo en el área estudiada, a partir de las variables ambientales derivadas del MDE y de imágenes satelitales.
- Producir modelos digitales de clases de suelo en el área estudiada, a partir de la integración de modelos de variación espacial de propiedades del suelo.
- Generalizar los resultados como una propuesta de cartografía digital para la generación de información de suelos y paisajes en áreas montañosas.

## CAPÍTULO II

### MARCO CONCEPTUAL

#### 2.1. ESTUDIOS DE LAS RELACIONES SUELO-PAISAJE

El levantamiento de suelos o más propiamente el inventario del recurso suelo, es una actividad sistemática que requiere de las bases científicas que involucran las relaciones entre el suelo y su entorno. Los estudios de las relaciones suelo-paisaje se fundamentan en modelos de morfogénesis y pedogénesis donde el suelo es parte del paisaje geográfico, y los atributos geológicos, geomorfológicos, hidrológicos y pedogenéticos se consideran como productos de procesos comunes de formación que funcionan de una manera integrada (e.g. Hudson, 1992; Zhu *et al.*, 2001; Olson, 2005; Thompson *et al.*, 2006). Los modelos considerados en los estudios de las relaciones suelo-paisaje incluyen:

- a) Modelos factoriales de suelos, entre los cuales destacan el modelo de los factores de estado (Jenny, 1941) y el modelo de energía para el desarrollo de los suelos (Runge, 1973).
- b) Modelos de procesos y sistemas dinámicos, que incluyen: b1) el modelo de procesos para la diferenciación de horizontes, basado en adiciones, remociones, translocaciones y transformaciones (Simonson, 1959); b2) el modelo de balance de masas (pérdidas y ganancias de sustancias en el suelo) sobre la base de respuestas geomorfológicas complejas (Brimhall y Dietrich, 1987), y b3) el modelo proceso-respuesta (Thornes y Brunson, 1977).
- c) Modelos de paisaje basados en la geomorfología, que involucran: c1) el concepto de catena (Milne, 1935), y c2) modelos de vertientes (*hillslope*) que se han desarrollado destacando las relaciones entre relieve, formas de terreno, hidrología y formación del suelo (Walker y Ruhe, 1968; Conacher y Dalrymple, 1977); c3) los modelos que consideran la naturaleza tridimensional de la relación suelo-paisaje (Hugget, 1975), y c4) los modelos geoespaciales multivariados (McBratney *et al.*, 2003).

La base para la definición de un modelo de relación suelo-paisaje es la ecuación de Jenny (1941), que identifica los cinco factores principales que contribuyen a la formación y desarrollo de los suelos: clima, organismos, relieve, material parental, y tiempo (*clorpt*). Recientemente, las adaptaciones de la ecuación formulada por McBratney *et al.* (2003) ha sido denominada modelo geoespacial multivariado *scorpan*, en la cual la estimación de un atributo o clase de suelo en una posición determinada es una función de otras propiedades de suelo medidas o estimadas en la misma localización (*s*), las condiciones climáticas (*c*), la acción de los organismos biológicos (*o*), los parámetros del relieve (*r*), el material

parental (*p*), el tiempo (*a*) y el espacio geográfico o la posición geográfica donde se estima el atributo del suelo (*n*). Este modelo incluye la naturaleza tridimensional de las relaciones suelo-paisaje como base de la cartografía digital de suelos y paisajes.

En general, el estudio de factores y procesos morfogénicos y pedogénicos constituye un sistema de bases teóricas que resume muchos de los principios básicos de la Ciencia del Suelo moderna. Aparte de considerar las relaciones empíricas del enfoque factorial, también se toma en cuenta la interpretación geomorfológica y fisiográfica, y las relaciones geoespaciales entre las propiedades y los atributos del paisaje. Estos aspectos conceptuales enmarcan la base de la presente investigación.

### **2.1.1. Base conceptual de los modelos de relaciones suelo-paisaje**

El enfoque suelo-paisaje, permite estudiar al suelo como parte del paisaje geográfico, basado en el desarrollo de modelos que relacionan las formas de terreno con las propiedades del suelo o sus clases (McLeod *et al.*, 1995; Mejías y Vera, 2001).

El enfoque surge principalmente de observaciones empíricas de la relación entre la distribución de los suelos y su posición en el paisaje. Esta relación es diferente entre unas áreas geográficas y otras. Establecer asociaciones suelo-paisaje requiere conocimientos de geomorfología y edafología. La comprensión de las relaciones suelo-paisaje en una determinada zona sirve de base para establecer los modelos de distribución de suelos, lo que facilita los trabajos en cartografía de suelos y ordenamiento territorial (Porta *et al.*, 1999).

Asimismo, el suelo como parte del paisaje es el producto de una interacción compleja de procesos físicos, químicos y biológicos, regidos por interacciones climáticas y geomorfológicas que ocurren con el tiempo. Debido a que el clima, la geología y el relieve varían de una zona a otra, las interacciones cambian y por lo tanto, cualquier modelo de predicción también debe cambiar. En estos aspectos, la geomorfología juega un papel importante en los procesos que determinan la distribución de los suelos en el paisaje y los procesos pedológicos pueden considerarse como parte integral de la evolución del paisaje (Hall, 1983). En este sentido, Walker (1989) destaca que sólo a través de los estudios de relaciones suelo-paisaje es posible entender la distribución de los suelos, su génesis y su variabilidad espacial.

El concepto suelo-paisaje afirma que si se conocen las relaciones entre cada suelo y su ambiente para una zona específica, entonces se puede deducir qué tipo de suelo puede estar en cada localidad en el paisaje, a partir de la identificación de las condiciones ambientales de ese sitio (Zhu *et al.*, 2001). Ello se debe principalmente a que los suelos evolucionan

conjuntamente con su medio ambiente y representan una parte significativa y funcional del paisaje (Dobos y Hengl, 2009).

El modelo suelo-paisaje permite a los edafólogos predecir los tipos de suelos y sus propiedades asociadas, a partir de la relación entre los suelos y las propiedades del paisaje y un conjunto limitado de observaciones del suelo. El modelo es construido sobre el concepto de unidades suelo-paisaje, las cuales son unidades de terreno naturales con formas y configuraciones observables, resultantes de las interacciones de los cinco factores que afectan la formación de los suelos: material parental, clima, organismos, relieve y tiempo (Hudson, 1992; De Bruin y Stein, 1998). De esta manera, el establecimiento de un modelo suelo-paisaje aparte de permitir el estudio del recurso suelo a través de un enfoque integral y una visión sistémica, sirve de base fundamental para el desarrollo de estrategias más eficientes en los procesos de planificación de los recursos naturales (McLeod *et al.*, 1995; Rijkse y Trangmar, 1995; Murphy *et al.* 2005; McKenzie *et al.*, 2003).

Investigadores como Schood y Veldkamp (2005), enfatizan el papel central del paisaje a diferentes escalas (multiescalar), y destacan que el paisaje es el principal factor conductor y responsable de muchos procesos a diferentes niveles (temporales y espaciales) en las ciencias geoambientales, y puede ser considerado como la consecuencia de la evolución geológica y los procesos geomorfológicos. El paisaje puede ser definido en términos de su génesis (formación, procesos), fisiografía (forma, configuración actual), litología y suelos (su composición), cobertura de la tierra (características superficiales), uso de la tierra (su uso, funciones humanas), y el manejo de la tierra (factor humano). Estos autores señalan que todas esas interrelaciones deben ser consideradas en el contexto del modelo suelo-paisaje multiescalar.

El modelo suelo-paisaje proporciona un marco geográfico al enfoque cuantitativo para analizar y predecir la distribución espacial de las propiedades del suelo basadas en la variabilidad de correlaciones ambientales, particularmente de determinados parámetros topográficos e hidrológicos (Moore *et al.*, 1991; McSweeney *et al.*, 1994). Los métodos implican normalmente: a) la caracterización del dominio fisiográfico local con el análisis de datos de MDE; b) la colección de muestras de suelos georeferenciadas y la compilación de datos de las propiedades del suelo deseadas y c) el desarrollo de modelos empíricos explícitos, cuantitativos, y generalmente sencillos. El enfoque proporciona los medios para cuantificar la distribución espacial de las propiedades del suelo con base en la variabilidad de atributos topográficos fuertemente correlacionados, que son más fáciles de recolectar con una mayor resolución que con el muestreo y la medición directa de propiedades del suelo (Thompson *et al.*, 2006).

De acuerdo con las referencias citadas, es importante señalar que la mayoría de las

investigaciones relacionadas con el modelo suelo-paisaje se han desarrollado a partir del modelo factorial de formación del suelo. Esto implica una estrecha relación entre la distribución de los suelos y las diferentes posiciones en el paisaje. De esta manera, en áreas con condiciones climáticas y geológicas relativamente uniformes, el relieve es el principal factor que controla la variabilidad espacial del suelo.

### **2.1.2. Estudios de relaciones suelo-paisaje en Venezuela**

En Venezuela se han propuesto varios sistemas de clasificación sistemática del paisaje, como apoyo al inventario de suelos (*e.g.* Freile, 1962; Steegmayer y Bustos, 1980; Zinck, 1981; Elizalde, 1983). Estos sistemas de clasificación han sido utilizados como medios para interpretar las relaciones suelo-paisaje. Adicionalmente, Elizalde y Jaimes (1989) proponen un enfoque sistémico del paisaje, en el cual los procesos desarrollados son el resultado del intercambio de flujos de materia y energía entre los componentes del paisaje (sistema pedogeomorfológico) y el ambiente circundante.

El modelo de relaciones suelo-paisaje propuesto por Elizalde y Jaimes (1989) fue desarrollado con la finalidad de analizar y comprender la génesis de los componentes del sistema pedogeomorfológico, partiendo de la concepción del paisaje como un sistema. Esto ha dado lugar a investigaciones y aplicaciones en el análisis de las relaciones suelo-paisaje y al estudio de los procesos pedogenéticos en el país, lo cual ha contribuido al avance de la Ciencia del Suelo a nivel nacional.

La concepción del modelo pedogeomorfológico ha sido aplicada en diversas investigaciones realizadas en el país, incluyendo la evaluación de la susceptibilidad a la erosión hídrica en la cuenca del río Tucutunemo (Fernández, 2001), la evaluación de tierras en la cuenca alta del río Guárico, mediante el uso de indicadores de calidad de suelo (Ospina, 2003), y en el estudio de suelos sulfato-ácidos de un área del Delta Superior del río Orinoco (Maza, 2010; Maza *et al.*, 2011; Rondón *et al.*, 2012).

Más recientemente, Elizalde (2012) ha propuesto una segunda aproximación de clasificación sistemática de paisajes vinculado con los conceptos de ecosistema y sistema pedogeomorfológico. El sistema jerárquico propuesto crea un marco para nuevos mapas geomorfológicos más detallados, constituye una referencia para la correlación regional o internacional de paisajes, y para la elaboración de mapas de unidades ecológicas, de suelos, de vegetación, de unidades agroecológicas o similares. Además, permite identificar la distribución espacial de los indicadores de riesgo ambiental de diferentes tipos de uso de la tierra y es adecuado para la selección de áreas de referencia de paisajes para la identificación de indicadores orientados a políticas agroambientales.

Zinck (2012) combinó varios enfoques de clasificación geomorfológica con la estructura de

un sistema taxonómico de geoformas desarrollado a partir de levantamientos geopedológicos en Venezuela y posteriormente utilizado en los Países Bajos (Holanda). El autor integró los criterios básicos para la construcción de un sistema taxonómico de geoformas con los resultados del ensayo exploratorio para detectar lineamientos de estructura jerárquica en el entorno geomorfológico, y obtuvo un sistema con seis niveles categóricos anidados. El sistema incluye información sobre el material parental de los suelos de las geoformas para sustanciar las relaciones entre geoforma y suelo, con la finalidad de documentar la litología en caso de substrato rocoso o las facies en caso de materiales no consolidados. Su inserción en la parte inferior del sistema se justifica por el hecho de que se necesita a menudo información de campo para complementar o precisar la información general suministrada por los mapas geológicos.

## **2.2. MODELOS DE VARIACIÓN ESPACIAL**

Los métodos utilizados para predecir cómo es el suelo, se fundamentan en los siguientes modelos o hipótesis de variación espacial: a) los modelos discretos que consideran al suelo como una colección de cuerpos naturales; b) los modelos continuos de variación espacial del suelo, y c) la combinación e integración de los enfoques anteriores, que ha generado un modelo mixto. Este último es un enfoque híbrido, con características potenciales de los dos modelos principales, el cual considera al suelo como un continuo con unidades discretas (Ibáñez *et al.*, 2005; Rossiter, 2005; Viloría, 2006).

A pesar de las fuertes discusiones para adoptar un modelo continuo para representar la variación de características del suelo y formas de terreno (Zhu 1997), la mayoría de los mapas disponibles adoptan un modelo coroplético (discreto). Incluso, algunos autores relacionados con el área de clasificaciones cuantitativas (Burrough *et al.*, 1997) han reconocido la necesidad de realizar por lo menos una estratificación inicial de la relación suelo-paisaje, empleando límites primarios y divisiones zonales basadas en diferencias de litología, formas de terreno y drenaje.

Los argumentos mencionados, indican que los métodos cartográficos disponibles, se han fundamentado bien sea en el modelo discreto o en el modelo continuo de variación espacial del suelo. Sin embargo, queda abierta la posibilidad de considerar la integración de ambos tipos de modelos, de manera que el suelo pueda ser concebido como un continuo conformado por un conjunto de unidades discretas o clases separadas entre sí por límites claros y precisos (Heuvelink y Webster, 2001; Hennemann y Rossiter, 2004; Rossiter, 2005; Ibáñez *et al.*, 2005; Morales y Viloría, 2006). Ambos enfoques se pueden combinar en un modelo general de variación espacial o un modelo mixto, donde cada uno incorpore sus supuestos específicos, con la finalidad de representar lo más objetivamente posible al manto del suelo (Heuvelink y Webster, 2001). Es decir, aunque el suelo existe como un

manto continuo, éste puede ser dividido en un número de elementos discretos o individuales para los propósitos de muestreo (Webster y Oliver, 1990).

La integración de ambos modelos es muy conveniente porque en el suelo la distribución espacial fluctúa entre un modelo discontinuo y continuo, debido a que las discontinuidades fisiográficas y la variación subsuperficial gradual a través de límites y dentro de unidades de suelo no están bien definidas (Lagacherie *et al.*, 1997). Este enfoque utiliza las características y potencialidades de los modelos de variación discretos y continuos, para estudiar la complejidad de la geografía del suelo y la naturaleza oculta, incierta, ruidosa y borrosa del mismo como cuerpo tridimensional complejo y dinámico, que varía tanto en el espacio como en el tiempo.

El modelo de unidades discretas requiere realizar una generalización temprana de gran volumen de datos primarios producidos en el levantamiento de suelos. Esto trae como consecuencia una mayor pérdida de información y una menor posibilidad de identificación de la variación sistemática. En cuanto al enfoque continuo, actualmente las herramientas computarizadas permiten manejar grandes volúmenes de datos (descripciones de perfiles de suelos, descripciones del paisaje, análisis de laboratorio, relaciones interpretativas), por lo que no es necesario generalizar desde el inicio del procesamiento de los datos y de esta manera, no existe pérdida de datos primarios.

La combinación de modelos discretos y continuos, permite el manejo cuantitativo de grandes volúmenes de datos y la interpretación de estructuras regionales (cambios en el ambiente de formación de suelo). Además, la generalización de los datos post-procesamiento permite presentar una información más comprensible.

### **2.3. CARTOGRAFÍA DIGITAL DEL PAISAJE**

La clasificación automática de unidades geomorfológicas del terreno se centra principalmente en parámetros morfométricos que pueden describir la forma superficial del terreno (Jamieson *et al.*, 2004; Giles y Franklin, 1998; Miliarensis, 2001; Bue y Stepinski, 2006; Evans, 2012). Este proceso de clasificación también se conoce como geomorfometría (Hengl y Evans, 2009), geomorfología cuantitativa (Pike *et al.*, 2009), o cartografía digital geomorfológica (Bishop *et al.*, 2012).

#### **2.3.1. Geomorfometría**

La geomorfometría consiste en la aplicación de métodos matemáticos y estadísticos al estudio de las formas del relieve terrestre. Se basa en la aplicación de métodos cuantitativos al análisis del paisaje para expresarlo en forma paramétrica, (Hengl y Evans, 2009; Pike *et al.*, 2009). La geomorfometría se refiere a la descripción cuantitativa de las formas de los elementos del terreno, y junto con la litología y los procesos genéticos delimitan una unidad



geomorfológica, equivalente a una zona con formas, materiales y génesis homogéneas (Dikau *et al.*, 1995; Gustavsson *et al.*, 2006).

De acuerdo con Rasemann, *et al.* (2004), esta ciencia se encarga de la descripción y el análisis cuantitativo de las características topológicas y geométricas del paisaje, para el reconocimiento y la cuantificación de las formas de terreno mediante la utilización de las matemáticas, la geociencia y la computación.

La geomorfometría involucra los temas relacionados con: a) el muestreo de los atributos de la superficie del terreno; b) la generación de modelos digitales de elevación; c) el procesamiento y la evaluación del error de MDE; d) la generación de parámetros, índices, y objetos de la superficie del terreno, y e) la producción de información geomorfométrica y la resolución de problemas utilizando los parámetros de la superficie del terreno. En general, es la base de la cartografía digital geomorfológica (Bishop *et al.*, 2012).

### **2.3.2. Geomorfometría, MDE y atributos superficiales del terreno**

Los aspectos conceptuales descritos anteriormente son la plataforma del análisis morfométrico del MDE, el cual permite evaluar las formas de terreno, la definición de unidades geomorfométricas, la descripción y caracterización de la variabilidad espacial de procesos específicos que ocurren en el paisaje, y la predicción de atributos y clases de suelo (Moore *et al.*, 1991; McKenzie *et al.*, 2000; Bishop y Minasny, 2005). Estos atributos topográficos aportados por las estructuras numéricas, constituyen un amplio conjunto de variables cuantitativas continuas que se denominan modelos digitales multivariados del terreno y constituyen la base fundamental para la obtención de clases geomorfométricas y la definición de unidades geomorfológicas.

La información acerca del relieve o la forma superficial del terreno es uno de los requerimientos fundamentales para el estudio del modelado en las ciencias ambientales y en la geomorfología aplicada. El análisis del terreno es el conjunto de actividades que conducen a la recopilación de características o cualidades del terreno (Townshend, 1981). La cuantificación y el análisis de los atributos del terreno obtenidos a partir de un MDE, también se conoce como la parametrización de la superficie del terreno o la descripción numérica de la superficie continua del paisaje (Shary *et al.*, 2002).

El análisis automatizado del terreno se basa en un MDE, el cual cada día es más utilizado en las investigaciones geomorfológicas, debido a que éste proporciona bases topográficas detalladas, y por consiguiente información geomorfológica potencial acerca de terrenos de difícil accesibilidad. El MDE es la base para el cálculo de los parámetros superficiales del terreno, entre los cuales se incluyen los atributos topográficos primarios y secundarios.

Los atributos topográficos primarios son aquellos que se calculan directamente de los datos de elevación e incluyen fundamentalmente variables como altitud, pendiente, orientación, tipos de curvatura y el área de captación. Los atributos secundarios o compuestos implican combinaciones de dos o más atributos primarios, y son índices que describen o caracterizan la variabilidad espacial de las condiciones topográficas en las cuales ocurren procesos específicos en el paisaje, tales como la distribución del agua en el suelo, el potencial para la erosión laminar, etc. (Moore *et al.*, 1991). Estos contribuyen a la descripción de los patrones topográficos como una función del potencial, para la ocurrencia de los procesos de redistribución del agua en el paisaje y en la modificación de la cantidad de radiación solar recibida en la superficie, con evidentes consecuencias hidrológicas, geomorfológicas y ecológicas.

### **2.3.3. Métodos de evaluación de la cartografía digital del terreno**

Según Arrell *et al.* (2007), los dos principales enfoques para el reconocimiento de formas de terreno son la clasificación deductiva (utilizando clases predeterminadas) y la inductiva (generando clases a partir de los datos).

El enfoque deductivo examina el MDE como una plantilla del arreglo de alturas relativas dentro de un área fija (e.g. ventana 3x3 celdas) y asigna el punto central de esa área a un limitado conjunto de clases morfométricas preestablecidas. El número de clases varía dependiendo de la implementación, de un mínimo de 6 clases reconocidas por Evans (1980), Peucker y Douglas (1975) y Wood (1996) a las 11 clases incluidas en el trabajo de Pellegrini (1995). Este enfoque determinístico permite sin duda la asignación de un píxel a una clase, pero Wood (2002) ha demostrado que muchos parámetros morfométricos están sujetos a variaciones con la resolución del MDE sobre el que es ejecutado el análisis. Por lo tanto, la clasificación es consistente con la concepción de la superficie del terreno como una composición dependiente de la escala. Los elementos independientes de la escala donde las clases morfométricas específicas son identificables como objetos discretos, son poco comunes (Fisher y Wood, 1998). Fisher *et al.* (2004) exploraron esta observación utilizando la incertidumbre como un método para generalizar los grados de pertenencia borrosa a las clases.

El método inductivo ha sido empleado en clasificaciones automatizadas no supervisadas de formas de terreno, utilizando como entrada una variedad de derivadas geomorfométricas. Entre los autores que han implementado este enfoque destacan: Irvin *et al.* (1997), de Bruin y Stein (1998), Lagacherie *et al.* (1997), Burrough *et al.* (2000), Macmillan *et al.* (2000) y Arrell *et al.* (2007).

Diversos investigadores han realizado importantes trabajos relacionados con la

## ACTA 2

determinación automatizada o semi-automatizada de formas de terreno, cuyos aspectos fundamentales se señalan a continuación: Bolongaro-Crevenna *et al.* (2005) demostraron que existe una relación estricta entre los atributos morfométricos y los procesos geomorfológicos. Burnett y Blaschke (2003) proporcionan una metodología para descomponer, modelar y clasificar entidades espaciales basadas en la segmentación multi-escalar y el modelado de relaciones de objetos. Adediran *et al.* (2004) emplearon un método basado en la aplicación de estadística multivariada y algoritmos de procesamiento de imágenes a datos de relieve, para la discriminación de unidades morfológicas. Luoto y Hjort (2006), compararon la precisión de diferentes enfoques en el modelado predictivo geomorfológico de paisajes subárticos. Drăguț y Blaschke (2006) presentaron un sistema de clasificación automatizado de elementos de formas de terreno basado en análisis de imágenes orientado a objeto. Van Asselen y Seijmonsbergen (2006) presentaron un método experto semi-automatizado para definir las unidades geomorfológicas de un MDE de láser de alta resolución. Asimismo, Iwahashi y Pike (2007) desarrollaron un sistema de clasificación no supervisada del relieve mediante un procedimiento iterativo que considera tres firmas geométricas: pendiente, textura superficial y convexidad local.

La mayoría de las investigaciones geomorfológicas mencionadas han demostrado que los datos derivados de MDE son útiles en la determinación de cualquier forma de superficie de terreno. Asimismo, los resultados señalan que el modelado predictivo geomorfológico tiene la ventaja de proporcionar información relevante y útil acerca de los procesos que ocurren en la superficie, algo difícil de lograr utilizando métodos de levantamiento convencional.

## **2.4. MODELO DIGITAL DE ELEVACIÓN Y RELACIONES SUELO-PAISAJE**

### **2.4.1. Modelos digitales de elevación y variables topográficas**

El relieve desempeña el papel de factor de redistribución de energía y sustancias, lo cual hace que las unidades morfológicas del relieve tengan una notable influencia sobre numerosas variables que intervienen en la dinámica de los ecosistemas (Priego *et al.*, 2004). Por esto el relieve es frecuentemente usado como base para la delimitación física de las unidades de paisaje. Y tal como se señaló en las secciones anteriores, el suelo es parte del paisaje geográfico, existiendo una relación muy estrecha entre las propiedades de los suelos y la posición de éstos en el paisaje. Así, por ejemplo, las propiedades de un suelo reflejan su grado de evolución, lo cual sólo puede ser entendido si se considera, entre otros aspectos, la posición de ese suelo dentro de la unidad de paisaje a la cual pertenece.

En la mayoría de los paisajes, los procesos de desarrollo del suelo son afectados por el flujo del agua superficial y subsuperficial. La comprensión del vínculo entre los patrones del suelo y el relieve permite la predicción de atributos del suelo, según la posición de éste en

el paisaje (Moore *et al.*, 1993). Las evaluaciones cuantitativas de propiedades del suelo junto con los MDE, pueden proporcionar un enfoque conveniente para la predicción espacial de muchas características del suelo, y pueden dar ideas acerca de los atributos del terreno que influyen el movimiento de sedimentos en las laderas. Ello permitiría el establecimiento de modelos suelo-paisaje basados en características espaciales del entorno, y de mayor comprensión como base para la predicción de atributos edáficos (Cantón, 1999; Gobin *et al.*, 2001, Soil Survey Division Staff, 1993; Pachepsky *et al.*, 2001; Zapata, 2006).

En general, hay que tener presente el vínculo entre los suelos y el paisaje, ya que existe una alta coincidencia entre sus patrones de distribución. De esta manera, el reconocimiento del patrón de uno de ellos permite hacer predicciones del otro, independientemente de la escala de apreciación de dichas relaciones. En este sentido, las relaciones entre los procesos pedogeomorfológicos y la distribución espacial de las propiedades del suelo permiten la predicción de las propiedades del suelo utilizando la información del terreno (Zhu *et al.*, 2001; Park y Burt, 2002).

Entre los modelos discretos propuestos para relacionar las propiedades de los suelos con la posición en el paisaje en áreas montañosas, destaca el modelo *Hillslope* planteado por Walker y Ruhe (1968) y Conacher y Dalrymple (1977). Este modelo se basa en la separación de laderas en diferentes secciones, tales como: cima, hombro, espalda, ladera superior e inferior, pie de ladera. Se ha demostrado que las propiedades del suelo dentro de una sección varían menos que entre secciones, de manera tal que propiedades de suelo diferentes pueden asignarse a cada sección. Estas propiedades pueden ser correlacionadas con las secciones específicas utilizando ecuaciones de regresión (Ovalles, 1984; Ovalles y Collins, 1986).

También, la utilización de variables topográficas derivadas de algunas expresiones matemáticas de las formas superficiales del terreno, tales como pendiente, curvatura y orientación, ha promovido el desarrollo de modelos continuos o mixtos para relacionar las propiedades del suelo con la posición en el paisaje. Estas variables pueden ser computarizadas directamente de los nodos de una grilla y utilizadas para el establecimiento de correlaciones estadísticas con las propiedades del suelo.

En general, tanto los atributos topográficos como los índices derivados de un MDE son relevantes en el estudio del paisaje. Los elementos de formas de terreno o geoformas pueden extraerse de los MDE como constituyentes físicos del relieve, utilizando varios enfoques, incluyendo: a) la combinación de parámetros morfométricos (Dikau, 1989; Shary *et al.* (2002); Ballantine *et al.* (2005) Adediran *et al.*, 2004; Martínez-Zavala *et al.*, 2005; Yang *et al.*, 2007; Iwahashi y Pike, 2007; Saadat *et al.*, 2008); b) lógica borrosa y clasificación no supervisada (Irvin *et al.*, 1997; Macmillan *et al.*, 2000; Burrough *et al.*,

2000; Schmidt y Hewitt, 2004; Arrell *et al.*, 2007); c) clasificación supervisada (Brown *et al.*, 1998; Hengl y Rossiter, 2003; Bolongaro-Crevenna *et al.*, 2005; Ballantine *et al.*, 2005; Reuter *et al.* 2006; Prima *et al.*, 2006); d) clasificación supervisada con algoritmos de agrupamientos probabilísticos (Stepinski y Collier, 2004; Stepinski y Vilalta, 2005); e) estadísticas descriptivas multivariadas (Evans, 1972; Dikau, 1989; Dehn *et al.*, 2001); f) clasificación utilizando redes neuronales artificiales (Ehsani y Quiel, 2007; Ehsani, 2008); y g) clasificación no supervisada mediante aplicaciones neuro-borrosas (Viloria, 2007; Valera y Viloria, 2009; Valera *et al.*, 2010; Viloria *et al.*, 2012; Valera, 2012).

#### **2.4.2. Modelos digitales de elevación y atributos del suelo**

Los parámetros derivados de los MDE se han utilizado en los modelos de variabilidad del suelo para una amplia gama de aplicaciones. Moore *et al.* (1991) fueron pioneros en la utilización de datos derivados del MDE en aplicaciones geomorfológicas, hidrológicas y biológicas. El almacenamiento y la dinámica del carbono orgánico del suelo ha sido frecuentemente el foco del esfuerzo del modelo suelo-paisaje (Bell *et al.*, 2000; Chaplot *et al.*, 2001). También se han modelado con éxito delineaciones hídricas e hidromórficas del suelo (Thompson *et al.*, 1997; Chaplot *et al.*, 2000; Chaplot *et al.*, 2003), la productividad primaria neta (Gessler *et al.*, 2000), drenaje del suelo (Bell *et al.*, 1992), la retención de humedad en el suelo (Pachepsky *et al.*, 2001) y las clases de drenaje (Peng *et al.*, 2003).

La información generada de los modelos digitales se ha aplicado con éxito para predecir la variabilidad del suelo en superficies de estudio menores a 100 ha (Moore *et al.*, 1993; Voltz *et al.*, 1997; Gessler *et al.* 2000; Florinsky *et al.*, 2002; Thompson *et al.*, 2001). En algunos estudios se ha demostrado que la combinación de uno a cinco atributos del terreno derivados de un MDE puede explicar entre 20 % y 88 % de la variación de las propiedades seleccionadas del suelo (Thompson *et al.*, 2006). Otros investigadores entre los cuales destaca Ziadat (2005), han aplicado estos modelos en superficies mayores (14.800 ha), apoyados en técnicas estadísticas. Este autor concluye que los atributos del terreno podrían ser utilizados para representar la predicción de los atributos del suelo, dado los procesos constantes y uniformes que en algunos lugares influyen la variación relacionada con el suelo y el paisaje.

Los argumentos anteriores dejan claro que la correspondencia entre el relieve y las propiedades del suelo, son los factores principales que suministran la información espacial, tanto para la predicción de atributos edáficos como para el estudio de las relaciones suelo-paisaje, a partir de los atributos topográficos derivados del MDE.

En síntesis, las variables topográficas, los atributos edáficos y sus interrelaciones, constituyen parte del marco conceptual y metodológico aplicado en esta investigación. Esta

plataforma es fundamental en el estudio de las relaciones suelo-paisaje, en el análisis de los patrones espaciales de distribución de suelos y en la predicción de propiedades edáficas en paisajes de cuencas altas.

## **2.5. TÉCNICAS DE INTELIGENCIA ARTIFICIAL**

La inteligencia artificial (IA) o inteligencia computacional es la rama de la informática que desarrolla procesos que intentan imitar a la inteligencia humana. La principal aplicación de esta ciencia es la creación de máquinas para la automatización de tareas que requieran un comportamiento inteligente. Es decir, se basa en un conjunto de algoritmos, cuyo propósito es imitar el razonamiento humano a través de una lógica deductiva o manipulación de símbolos (Schalkoff, 1990; Martín del Brío y Sanz, 2007). El objetivo de la IA es lograr la construcción y la comprensión de entidades inteligentes, que permitan resolver los mismos problemas que resuelven los humanos (Russell y Norvig, 2004).

La IA se encarga principalmente de estudiar y crear algunos programas que sean capaces de imitar el razonamiento del ser humano. Los sistemas con IA deben ser capaces de responder y diagnosticar problemas que se presentan en la vida diaria (Elguea, 1987).

Dentro de la IA se encuentran algunos modelos alternativos o técnicas emergentes como la lógica borrosa o difusa (LB) y las redes neuronales artificiales (RNA), las cuales están inspiradas en las soluciones que la naturaleza ha encontrado durante millones de años de evolución a numerosos problemas que tratan cantidades masivas de información, de tipo redundante, imprecisa y ruidosa (Martín del Brío y Sanz, 2007), tales como la que presenta el complejo multivariado suelo-paisaje.

### **2.5.1. Redes neuronales artificiales**

Una red neuronal artificial (RNA) es un modelo computacional inspirado en redes neuronales biológicas, que pueden ser consideradas como un sistema de procesamiento de información con características particulares como aprendizaje a través de ejemplos, adaptabilidad, robustez, capacidad de generalización y tolerancia a fallas. La RNA puede ser definida como una estructura distribuida, de procesamiento paralelo, formada de neuronas artificiales (llamados también elementos de procesamiento), interconectados por un gran número de conexiones (sinapsis), las cuales son usadas para almacenar conocimiento que estará disponible para su posterior utilización (Wasserman, 1989; Hilera y Martínez, 1995).

Para Atkinson y Tanall (1997), una RNA es un modelo matemático complejo, basado en ecuaciones no lineales, en el que se introducen unas variables de entrada (*e.g.* atributos topográficos derivados de un MDE) y se obtiene una respuesta o datos de salida (*e.g.* unidades o clases geomorfométricas, clases de suelo). La red neuronal se ajusta con la

finalidad de obtener resultados parecidos a los introducidos en un proceso previo de aprendizaje o entrenamiento

Las RNA simulan el proceso de resolución de problemas espaciales por parte del cerebro humano. Aplicado a la clasificación basada en la información multispectral derivada de imágenes satelitales, o multivariantes derivadas de MDE, suele ser más eficiente y requiere menos entrenamiento que los algoritmos de clasificación estadísticos o tradicionales. Además, en su diseño y desarrollo no hay que asumir ninguna hipótesis estadística de partida como funciones de distribución de variables, ni dependencias multivariantes (Kimes *et al.*, 1998).

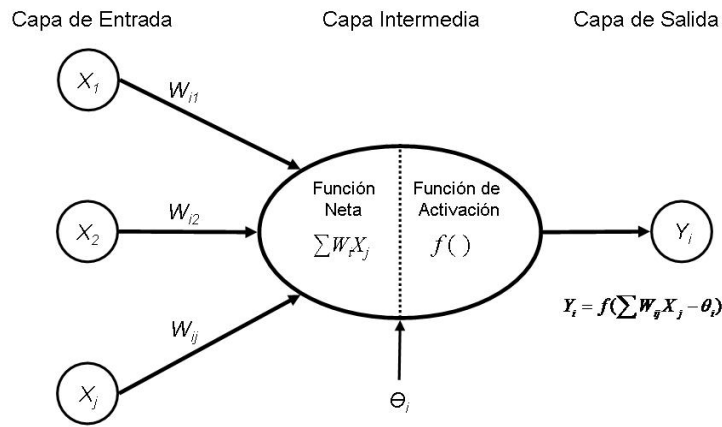
Los enfoques anteriores son resumidos por Martín del Brío y Sanz (2007), quienes señalan que las RNA son sistemas de procesamiento, que copian esquemáticamente la estructura neuronal del cerebro para tratar de reproducir sus capacidades. De esta manera, las RNA son capaces de aprender de la experiencia a partir de las señales o datos provenientes del exterior, dentro de un marco de computación paralela y distribuida, fácilmente implementable en dispositivos *hardware* específicos.

#### ***2.5.1.1. Funcionamiento de una red neuronal artificial***

Las RNA están formadas por un conjunto de elementos simples de procesamiento llamados nodos o neuronas que están organizadas en capas: a) una primera capa de entrada ( $X_1 \dots X_j$ ), que recibe información del exterior, b) una capa oculta (intermedia) encargada de realizar el trabajo de la red, y c) una capa de salida ( $Y_i$ ), que proporciona el resultado del trabajo de la red al exterior (Figura 2-1). Cada neurona cuenta con una cantidad de variables de entrada ( $X_j$ ) y está conectada con otras neuronas mediante enlaces de comunicación, cada uno de los cuales tiene asociado un peso sináptico ( $W_{ij}$ ), que representa la información que será utilizada por la red neuronal para resolver un problema determinado. A su vez dispone de una salida ( $Y_i$ ) que transmite la información al exterior o hacia otras neuronas. (Sotolongo *et al.*, 2000).

En general, la neurona de salida  $Y_i$  recibe las señales de las neuronas de entrada ponderadas por los pesos de las conexiones ( $\sum W_{ij}X_j$ ). La suma de estas señales ponderadas proporciona la entrada total o neta de la neurona y, mediante la aplicación de una función matemática (función de activación), sobre la entrada neta, se calcula un valor de salida, el cual es enviado a otras neuronas. Si el resultado de la función  $f$  es mayor que el valor umbral ( $\theta_i$ ), la neurona se activa y emite una señal hacia las neuronas de la capa siguiente. Pero, si por el contrario, el resultado es menor que el valor umbral, la neurona permanece inactiva y no envía ninguna señal.





**Figura 2-1.** Esquema del funcionamiento general de una neurona artificial.

Del funcionamiento de una RNA, es necesario hacer énfasis en las tres grandes partes que la constituyen. La primera parte es la *función de activación*, la cual define la salida de una neurona en términos del nivel de activación de la neurona de entrada. Esta función umbral es generalmente alguna forma de función matemática no lineal (Buendía *et al.*, 2002).

La segunda parte se compone de la *topología o arquitectura* de las redes neuronales, que consiste en la organización y disposición de las neuronas en la red formando capas o agrupaciones de neuronas más o menos alejadas de la entrada y salida de la red. En este sentido, los parámetros fundamentales de la red son: a) el número de capas, b) el número de neuronas por capa, c) el grado de conectividad, y d) el tipo de conexiones entre neuronas (Haykin, 1994; Salas, 2000).

La tercera parte se corresponde con la fase de *aprendizaje y/o entrenamiento* considerado como el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada a través de una época (se denomina época cuando todo el conjunto de entrenamiento ha sido presentado a la red). Los cambios que se producen durante el proceso de aprendizaje se reducen a la modificación y creación de conexiones entre neuronas (Haykin, 1996). En el entrenamiento de una red neuronal tanto el peso sináptico de las conexiones como el valor umbral para cada neurona se modifican (según un algoritmo de aprendizaje), con el fin de que los resultados generados por la red coincidan con los resultados esperados. En esta fase de aprendizaje el valor umbral se expresa como un peso sináptico más, pero asociado a una neurona activa. Esta neurona activa, se denomina "*bias*", y se sitúa en la capa anterior a la señal de salida.

El proceso de aprendizaje consiste en encontrar los pesos que codifican los conocimientos. Una regla de aprendizaje hace variar el valor de los pesos de una red hasta que estos adoptan un valor constante, cuando esto ocurre se dice que la red ha sido ajustada y que ha

aprendido. Esta característica es fundamental, ya que permite a la red responder correctamente ante información novedosa, distorsionada o incompleta (Palmer y Montaña, 1999).

### **2.5.1.2. Arquitectura o topología de las RNA**

Las neuronas que componen una RNA se organizan de forma jerárquica formando capas. (Hilera y Martínez, 1995; Palmer y Montaña, 1999). Dependiendo de la estructura de la red neuronal existen varias tipologías, algunas de estas son: Perceptrón (McCulloch y Pitts, 1943), adaline (Widrow, 1960), perceptrón multicapa (Rosenblatt, 1958), máquina de Bolzman (Hinton *et al.*, 1984), máquina de Cauchy (Takefuji y Szu, 1989), propagación hacia atrás (*backpropagation*) (Rumelhart *et al.*, 1986), redes de Elman (1990), redes de Hopfield (1982), red de contrapropagación (Hecht-Nielsen, 1987), redes de neuronas de base radial (Chen *et al.*, 1991), redes de neuronas de aprendizaje competitivo (Rumelhart y Zipser, 1985), mapas Autoorganizados (Kohonen, 1982), redes ART (teoría de resonancia adaptativa, *Adaptive Resonance Theory*) (Carpenter y Grossberg, 1988).

La arquitectura de red más ampliamente utilizada es la denominada red de retropropagación (*Backpropagation*). El 90% de las aplicaciones de redes neuronales utilizan esta arquitectura (Welstead, 1994). Sin embargo, en esta investigación se hace énfasis en las redes de mapas auto-organizados, ya que la base teórica de este tipo de red en combinación con la lógica borrosa, conforma el fundamento conceptual de las redes neuro-borrosas.

### **2.5.1.3. Los mapas auto-organizados de Kohonen**

Los mapas auto-organizados o SOM (*Self-Organizing Map*) desarrollados por Kohonen (1982), son un método de aprendizaje no supervisado para la clasificación de datos multidimensionales. La principal característica de esta técnica es la creación de mapas bidimensionales, en los que los contenidos que estén relacionados aparecerán juntos, y cuanto más similar sean dos datos más próximos estarán en el mapa.

Los SOM, son apropiados para la identificación de unidades de paisaje cuando se utilizan los MDE, ya que son capaces de reducir la dimensionalidad de datos complejos e identificar patrones, permitiendo una fácil visualización, el análisis y la comprensión de clases o unidades geomorfológicas conformadas por grandes agrupaciones de píxeles. En el campo geomorfológico, el insumo de la red es el conjunto de atributos topográficos para cada píxel. Este análisis permite: a) la agrupación de un conjunto píxeles con valores de atributos topográficos en función de características similares, b) la producción de unidades geomorfológicas con el centro de las clases, para cada unidad generada, c) el cálculo de estimados de la distancia relativa entre cada grupo, y d) la representación bidimensional de la clasificación en forma de un mapa, que muestra la distribución espacial de las clases

geomorfométricas, con valores y variabilidad espacial similar (Klein *et al*, 2007; Ehsani y Quiel, 2007; Ehsani y Quiel, 2009).

Esta red intenta imitar la capacidad que tiene el cerebro de formar mapas característicos de la información recibida del exterior. El objetivo de las investigaciones de *Kohonen* (1982), fue demostrar que un estímulo externo, suponiendo una estructura propia y una descripción funcional del comportamiento de la red, era suficiente para forzar la formación de mapas.

La red neuronal Kohonen contiene solamente una capa de neuronas y una capa de entrada, que se ramifica para todos los nodos. Pertenece a la categoría de redes competitivas o mapas de auto organización, es decir tiene un aprendizaje no supervisado. Tiene también funciones de activación lineales y flujo de información unidireccional (red en cascada). La red cuenta con  $N$  neuronas de entrada y  $M$  de salida, cada una de las neuronas de entrada está conectada a todas las de salida. Por ejemplo, cada neurona de entrada puede corresponder a una variable derivada de un MDE (altitud, pendiente, orientación, curvatura) y cada neurona de salida puede corresponder a una clase geomorfométrica. Las unidades de entrada reciben datos continuos normalizados. La red clasifica los patrones de entrada en grupos de características similares, de tal manera que cada grupo activa siempre las mismas salidas (Cardona, 2001).

Según Kalteh *et al.* (2007) el procedimiento requerido para la aplicación de los SOM puede dividirse en tres categorías: a) agrupación y normalización de datos; b) entrenamiento de la red, y c) extracción de la información de los SOM entrenados. Después de realizado el ajuste de los SOM, el mapa resultante puede ser post-procesado basado en la visualización, en la clasificación y en la propuesta de modelos locales.

La emergencia de algoritmos como los SOM en combinación con la creciente disponibilidad de MDE de alta resolución de fuentes fidedignas, ha contribuido al desarrollo y a la aplicación de procedimientos para la identificación automatizada de formas de terreno, cada vez más viables y económicos (Ehsani y Quiel, 2007). Es decir, desde el punto de vista geomorfológico, los SOM entrenados son una valiosa herramienta para la visualización y organización de grandes cantidades relativas de datos de atributos topográficos, y para la definición y delimitación de unidades de paisaje.

### **2.5.2. Teoría de los conjuntos borrosos o difusos**

La teoría de los conjuntos borrosos (*fuzzy set theory*) desarrollada por Zadeh (1965) es una generalización del álgebra booleana (Burrough, 1989), y es particularmente útil para procesar datos e información cuyos atributos tienen zonas de transición gradual. En un agrupamiento convencional los conjuntos son estrictamente excluyentes, separados por límites bien definidos. Los conjuntos borrosos, en cambio, tiene límites graduales entre sí.

De manera más formal, un conjunto borroso  $A$  de  $x$  (donde  $x$  constituye el universo en consideración) es formado por el par  $(A, \mu_A(x))$ , donde  $x$  es un elemento de  $A$ , y  $\mu_A$  es una función cuya imagen pertenece al intervalo cerrado  $[0, 1]$ . Esta última es denominada función de pertenencia de  $x$  al subconjunto  $A$  (o función de asociación o membresía). Un valor de  $\mu_A(x)$  cercano a 1 significa que el grado de pertenencia de  $x$  en  $A$  es alto. La pertenencia es absoluta cuando  $\mu_A(x)$  es igual a 1 y la no pertenencia absoluta corresponde a  $\mu_A(x)=0$  (Lamedada y Rojas, 2004; Orellana, 2004).

La lógica borrosa (*fuzzy logic*) ofrece las herramientas teóricas para tratar conceptos expresados en lenguaje natural. Por lo tanto, describe la posibilidad de que un individuo sea miembro de un conjunto determinado, cuya opción puede estar basada en el conocimiento subjetivo de un experto (Burrough y McDonnell, 1998).

La teoría de conjuntos borrosos parte de la teoría clásica de conjuntos, añadiendo una función de pertenencia al conjunto, que se define como un número real entre 0 y 1. El conjunto borroso está asociado a un determinado valor lingüístico, definido por una palabra, adjetivo o etiqueta lingüística. Para cada conjunto borroso se define una función de pertenencia o inclusión  $\mu_A(x)$ , que indica el grado en que la variable  $x$  está incluida en el concepto representado por la etiqueta  $A$  (Martín del Brío y Sanz, 2007).

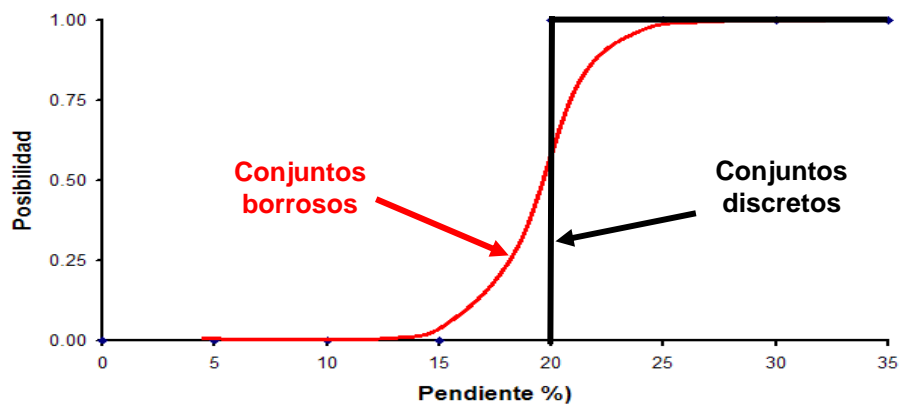
Los métodos de clasificación borrosa comenzaron a suplantar los métodos discretos de clasificación multivariada, debido a que aquellos proporcionan un solapamiento para las clases, lo cual es a menudo una característica de los grupos naturales en las ciencias ambientales. Estos métodos de clasificación borrosa tienen la capacidad de modelar la variación continua de los suelos de acuerdo con los valores de la función de membresía. En este sentido, la teoría de conjuntos borrosos (*fuzzy set*) permite la definición de límites suaves y clases borrosas de formas de terreno para ser representadas explícitamente, proporcionando más información acerca de la estructura del paisaje que las clases de formas de terreno discretas.

En este tipo de clasificación, el grado con el que un individuo pertenece a una clase determinada no es expresado en términos binarios (0 ó 1), sino de acuerdo a un valor de membresía (función de pertenencia o inclusión). La función de membresía es continua y varía en forma gradual desde 0 hasta 1 ó entre 0 y 100 (Burrough *et al.*, 2000).

Lo descrito anteriormente significa que con el uso de la lógica clásica o tradicional es muy difícil modelar cuando la información con respecto a un fenómeno es totalmente desconocida, y no se debe aplicar cuando tal información es imprecisa e incompleta; mientras que la lógica borrosa es una técnica cartográfica que ayuda a la representación de la información imprecisa, incompleta o incierta.

La teoría matemática de la lógica borrosa se basa en la existencia de diferentes grados de pertenencia entre lo falso y lo verdadero, es decir, valores como “moderado”, “bajo”, “alto”, entre otros (variables lingüísticas), en lugar de expresarse como “sí” o “no” similar al concepto del álgebra *booleana*. Mediante la lógica borrosa, un problema se puede representar en términos de conjuntos borrosos, los cuales pueden derivarse de procedimientos cuantitativos o cualitativos (Prakash, 2003; Ramírez, 2007).

A manera de ejemplo, al evaluar si una pendiente es fuerte o no, es posible definir una función de membresía borrosa de manera tal que una pendiente de 10% tiene una membresía de 0, y una pendiente de 25% tiene una membresía de 1; pero entre 10% y 25% los valores se incrementan gradualmente en la escala de 0 a 1 (Figura 2-2) (Eastman, 2003; Gutiérrez y Jegat, 2005).



**Figura 2-2.** Límites difusos o de transición gradual y límites discretos.  
**Fuente:** Eastman (2003).

El ejemplo considerado contrasta con los conjuntos precisos clásicos que tienen límites distintivos y discretos; aunque un conjunto preciso también puede ser considerado como un caso especial de los conjuntos borrosos donde la membresía cambia instantáneamente de 0 a 1. Sin embargo, la teoría de lógica borrosa proporciona una excelente base matemática para entender problemas de decisión y para construir reglas de decisión en combinación con la evaluación de criterios (Gutiérrez y Jegat, 2005).

La mayoría de los estudios de clasificaciones automatizadas emplean el algoritmo denominado *fuzzy-c-means* (c-medias borroso). El agrupamiento borroso (FCM) (Bezdek, 1981), basa su clasificación en el criterio del error mínimo cuadrático, forma grupos borrosos de los valores del vector de entrada y da un valor de pertenencia entre 0 y 1 a cada uno de los grupos borrosos, generando una matriz con el grado de pertenencia a cada grupo borroso. Cada grupo tiene un centro en el espacio, para cada dato el valor de pertenencia a los diferentes grupos depende de la distancia a los centros de las clases. El algoritmo agrupa los datos de acuerdo a su proximidad, por lo que se encuentra influenciado por la

## **DEDICATORIA**

Al Dios Todopoderoso  
A la memoria de mis padres: Ligia Rosa y Manuel  
A mi esposa, Lidia Yaxira  
A mis hijas, Liangie Mayela y  
Karelys Nazareth

forma como se define la distancia entre los datos, es decir, las reglas utilizadas y el rango de variación de las variables.

En el ámbito geomorfológico, las clases morfométricas del terreno o clases geomorfométricas obtenidas de la combinación de atributos derivados de los MDE, son más informativas y útiles si son utilizados los conjuntos borrosos, ya que en este contexto, cualquier píxel puede contener elementos de un número diferente de clases. En efecto, la función de pertenencia de un determinado píxel a una clase morfométrica indica el grado de similitud de ese píxel al concepto central de dicha clase.

Las unidades de relieve o clases geomorfométricas, tales como crestas, picos, laderas, canales, y zonas planas, han sido ampliamente extraídas de MDE (Wood, 1996). Sin embargo, la definición de las clases geomorfométricas como objetos claramente delimitado ha sido cuestionada y se han reconceptualizado como vagos, imprecisos, o como conjuntos borrosos (Irvin *et al.*, 1997; Lagacherie *et al.*, 1997; Fisher, 2000; Fisher *et al.*, 2004). La clase morfométrica es representada por el grado de membresía o función de pertenencia (en un rango de valores de cero a uno) que cualquier celda tiene para cada una de las formas de terreno identificadas. Un grado de membresía de uno (1) estaría asociado con una celda que satisface exactamente el valor ideal de los atributos (el concepto central) de una clase particular, y un valor de cero (0) indicaría que la celda no tiene similitud o membresía para esa clase (Arrell *et al.*, 2007).

En el campo edafológico, el concepto central de las clases de suelo está representado por el núcleo central de la clase, en la que los individuos están relacionados en diferentes grados. De esta manera, los perfiles de suelo más próximos al núcleo son similares a él, y los ubicados en el margen serán disimilares (Morales y Vilorio, 2006).

### **2.5.3. Sistemas neuro-borrosos: integración de RNA y LB**

La combinación de redes neuronales artificiales y la teoría de los conjuntos borrosos han dado origen a un enfoque de gran fortaleza conocido como sistema neuro-borroso (SNB). En la literatura consultada, existe cierto grado de confusión en la terminología empleada, utilizándose, entre otros, los siguientes términos: *Hybrid neural net*, *regular fuzzy neural net*, *hybrid fuzzy neural net*, *fuzzy neural net*, *neural fuzzy net*, *neuro-fuzzy net*. Cada autor utiliza distintos términos para conceptos similares, pudiendo una misma expresión tener diferente significado dependiendo del autor consultado (Puente *et al.*, 2004).

Los enfoques neuro-borrosos constituyen una vía para encontrar de modo heurístico o empírico los parámetros de un sistema borroso a través del procesamiento de datos, sometidos a un algoritmo de entrenamiento. Deben considerarse como herramientas que ayudan a desarrollar modelos borrosos, pero no los crean de forma automática. La

implementación de un SNB siempre debe permitir al usuario supervisar e interpretar el proceso de aprendizaje. Al igual que con las redes neuronales, el éxito del proceso de aprendizaje no está garantizado. Además, la solución que se persigue no debe ser evaluada exclusivamente en términos de obtención de un alto porcentaje de aciertos en la clasificación de nuevos casos, sino en términos de interpretabilidad y sencillez del modelo de clasificación resultante (Nauck y Kruse, 1998).

En general, estos modelos combinan la capacidad de aprendizaje y adaptación de las RNA y el poder de la interpretación lingüística de la LB. Esto conlleva a la obtención final de un método que permite combinar información auxiliar de todo tipo, ya sean variables continuas derivada de MDE (*e.g.* elevación, pendiente), información espectral de imágenes de satélite (*e.g.* índice de vegetación) o variables discretas derivadas de mapas temáticos (*e.g.* litología, tipos de suelo) en el proceso de clasificación. Como resultado, cada objeto (o píxel) puede presentar cierto grado de pertenencia a diferentes clases.

#### ***2.5.3.1. Red de agrupamiento borroso de Kohonen (FKCN)***

La integración de las redes neuronales autoorganizadas (Kohonen, 1982) y los conjuntos borrosos c-medias (FCM) (Bezdek *et al.*, 1992), ha sido posible debido a que ambas técnicas presentan un denominador común basado en su orientación hacia el tratamiento de tareas. Estas tareas involucran el procesamiento de cantidades masivas de información de tipo redundante, imprecisa, incierta y con ruido. Ambos enfoques son útiles para modelar procesos complejos no lineales (Martín del Brío y Sanz, 2007).

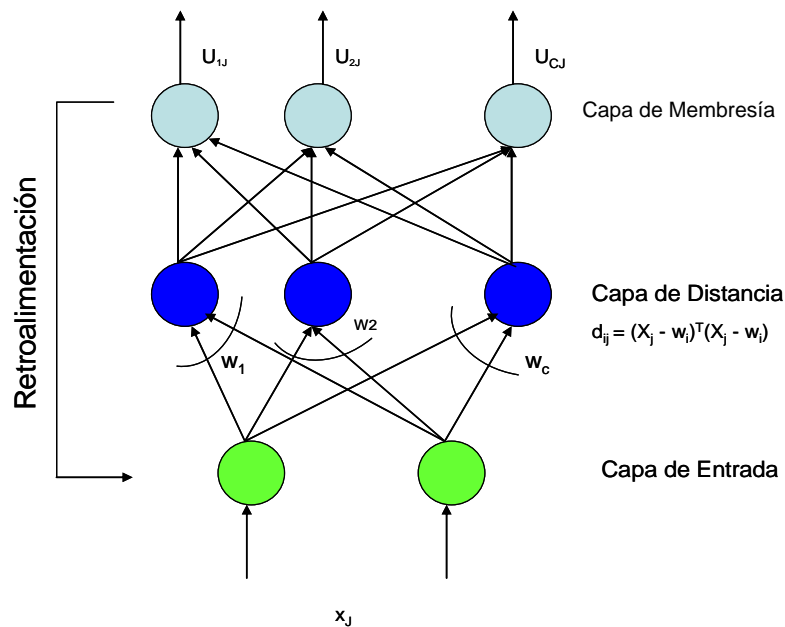
Los mapas autoorganizados de Kohonen (1982) han llamado mucho la atención debido principalmente a la sencillez y eficiencia de su aprendizaje. Sin embargo, según Lin y Lee (1996) el algoritmo SOM tiene algunas deficiencias. En primer lugar, el algoritmo se basa en un procedimiento heurístico (emplea iteraciones de ensayo y error), por lo que la condición de parada no se fundamenta en un modelo de optimización del proceso y de los datos. En segundo lugar, los pesos finales de las neuronas dependen estrictamente del orden en el cual se presentan los datos de entrada a la red. Finalmente, el algoritmo sólo ajusta en un instante de tiempo  $t$  el peso de aquellas neuronas que resultaron activadas por una entrada  $X_j$ .

Para solventar los inconvenientes descritos, Bezdeck *et al.* (1992) presentaron una propuesta basada en la integración del algoritmo FCM con el coeficiente de aprendizaje y las estrategias de adaptación de los SOM. La integración de ambos algoritmos es una manera de solucionar las dificultades planteadas del algoritmo SOM, ya que FCM se basa en un proceso de optimización y los SOM no. Además, con esta integración el algoritmo



SOM genera salidas con valores borrosos, en lugar de clases con límites abruptos (Viloria, 2007).

La estructura topológica de la red neuronal borrosa de Kohonen (FKCN) (Lin y Lee, 1996), consta de tres capas: a) la de entrada, b) la de distancias, y c) la de funciones de membresías. La capa de entrada contiene los valores normalizados de las variables empleadas; la capa de distancias tiene un número de neuronas equivalente al número preestablecido de clases, y la capa de membresía es una capa extra a la capa de salida de la red original de Kohonen (Figura 2-3).



**Figura 2-3.** Estructura de la red neuronal borrosa de Kohonen (FKCN).  
**Fuente:** Lin y Lee (1996).

En la capa de distancia se calcula la separación  $d_{ij}$  existente entre un patrón de entrada  $X_j$  y el peso del nodo  $w_i$ , con  $i = 1, 2, \dots, c$ , donde  $c$  representa el número de clases del modelo a estimar. Posteriormente, la capa de membresía grafica las distancias  $d_{ij}$  en valores de membresía  $U_{ij}$ , donde  $U_{ij}$  representa el grado de pertenencia de un patrón de entrada  $X_j$  a una clase  $c$ . Debe tenerse en cuenta que a lo largo del proceso de aprendizaje ocurre una retroalimentación de la capa de funciones de pertenencia a la capa de distancia, con la finalidad de ajustar los centros de clase.

## 2.6. APLICACIONES DE TÉCNICAS IA EN LA CIENCIA DEL SUELO

### 2.6.1. Las redes neuronales artificiales en la Ciencia del Suelo

La mayoría de las publicaciones acerca de las aplicaciones de las RNA están referidas al campo de los sensores remotos (Skidmore *et al.*, 1997) y a la evaluación de la estabilidad

de pendientes (Guzzetti *et al.*, 2008). Existen pocos ejemplos de la aplicación de RNA en la evaluación de la distribución espacial de procesos geomorfológicos y formas de terreno (Leverington y Duguay 1997; Brown *et al.* 1998). Sin embargo, los modelos de RNA son más poderosos y ventajosos que la regresión múltiple cuando se modelan relaciones con influencias complejas y no lineales como los procesos geomorfológicos (Lek y Guegan, 1999) y las relaciones suelo-paisaje (Zhu, 2000).

Algunos ejemplos generales donde destaca la aplicación de las RNA en el ámbito geomorfológico son: a) discriminación de unidades litológicas usando datos de imágenes Landsat TM (Mather *et al.*, 1998), b) clasificación del uso de tierras urbanas (Özkan y Sunar Erbek, 2005), c) planificación ecológica (Lek y Guegan, 1999); d) análisis morfométrico de atributos (Ehsani y Quiel, 2007), e) análisis de elementos del paisaje (Ehsani y Quiel, 2009; Ehsani, 2008).

En el ámbito edafológico, las investigaciones más destacadas incluyen: a) cartografía digital de suelos (Zhu, 2000; Behrens *et al.*, 2005), b) clasificación de suelos con fines ingenieriles (Cal, 1995), c) clasificación de suelos con fines agrícolas (Fidencio *et al.*, 2001; Chagas *et al.*, 2007; Dragovic y Onjia, 2007), d) modelo de relaciones suelo-paisaje (Zhu, 2000), e) determinación de perfiles de suelos representativos (Caglar y Arman, 2007), f) evaluación de la calidad biológica del suelo (Mele y Crowley, 2008); g) predicción de propiedades del suelo (Sarmadian *et al.*, 2009); h) contenido de carbono orgánico (Somaratne *et al.*, 2005; Takata *et al.*, 2007); i) capacidad de intercambio catiónico del suelo (Keshavarzi y Sarmadian, 2010); j) distribución de tamaño de partículas (Zhai *et al.*, 2006; Zhao *et al.*, 2009), k) contenido de humedad en los suelos (Pachepsky *et al.*, 1996; Schaap y Leij, 1998; Koekkoek y Booltink, 1999; Elshorbagy y Parasuraman, 2008); l) pérdida de sedimentos: (Sarangi y Bhattacharya, 2005); m) estimación de la erosión del suelo y la concentración de nutrientes (Kim y Gilley, 2008); n) en agricultura de precisión para evaluar rendimientos de maíz y soya (Kaul *et al.*, 2005), y rendimiento y calidad del grano de maíz (Miao *et al.*, 2006); y o) en el desarrollo de funciones de pedotransferencias de propiedades hidráulicas (Tamari *et al.*, 1996; Schaap *et al.*, 1998; Akbulut, 2005; Agyare *et al.*, 2007; Minasny y McBratney, 2002; Minasny *et al.*, 2004; Merdum *et al.*, 2006; Lamorski *et al.*, 2008; Baker y Ellison, 2008).

La mayoría de estos estudios utilizan RNA basadas en perceptrón multicapa y los mapas auto-organizados (SOM). Las ventajas de su aplicación incluyen primero, el hecho de no estar limitadas por la disponibilidad de expertos en un área determinada, y segundo, son unas herramientas eficientes para analizar rasgos geomorfométricos, aparte de que los resultados son reproducibles y consistentes.

### 2.6.2. La lógica borrosa en la Ciencia del Suelo

Dentro de las aplicaciones más relevantes de la lógica borrosa, relacionadas con el campo geomorfológico y edafológico, basadas en los atributos generados de MDE, se tienen las siguientes investigaciones:

a) Clasificación de unidades geomorfológicas (Irvin *et al.*, 1997; Burrough *et al.*, 2000; Macmillan *et al.*, 2000; Triantafilis *et al.*, 2001; Schmidt y Hewitt, 2004; b) análisis morfométrico del paisaje (Arrel *et al.*, 2007; Klingseisen *et al.*, 2007; Carvalho y Bayer 2008); c) clasificación de suelos (McBratney *et al.*, 1992; Odeh *et al.*, 1992; Mazaheri *et al.*, 1995; Lagacherie *et al.*, 1997; Bragato, 2004; Chen *et al.*, 2005; Lagacherie y McBratney, 2006; Boruvka *et al.*, 2008; Bhargavi y Tech, 2010); d) levantamiento de suelos y evaluación de tierras (Burrough, 1989; Kollias *et al.*, 1999); e) modelos de relaciones suelo-paisaje (Zhu *et al.*, 1997; De Bruin y Stein, 1998); f) cartografía digital de suelos (Zhu, 2001; Shi *et al.*, 2004; Burt *et al.*, 2006; Balkovic *et al.*, 2007; Zhu *et al.*, 2008; Yang *et al.*, 2011; Sun *et al.*, 2012); g) predicción de propiedades del suelo (Zhu *et al.*, 2006; Zhu *et al.*, 2008; Yang *et al.*, 2007); h) zonas para el manejo de sitio específico (Li *et al.*, 2007); i) ecología de paisaje (Burrough *et al.*, 2006); j) evaluación de la calidad de los suelos (Oberthür *et al.*, 2000; Torbert *et al.*, 2008), y k) en la evaluación del cambio de cobertura de la tierra (Zachwatowicz, 2011).

Las investigaciones referidas destacan la importancia del enfoque borroso en los estudios de las relaciones suelo-paisaje, ya que consideran las características de vaguedad, incertidumbre y transición gradual en la variación de los suelos en las diversas formas de terreno. Es decir, el agrupamiento de datos de los atributos del terreno mejora el modelo de relación suelo-paisaje convencional, debido a que permite la representación de la borrosidad inherente a las unidades suelo-paisaje.

De igual manera, muchos de los estudios destacan que los métodos de clasificación numérica basados en agrupamientos borrosos, proporcionan una alternativa a la delineación manual de formas de terreno por fotointerpretación, ya que esta última es subjetiva y requiere mucho conocimiento del paisaje en estudio.

Las unidades obtenidas por clasificación borrosa proporcionan una base para el desarrollo de procedimientos agronómicos y opciones de manejo que podrían aplicarse a la mayoría de los paisajes agrícolas de interés. También pueden ser utilizadas en el modelado de paisajes, para apoyar la cartografía de suelos, la delineación de mapas de vegetación y la ecología de paisaje, en aspectos de uso y manejo de la tierra, en aplicaciones relacionadas a la agricultura de precisión y en estudios de degradación de los suelos.

### **2.6.3. Los Sistemas neuro-borrosos en la Ciencia del Suelo**

Son escasos los trabajos de investigación realizados en el ámbito geomorfológico y edafológico que toman en cuenta la combinación de las redes neuronales artificiales y la teoría de los conjuntos borrosos. En Venezuela, se ha dado un paso importante en la utilización de tecnologías de inteligencia artificial computacional en el área de clasificación de paisajes y predicción de atributos del suelo con los trabajos de Viloria (2007), Valera y Viloria (2009), Valera *et al.* (2010), Núñez (2011) y Valera (2012).

Viloria (2007) implementó el algoritmo neuro-borroso en un sistema de clasificación del terreno, el cual permite la obtención de clases digitales del paisaje. El estudio se realizó con datos derivados de una imagen satelital y un MDE de un sector de la cuenca del río Caramacate, estado Aragua. A partir de estos datos se clasificó el paisaje del área de estudio por medio de tres modelos de redes neuronales: una red de agrupamiento borroso de Kohonen (FKCN), y dos mapas autoorganizados (SOM) con distintas funciones de vecindad. La comparación visual de la clasificación provista por el experto con los tres modelos obtenidos, reveló que el mejor modelo correspondió al FKCN. Este último constituye una representación continua del paisaje en dos sentidos: primero, porque el paisaje es representado por una matriz de píxeles, y segundo porque la clasificación de cada píxel es borrosa, representada por un vector de similitud para diferentes clases. Este estudio constituye una aplicación de apoyo al experto, mediante una clasificación no supervisada que le suministra una visión integral de la estructura del paisaje.

En general, las investigaciones demuestran la alta capacidad de las RNA y de los conjuntos borrosos para el análisis morfométrico del terreno como apoyo a la evaluación morfológica, y para la predicción de propiedades edáficas en áreas de interés con escasa disponibilidad de datos de suelo, siempre y cuando exista adecuada información auxiliar derivada de MDE e imágenes de satélite.

## **2.7. CARTOGRAFÍA DIGITAL DE SUELOS**

La cartografía digital de suelos - CDS (*Digital Soil Mapping*), se ha definido como la creación y generalización de los sistemas de información espacial de suelos a través de métodos numéricos que infieren las variaciones espaciales y temporales de tipos y propiedades de suelos, a partir de la relación entre observaciones y variables ambientales (Lagacherie y McBratney, 2006).

La CDS es la operacionalización de ideas basada en sistemas computarizados para predecir la distribución de los suelos en el paisaje (Hewitt, 1993; Lagacherie *et al.*, 1997; McBratney *et al.*, 2003; Scull *et al.*, 2003). También se le conoce como modelado predictivo del suelo (Hewitt, 1993; Scull *et al.*, 2003) y levantamiento cuantitativo de

suelos (McKenzie y Ryan, 1999) o cartografía y modelado digital de suelos (Grunwald, 2006; Bui, 2007). En el presente siglo pareciera que esta técnica constituye un cambio en el paradigma suelo-paisaje (McBratney *et al.*, 2003).

La base de la CDS está en la relación entre covariables ambientales (valores predictores o independientes) y la distribución de los suelos (propiedades o clases) en el paisaje. La historia de la CDS está marcada por la adopción de nuevas herramientas y técnicas para analizar, integrar y visualizar conjuntos de datos del suelo y el medio ambiente. La introducción de las tecnologías digitales, la velocidad de procesamiento de datos, el manejo de datos espaciales, los métodos cuantitativos para describir los patrones y los procesos del suelo y los métodos científicos de visualización han proporcionado nuevas oportunidades para predecir las propiedades y los procesos del suelo (Grunwald, 2006).

Para representar la variabilidad del suelo se utilizan las técnicas de cartografía o el modelado de patrones edáficos. El modelado es definido como "el uso de ecuaciones matemáticas para simular y predecir los acontecimientos reales y los procesos, con el apoyo de métodos de campo y laboratorio, junto con los sistemas de inferencia espacial y no espacial del suelo" (Grunwald, 2009).

### **2.7.1. Aplicaciones de las técnicas de cartografía digital de suelos**

Las técnicas de CDS emergen como una alternativa para apoyar las técnicas tradicionales de levantamientos de suelos. Sin embargo, la mayoría de sus aplicaciones se han realizado en levantamientos de suelos, detallados o semidetallados, donde la variabilidad de los suelos está determinada por un limitado número de factores formadores del suelo (Mora *et al.*, 2008).

Con el desarrollo avanzado de la computación y las tecnologías de la información, conjuntamente con la disponibilidad de nuevos tipos de sensores remotos, se ha desarrollado un enfoque más cuantitativo que puede reemplazar las técnicas de inventario tradicionales. Las nuevas técnicas incluyen el modelado de superficies continuas basadas en los factores de formación de suelos, así como también la evaluación de la exactitud y la incertidumbre de las predicciones (McBratney *et al.*, 2000).

En la CDS es posible la utilización de un número limitado de observaciones del suelo, ya que posteriormente, esas observaciones son relacionadas con información auxiliar derivada de MDE e imágenes de satélite, que representa los factores formadores del suelo. En este sentido, los MDE representan el relieve, las imágenes de satélite representan la cobertura del terreno y el clima, y los mapas geológicos representan el material parental y posiblemente al tiempo. Ahora, esas relaciones pueden ser utilizadas para predecir

propiedades del suelo para áreas completas, en las cuales exista disponibilidad de información auxiliar (Stoorvogel *et al.*, 2008; Mora *et al.*, 2008).

La CDS depende de la identificación de las relaciones entre las variables edáficas y las covariables auxiliares. Estas últimas tienen componentes de variación que corresponden a diferentes escalas, tanto que un análisis de variación de la escala espacial puede ser de gran valor práctico y suministrar información para una apropiada predicción de las propiedades del suelo. La CDS hace extensivo el uso de información auxiliar, tal como la contenida en las imágenes de sensores remotos y los MDE; sin embargo, no puede ejecutarse sin tomar en cuenta las muestras del suelo propiamente dichas (Lark, 2005).

En las primeras aplicaciones las observaciones del suelo eran relacionadas con mapas de atributos del terreno utilizando modelos de regresión simple, pero posteriormente los predictores fueron ampliados a un arreglo de variables ambientales, dando origen al término denominado correlación geográfica (Kellogg, 1959), correlación ambiental (Gessler *et al.*, 1995; McKenzie y Ryan, 1999), modelo de similitud de suelos (Zhu, 1997), técnicas *clorpt* (McBratney *et al.*, 2000) y modelo geoespacial *scorpan* (MacBratney *et al.*, 2003). Alternativamente, se han desarrollado métodos híbridos de la combinación de técnicas geoestadísticas y modelos de correlación ambiental, donde las observaciones o los residuales de la regresión son interpolados utilizando *co-kriging* o *kriging* de regresión (Heng *et al.*, 2004).

La mayoría de las investigaciones relacionadas con CDS provienen de Europa occidental, Estados Unidos, Canadá y Australia, donde existe una amplia disponibilidad de inventario de suelos. Sin embargo, hay una emergencia en cuanto al impulso de los levantamientos de suelo en muchos países de zonas tropicales, donde el inventario nacional no ha progresado tanto como en muchos de los países altamente industrializados (Mora *et al.*, 2008).

Al respecto, es prioritario, como ya se ha dicho, desarrollar y poner en práctica metodologías que permitan de manera rápida y efectiva la captura de información acerca de la variabilidad espacial de los suelos y reducir los requerimientos de muestreos intensivos y costosos. La interrogante clave al respecto es si las técnicas de CDS son apropiadas para levantamiento de suelos a nivel semidetallado, en áreas montañosas de difícil acceso, con limitada disponibilidad de datos sobre suelos, paisaje y ambiente, y una considerable variación de los suelos causada por la interacción de diferentes factores de formación de suelos, donde el terreno presenta evidencias de erosión causadas por movimientos en masa, como es el caso de la cuenca del río Caramacate.

### 2.7.2. Cartografía digital de suelos y densidad de muestreo

Un aspecto clave en cualquier sistema de inferencia de propiedades del suelo es la densidad de muestras de suelos, por su influencia sobre la calidad y costo de la información. Los levantamientos tradicionales consideran un número estipulado de muestras, usualmente una observación en campo por 1-4 cm<sup>2</sup> de mapa o 0,5 observaciones/cm<sup>2</sup> de mapa (Dent y Young, 1981; Gunn *et al.*, 1988; Rossiter, 2008), lo que significa que para una escala 1:50.000 se requiere una observación por cada 25-100 ha, o un promedio de una observación por cada 50 ha (1 observación/0,5 km<sup>2</sup>).

La literatura proporciona un gran número de ejemplos donde la CDS es representada como una técnica eficiente de levantamiento de información de suelos. Sin embargo, en muchos de esos casos las técnicas son aplicadas en áreas pequeñas (menores que 100 ha) con al menos 200 observaciones por km<sup>2</sup> (Bhatti *et al.*, 1991; McBratney *et al.*, 2000; Florinsky *et al.*, 2002; Kravchenko *et al.*, 2006), o para levantamientos de suelos semidetallados en áreas menores que 150 km<sup>2</sup>, en la que el número de observaciones por km<sup>2</sup> varía de 1 a 20 (Gessler *et al.*, 2000).

Las investigaciones realizadas por Zhu *et al.* (2008) y Yang *et al.* (2011) con la aplicación del algoritmo c-medias borroso y el enfoque SoLIM en una cuenca de 60 km<sup>2</sup> (Heilongjiang, China) consideraron un total de 68 observaciones. Para ello emplearon los valores de similitud cercanos a la unidad como áreas representativas de las relaciones entre el suelo y las condiciones ambientales. El mapa de suelos producido a escala 1:24.000 arrojó un 76% de confiabilidad.

Brungard (2009) utilizó el esquema de muestreo hipercubo latino condicionado (Minasny y McBratney, 2006), que es un método aleatorio estratificado basado en la distribución de covariables ambientales. El área de estudio fue de 275 km<sup>2</sup> de terreno montañoso, de alta complejidad geológica, topográfica y de vegetación. Se evaluaron 50 a 300 sitios de muestreo empleando cinco covariables ambientales (pendiente, índice de humedad, aspecto, índice de vegetación, y un mapa de cobertura vegetal combinado con geología). En este caso se concluyó que 300 observaciones eran suficientes para la obtención óptima de un mapa de clases de suelo, lo cual corresponde a una densidad de 1 observación por km<sup>2</sup>.

En otras situaciones, la CDS ha sido aplicada sencillamente para la predicción de propiedades individuales del suelo con apoyo de los atributos topográficos derivados de MDE y la aplicación de diferentes métodos estadísticos. McKenzie y Ryan (1999) y Ryan *et al.* (2000) utilizaron la correlación ambiental de los factores de formación del suelo y 165 muestras de suelo en dos áreas diferentes de 500 km<sup>2</sup> en Australia, y se encontró un R<sup>2</sup> de 0,54 y 0,39 para el %CO. El primer caso se refiere a un bosque en una zona montañosa,

mientras que en el segundo caso, el relieve y el uso del suelo son de menor complejidad. En ambas situaciones se utilizó un MDE con una resolución espacial de 25 m.

Por otro lado Ziadat (2005) realizó un ejercicio de CDS emprendido en Jordania en un área de 148 km<sup>2</sup>, con una densidad de 15 observaciones por km<sup>2</sup>. Se utilizaron los parámetros del relieve procedentes de un MDE de 20 m de resolución, con métodos de regresión lineal por pasos y un algoritmo de clasificación no supervisada, para predecir la profundidad del suelo, la capacidad de retención de agua, tipo y textura del suelo. Sin embargo, se encontró que la capacidad de los atributos del terreno para predecir los atributos del suelo fue poco significativa.

Mora *et al.* (2008) aplicaron las técnicas de CDS en una cuenca de alta complejidad topográfica, en la provincia de Kenia (13.500 km<sup>2</sup>), utilizando variables auxiliares y 95 muestras de suelo (1 observación/km<sup>2</sup>). El mapa de reconocimiento de suelos para la evaluación final de propiedades individuales (%A, %CO) fue elaborado utilizando *kriging* de regresión, mostrando una mejoría significativa al compararlo con mapas de un levantamiento del suelo existente.

De acuerdo con las investigaciones relacionadas con las técnicas de CDS, no existe una fórmula universal capaz de resolver las situaciones existentes a nivel local o regional, ni un procedimiento genérico para la obtención óptima de un número determinado de observaciones en un área de interés, con la finalidad de extraer la máxima información al mínimo costo. Sin embargo, conviene considerar el aporte de Scull *et al.* (2003) quienes señalan que en CDS se requieren menos observaciones que en el levantamiento convencional, porque no es necesaria la búsqueda de los límites de los suelos en el campo.

Hay que tener presente que un enfoque cuantitativo implica la utilización de la potencialidad de las bases de la CDS, haciendo uso eficiente de las variables auxiliares derivadas de MDE y sensores remotos. Por ello, la actividad de levantamiento en zonas montañosas debe ser complementada con un número determinado de observaciones del suelo; más aún en situaciones donde la intervención humana y los procesos erosivos causados por los movimientos en masa tienen una alta influencia en el desarrollo y el grado de evolución de los suelos. Hasta el presente, no existe un acuerdo, ni reglas sobre el número adecuado de observaciones necesarias en CDS.

### **2.7.3. Sistemas de inferencia espacial de propiedades del suelo**

La inferencia espacial de propiedades del suelo es un procedimiento o conjunto de procedimientos para la aplicación de un modelo de relaciones suelo-paisaje para predecir propiedades o clases de suelo, a partir de variables auxiliares y puntos de muestreo de suelos (Dobos *et al.*, 2006).



## AGRADECIMIENTO

### INSTITUCIONES:

Al Fondo Nacional de Ciencia, Tecnología e Innovación FONACIT, por el financiamiento parcial de los estudios de Doctorado en Ciencia del Suelo, a través del programa “Misión Ciencia”, y a la Gerencia de Financiamiento y Formación de Talentos, por su apoyo para la realización de pasantías de investigación en el exterior.

Al Laboratorio de Agrología del Instituto de Edafología de la Facultad de Agronomía de la Universidad Central de Venezuela, por el apoyo logístico.

Al Postgrado en Ciencia del Suelo de la Facultad de Agronomía de la Universidad Central de Venezuela, por la formación académica recibida.

Al Departamento de Medio Ambiente y Ciencia del Suelo de la Universidad de Lleida, Cataluña, España, por permitir la realización de pasantías en el área de Sistemas de Información Geográfica.

Al Centro de Investigación y Extensión en Suelos y Aguas de la Universidad Rómulo Gallegos (CIESA-UNERG), por el apoyo en cuanto a recursos, tanto humanos como materiales.

### INDIVIDUALIDADES:

Al Doctor Jesús Vilorio Rendón, quien en su condición de tutor, apoyó en forma incansable la formulación del proyecto de tesis, realizó innumerables revisiones y críticas constructivas para lograr la culminación de esta tesis, y contribuyó con su empeño al desarrollo metodológico y a la propuesta final lograda con la investigación. Padre científico de esta tesis doctoral por sus aportes en el ámbito de la cartografía digital de suelos y paisaje.

A la Doctora Clara Rondón de Rodríguez, quien con sus asesorías, guiaron con oportunas sugerencias, apropiadas observaciones y acertadas recomendaciones, el desarrollo y la fase final de esta investigación.

Mención especial al Doctor Francisco Ovalles (†), por sus recomendaciones en el desarrollo del proyecto de tesis, en los avances de la investigación y en la motivación acerca del interesante mundo de la variabilidad espacial de los suelos.

Al Doctor José Antonio Martínez, por sus incondicionales asesorías acerca de la aplicación de los sistemas de información geográfica en la cartografía digital de suelos y paisaje.

A la Doctora María Corina Pineda, compañera del postgrado en Ciencia del Suelo, pilar fundamental en la discusión de resultados obtenidos en las diferentes etapas de desarrollo de la investigación, y también amiga retadora de pensamientos, cuyas críticas constructivas siguen fortaleciendo e impulsando la línea de investigación relacionada con la cartografía digital de suelos y paisaje en el país..

El objetivo de un sistema de inferencia espacial (SIE) de propiedades del suelo es la producción de cartografía digital de propiedades y clases de suelo, asistida por computadores y SIG. De acuerdo con Dobos *et al.* (2006) estos sistemas se componen de tres elementos fundamentales: a) un modelo predefinido de formación del suelo, b) un conjunto de valores de propiedades del suelo obtenidas a través de mediciones directas, y c) un conjunto de variables auxiliares que tienen un impacto significativo sobre la formación del suelo y la distribución espacial de sus propiedades.

La base para la definición del modelo de formación edáfica es la conocida ecuación de factores de estado (Jenny, 1941), que identifica los cinco factores principales que contribuyen a la formación y desarrollo de los suelos (*clorpt*): el clima, los organismos biológicos, el relieve, el material parental y el tiempo.

Desde el punto de vista de un SIE, los cinco factores identificados por la ecuación de factores de estado se han extendido con la inclusión de la posición geográfica. Asimismo, algunas propiedades edafológicas son difíciles de medir, pero pueden estimarse con aceptable precisión mediante su correlación con otros parámetros del suelo medidos en la misma localización. Esta visión es la base del modelo *scorpan* formulado por McBratney *et al.* (2003), quienes identificaron siete factores para la predicción espacial de propiedades edafológicas.

$$Sac = f(s, c, o, r, p, a, n) \quad (2-1)$$

donde *Sac* es el atributo del suelo o clase de suelo estimada en una posición determinada, *s* representa otras propiedades del suelo medidas o estimadas en la misma localización, *c* son las condiciones climáticas, *o* expresa la acción de los organismos biológicos (vegetación, fauna o actividades humanas), *r* son los parámetros del relieve, *p* es el material parental, *a* es el tiempo y *n* es el espacio geográfico o la posición geográfica donde se estima el atributo del suelo.

Todos los factores mencionados son referidos como covariables ambientales, cuya información se puede derivar utilizando técnicas basadas en SIG y sensores remotos. De igual manera, McBratney *et al.*, (2003) señalan que las propiedades o las clases de suelo se pueden predecir de sus propiedades medidas o inferidas en combinación con los factores *clorpt*, los cuales se pueden derivar utilizando sensores remotos o el conocimiento de expertos. Bajo este enfoque el factor tiempo es difícil describirlo en cada lugar a través del paisaje; sin embargo cierta información de dicho factor es expresada implícitamente en otros factores de formación ambiental, tales como las posiciones topográficas.

En la ecuación (2-1), la *f* se refiere a la relación entre las condiciones de los suelos y el ambiente de formación de los mismos a través del tiempo, lo que en este caso es

equivalente a la relación entre las propiedades de los suelos y las covariables ambientales (modelo suelo-paisaje). Las relaciones suelo-paisaje  $f$  reflejan el concepto central de una categoría de suelo, así como también las transiciones hacia dicho concepto (Qi *et al.*, 2006).

Bajo el enfoque de CDS las relaciones suelo-paisaje son de dos tipos: a) en el dominio espacial o espacio geográfico, y b) en el dominio paramétrico o atributivo (Zhu *et al.*, 2008). Las relaciones en el dominio espacial exploran la autocorrelación espacial de los atributos del suelo, y la correlación cruzada entre atributos del suelo y covariables ambientales (*e.g.* aplicación de métodos geoestadísticos). En el dominio paramétrico se consideran las relaciones entre las condiciones del suelo y las covariables presentes en el modelo *scorpan* (*e.g.* métodos de extracción del conocimiento). Sin embargo, en las dos situaciones mencionadas los requerimientos de datos de suelos deben obtenerse de muestreos intensivos, ya que la mayoría de los métodos tienen aplicaciones restringidas en áreas con datos limitados (Zhu *et al.*, 2008).

#### **2.7.4. Modelos de predicción espacial**

Las técnicas de cartografía digital de suelos incluyen diferentes modelos para la predicción espacial. Estos modelos pueden ser organizados en tres categorías que agrupan diversos métodos de predicción espacial (Hengl, 2009):

1) Modelos mecánicos (deterministas): utilizan parámetros arbitrarios o empíricos. No producen una estimación del error del modelo y por lo general no se fundamentan en supuestos estrictos sobre la variabilidad de los datos interpolados. Las técnicas más comunes que pertenecen a este grupo son: interpolación lineal por triangulación, promedios ponderados basados en el inverso de la distancia o en el cuadrado de ésta, polinomios por cuadrados mínimos, regresión con las coordenadas, vecinos cercanos y curvas cúbicas (splines).

2) Modelos estadísticos lineales (de probabilidad): Los parámetros del modelo son estimados comúnmente en forma objetiva, aplicando la teoría de probabilidad. Las predicciones son acompañadas con una estimación del error de predicción. El éxito en la aplicación de estos modelos requiere que los datos de entrada cumplan con estrictos supuestos de estacionaridad estadística de la variable interpolada. Existen al menos cuatro grupos de modelos estadísticos lineales: kriging, correlación ambiental, modelos bayesianos y modelos híbridos (por ejemplo, kriging de regresión).

3) Sistemas expertos: Estos modelos pueden ser completamente subjetivos (por lo tanto irreproducibles) o completamente basados en los datos. Las predicciones suelen ser diferentes cada vez que se corre el modelo. Los sistemas expertos también se basan en gran medida en la teoría de probabilidad (en especial las estadísticas bayesianas), sin embargo,

son diferentes conceptualmente a las técnicas estadísticas lineales. Existen al menos tres grupos de sistemas expertos: Los basados principalmente en el conocimiento de expertos (por ejemplo, mapas hechos a mano); los algoritmos de aprendizaje automático (por ejemplo, basado en redes neuronales) y los sistemas expertos basados en lógica borrosa, los cuales introducen algunas reglas para tratar la incertidumbre.

Estos métodos son considerados como enfoques para la predicción espacial del suelo y son descritos en McBratney *et al.*, (2000), McBratney *et al.* (2003), Scull *et al.* (2003) y Hengl (2009), de cuyas investigaciones se han extraído los principales fundamentos expuestos a continuación.

#### **2.7.4.1. Modelos geoestadísticos**

Las técnicas geoestadísticas como *kriging* y sus derivados se basan en la teoría de las variables regionalizadas, e intentan utilizar de forma óptima los datos tomando en cuenta la variación de las propiedades del suelo en el espacio, a través del modelo del semivariograma. Su objetivo es predecir los valores en puntos no muestreados, basado en el modelo de un proceso estacionario estocástico, para lo cual es preciso que los valores puntuales estén autocorrelacionados espacialmente. La versión estándar es denominada *kriging* ordinario (OK) y las predicciones se basan en la ecuación:

$$Z(s) = \mu + \varepsilon'(s) \quad (2-2)$$

lo que implica que el valor de  $Z$  en el punto  $s$  es igual a una función constante estacionaria <sup>1</sup> (media global,  $\mu$ ) más el componente aleatorio ( $\varepsilon'$ ) de la variación espacialmente correlacionada.

#### **2.7.4.2. Correlación ambiental**

En este enfoque se incluyen las técnicas de predicción espacial basadas en la información ambiental relacionada con los factores de formación del suelo (denominadas *clorpt*, correlación ambiental o *scorpan*). En estos casos, los parámetros estadísticos del modelo se estiman comúnmente de manera objetiva siguiendo la teoría de probabilidades.

Cuando las variables auxiliares explicativas o covariables están correlacionadas significativamente con la variable a estimar (autocorrelación espacial) y los valores puntuales no están autocorrelacionados espacialmente, las predicciones pueden obtenerse centrándose en la parte determinística de la variación:

---

<sup>1</sup>Un supuesto fundamental en el análisis geoestadístico es que el fenómeno es estacionario, para lo cual el nivel promedio de la variable debe ser constante en todos los puntos del área de estudio. Es decir, el valor promedio de la variable no debería mostrar una tendencia definida en alguna dirección.

$$Z(s) = f\{q_k(s)\} + \varepsilon \quad (2-3)$$

donde  $q_k$  son los predictores auxiliares y  $\varepsilon$  es el error de la predicción. Los predictores ambientales explicativos utilizados para cartografiar las variables ambientales son los parámetros de superficie del terreno, las imágenes de sensores remotos, y mapas de geología, suelo y uso de la tierra. Debido a la disponibilidad de muchos predictores auxiliares a bajo costo o gratuitos, este enfoque de predicción espacial es cada vez más importante (Pebesma, 2006; Hengl *et al.*, 2007).

Dentro de los grupos de modelos estadísticos utilizados para hacer predicciones espaciales con el apoyo de los factores ambientales se encuentran los siguientes (Hengl, 2009):

- *Modelos basados en clasificación.* Estos se utilizan principalmente cuando se trata de la estimación de variables discretas (por ejemplo, cobertura del terreno o tipos de suelo). También hay una diferencia si las reglas de clasificación utilizadas para crear la salida son *booleanas* (discretas) o *borrosas* (continuas). Las salidas del proceso de ajuste del modelo son los límites de clase (centros de clase y las desviaciones estándar) de las reglas de clasificación. En este grupo juega un papel importante la teoría de los conjuntos borrosos, cuando la información es de naturaleza compleja, incierta e imprecisa, la cual es parte inherente de los sistemas naturales como el suelo.
- *Modelos basados en árboles* (árboles de clasificación o regresión). Estos son fáciles de interpretar cuando se utilizan como predictores la combinación de variables continuas y discretas. Estos son ajustados con una sucesiva división de un conjunto de datos en grupos cada vez más homogéneos. La salida del proceso de ajuste del modelo es un árbol de decisión, que luego se puede aplicar para hacer predicciones de cualquier valor de una propiedad individual o clase en toda una zona de interés.
- *Modelos de regresión.* El análisis de regresión emplea un conjunto de funciones llamadas modelos lineales generalizados (*MLG*), los cuales asumen una relación lineal entre las entradas y las salidas. La salida del proceso de ajuste del modelo es un conjunto de coeficientes de regresión. Los modelos de regresión también pueden ser utilizados para representar las relaciones no lineales con el uso de modelos aditivos generales (*MAG*). La relación entre los predictores y los estimados se puede resolver utilizando un ajuste de datos de un solo paso o mediante el uso de técnicas de ajuste iterativo de datos (e.g. redes neuronales artificiales). Estas últimas son importantes cuando las relaciones entre los atributos del suelo y el paisaje son altamente complejas, y no requiere de supuestos acerca de la distribución de los datos.

El enfoque de regresión más común para la predicción espacial es la regresión lineal múltiple. Las predicciones se obtienen por el promedio de los pesos de los predictores:

$$\hat{Z}(s_0) = \sum_{k=0}^p \hat{B}_k \cdot q_k(s_0) \quad (2-4)$$

donde  $q_k(s_0)$  son los valores de las variables explicativas en una localización determinada,  $p$  es el número de predictores o variables explicativas, y  $\hat{B}_k$  son los coeficientes de la regresión estimados por el método de los mínimos cuadrados ordinarios o lineales.

#### 2.7.4.3. Modelos híbridos (deterministas-estocásticos)

Los modelos de predicción espacial híbridos se componen de una combinación de las técnicas descritas anteriormente. Al respecto, un modelo híbrido geoestadístico emplea simultáneamente la correlación con predictores auxiliares y la autocorrelación espacial. Los enfoques híbridos han sido desarrollados con combinaciones de modelos de regresión y predictores geoestadísticos. Estos enfoques producen una mayor integración de los datos de campo y variables auxiliares, haciendo un mejor uso de las observaciones de campo de alto costo y mejoran la precisión de los productos cartográficos derivados de los datos de MDE y teledetección. Existen varios modelos geoestadísticos híbridos donde destacan: *kriging* universal, *co-kriging*, *kriging* con tendencia externa, *kriging* factorial y *kriging* de regresión (McBratney *et al.*, 2003).

Los modelos híbridos son más genéricos que las técnicas *kriging* puras o las técnicas basadas en regresión, y pueden utilizarse para representar en el espacio los cambios discretos y continuos de procesos deterministas y estocásticos. También se pueden combinar modelos de estimación determinista, estadísticos y sistemas expertos. Sin embargo, las técnicas más utilizadas que toman en cuenta la geoestadística y la incorporación de la información auxiliar del terreno son el *co-kriging* y el *kriging* de regresión (Kalivas *et al.* 2002; McBratney *et al.*, 2003).

##### 2.7.4.3.1. Kriging de regresión

El método de interpolación de la regresión (*regression kriging*, RK) se ha convertido en una herramienta importante de la geoestadística, debido a la amplia disponibilidad de covariables de alta resolución espacial con el avance de la teledetección y las tecnologías de posicionamiento satelital (Sun *et al.*, 2012).

Las técnicas RK utilizan un modelo determinista para estimar el valor de la variable de interés. De esta manera, se emplean los valores medidos en campo o laboratorio para desarrollar la calibración del modelo, posteriormente se analiza la correlación espacial de los residuos, y finalmente se combina el ajuste estadístico y el modelo determinista (Hengl *et al.*, 2007). Es decir, el RK utiliza la regresión de la información auxiliar y luego emplea OK para interpolar los residuales del modelo de regresión (Hengl *et al.*, 2007).

Generalmente, si la variable destino es estacionaria y el  $R^2$  de la regresión con las variables auxiliares no es significativa, RK no debe ser utilizado (Zhu y Lin, 2010). Los modelos estadísticos en muchos casos son capaces de explicar un porcentaje significativo de la variación espacial de las propiedades del suelo, aunque generalmente queda una gran cantidad de variación residual restante. Si esta última tiene dependencia espacial se pueden interpolar los residuales y describir parte de la variación restante utilizando *kriging* (Hengl *et al.*, 2004).

En la actualidad el modelo RK puede desempeñar un papel importante en la geoestadística, debido a que muchas covariables están disponibles con el avance en la teledetección y tecnologías de posicionamiento (Sun *et al.* 2012). Muchos estudios han demostrado que RK es fácil de usar y su precisión a menudo supera a la regresión lineal ordinaria, al *kriging* ordinario (Minasny y McBratney, 2007) y al *co-kriging* (Eldeiry y García, 2010). Al respecto, varios estudios han comparado los métodos RK y OK para distintas propiedades del suelo. Bishop y McBratney (2001) encontraron que RK fue es más preciso en la predicción de la CIC del suelo y Hengl *et al.* (2007) considera que RK es mejor cuando la información auxiliar puede explicar parte de la variación de la variable estimada. Sin embargo, Kravchenko y Robertson (2007) consideran que RK no es capaz de superar al OK en la predicción de propiedades del suelo que presentan una fuerte dependencia espacial y un  $R^2 < 0,6$ .

En otros casos, el uso de datos auxiliares del terreno ha mejorado la predicción de la profundidad efectiva del suelo, aunque la mejoría ha sido relativamente pequeña debido a la baja correlación de la variable estimada con los atributos del terreno utilizados (Penížek y Borůvka. 2006).

Zhu y Lin (2010) realizaron comparaciones entre RK y OK en dos paisajes contrastantes, determinando que RK fue más favorable en el paisaje boscoso de pendiente escarpada (7,9 ha; n=60), y OK fue más eficiente en paisajes agrícolas relativamente planos (19,5 ha; n=70). Además indicaron que con un tamaño de la muestra relativamente pequeño (n<100), las propiedades del suelo en los dos paisajes podrían ser razonablemente interpoladas, para lo cual hay que utilizar el método de interpolación más adecuado.

Sun *et al.* (2012) realizaron pruebas con el algoritmo RK, empleando 985 observaciones en la predicción del pH del suelo, el contenido de arcilla y el contenido de carbono en la parte baja del valle Hunter de Nueva Gales del Sur en Australia. Los resultados de validación mostraron que el método RK no siempre presenta las mejores predicciones, pero en casos específicos puede ser altamente preciso. Los investigadores concluyeron que el rendimiento de RK depende de las relaciones entre el suelo y los factores ambientales.

De acuerdo con los planteamientos anteriores, se ha corroborado que las relaciones entre el desarrollo del suelo y las condiciones de formación a menudo son muy complejas, y en la mayoría de los casos son desconocidas para un área específica, por lo que dichas relaciones no pueden ser caracterizadas completamente con modelos de regresión generales, debido posiblemente a su naturaleza de alta complejidad. En este sentido, tanto las redes neuronales como la teoría de conjuntos borrosos, son capaces de aproximar virtualmente las relaciones entre el desarrollo del suelo y las condiciones ambientales, para una determinada área de estudio.

Alternativamente, los enfoques híbridos como el *kriging* de regresión, el cual combina un modelo de regresión lineal con el *kriging* ordinario de los residuos de la regresión, constituye un método no estacionario de gran utilidad para mejorar las predicciones de propiedades de los suelos sobre áreas extensas. Estos aspectos teóricos se combinan con la realidad de campo de la cuenca del río Caramacate, en los capítulos que se desarrollan *a posteriori*.



## CAPÍTULO III

### CLASIFICACIÓN DIGITAL DEL TERRENO CON REDES NEUROBORROSAS EN LA CUENCA DEL RÍO CARAMACATE, ESTADO ARAGUA

#### 3.1. INTRODUCCIÓN

Se ha señalado previamente que el inventario del recurso suelo se fundamenta en el estudio de las relaciones entre el suelo y su entorno, para lo cual se desarrollan modelos de relaciones suelo-paisaje que relacionan el suelo con las formas de terreno. La identificación de unidades geomorfológicas en zonas montañosas por métodos tradicionales (manuales) es a menudo lenta, por la complejidad del paisaje y su difícil acceso. Por otra parte, diferentes expertos tienden a describir la misma zona de estudio de distintas maneras, lo cual depende, en gran parte de su percepción personal, por lo que la clasificación obtenida tiende a ser subjetiva y no reproducible.

Existen numerosas alternativas para generar unidades geomorfológicas en forma rápida, precisa y objetiva, a partir de clasificaciones digitales del paisaje tomando en cuenta las diversas variables auxiliares derivadas de un modelo digital de elevación (MDE) e imágenes de satélite. Estas nuevas opciones técnicas, producto del desarrollo de la geomática y del análisis geomorfométrico, complementadas con la geomorfología, contribuyen a la creación de una imagen entendible de las unidades de paisaje para apoyar la toma de decisiones y la planificación del ordenamiento territorial. Entre las opciones disponibles se destacan las técnicas de análisis espacial basadas en inteligencia artificial, tales como la lógica borrosa (LB) y las redes neuronales artificiales (RNA), para producir modelos digitales de clasificación del terreno.

En este estudio se utilizó la técnica integral de clasificación no supervisada, denominada red de agrupamiento borroso de Kohonen o *Fuzzy Kohonen Clustering Network* (FKCN) (Lin y Lee, 1996). Esta técnica se basa en la combinación del modelo de mapas autoorganizados o *Self Organizing Map* (SOM) (Kohonen, 1982) con lógica borrosa (Bezdek *et al.*, 1992). Este algoritmo neuro-borroso ha sido implementado en un sistema de clasificación del terreno (Viloria, 2007). Como datos de entrada se utilizaron atributos derivados de un MDE e imágenes satelitales de un sector de la cuenca del río Caramacate.

El significado geomorfológico de las clases digitales de superficie de terreno (Clases DST) obtenidas, se interpretó a partir de la distribución espacial de las clases y de las descripciones de los centros de clases y las matrices de membresías, conjuntamente con la información sobre geología, clima y vegetación del área de estudio. Este método de clasificación no supervisada constituye una importante alternativa de apoyo al experto para la obtención de una visión más amplia e integral de la estructura del paisaje, a la vez que

servió como base para la posterior planificación y aplicación del método de muestreo de suelos para la evaluación de las relaciones suelo-paisaje en la cuenca del río Caramacate.

Para cumplir con el primer objetivo de esta investigación, este capítulo describe la generación de un modelo digital de superficie de terreno en un sector de la cuenca del río Caramacate, a partir de un MDE e imágenes satelitales, por medio de algoritmos basados en lógica borrosa y redes neuronales artificiales.

Para el alcance del objetivo planteado, se realizaron varias actividades que incluyeron la evaluación de la variación de las clases digitales a diferentes resoluciones espaciales, la selección del número óptimo de clases neuro-borrosas y el modelo más representativo de las clases DST. También se evaluó la influencia de una variable discreta (litología) y el número de variables ambientales cuantitativas en la definición de clases DST, y finalmente se correlacionaron dichas clases con un sistema tradicional de clasificación de paisajes.

## **3.2. MATERIALES Y MÉTODOS**

### **3.2.1. Área de estudio**

#### ***3.2.1.1. Localización geográfica***

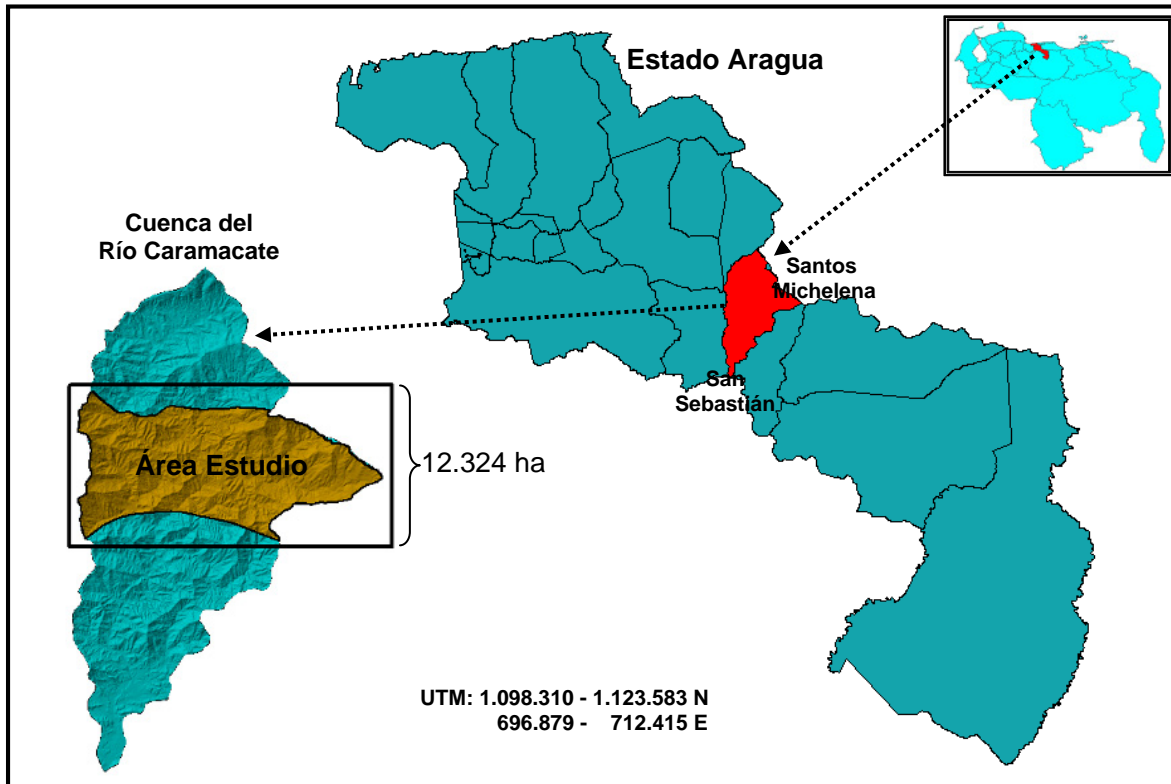
La cuenca del río Caramacate está ubicada entre los municipios Santos Michelena y San Sebastián de los Reyes del estado Aragua, geográficamente localizada entre las coordenadas UTM 696.879-712.415 Norte, y 1.098.310-1.123.583 Este, Huso 19 (Figura 3-1). Tiene una extensión aproximada de 17.888 ha y representa el 8,5% de la cuenca alta del río Guárico, de la cual es tributaria (Jácome, 2002; Ochoa, 2003). Esta cuenca, aunque abastece el 60% del agua que consume el área metropolitana de Caracas, está sometida a intensos procesos de degradación ambiental y carece de suficiente información para apoyar la implantación de planes de manejo.

#### ***3.2.1.2. Características físico-naturales***

En la cuenca del río Caramacate se presentan dos zonas bioclimáticas (Spósito y Elizalde, 2004): Bosque Seco Tropical en las zonas más bajas y Bosque Húmedo Premontano en las más altas. La cuenca forma parte de un área orogénica, metamorfizada, plegada y fallada, que constituye un sistema montañoso de orientación Este-Oeste de la región fisiográfica Cordillera de la Costa Central, provincia fisiográfica Serranía del Interior Central y subprovincia fisiográfica Complejo Montañoso de la Serranía del Interior Central (Ospina y Elizalde, 2004).

Desde el punto de vista geológico, la cuenca está ubicada sobre dos grandes grupos geológicos, donde destaca el grupo Villa de Cura (Formaciones Santa Isabel, El Carmen, El Caño y El Chino) perteneciente al Mesozoico metamorfizado, aparentemente desde el

Cretáceo y Paleoceno, dispuesta de Sur a Norte en la cuenca. El otro grupo geológico aflora en la depresión colinada de San Sebastián de los Reyes, correspondiente al piedemonte, perteneciente al Cretáceo Superior metamorfozido y Terciario sedimentario (Shagam, 1960).



**Figura 3-1.** Ubicación relativa del área de estudio dentro de la cuenca del río Caramacate, estado Aragua.

El paisaje está constituido principalmente por laderas de montaña, escarpadas, con pendientes medias del orden del 45% y altitudes que oscilan entre 334 y 1405 msnm (Ochoa *et al.*, 2004). Presenta un relieve accidentado, donde más del 70% de su superficie tiene una pendiente superior al 15%, y la tendencia general en la evolución de los cambios de cobertura de la tierra, refleja claramente el incremento de herbazales y la reducción de la superficie cubierta por bosques (Jácome, 2002).

El clima manifiesta la influencia de la topografía y muestra un gradiente de temperatura decreciente de Sur a Norte de 26 a 22 °C, en la medida que aumenta la altura del terreno. De igual manera, la precipitación media anual aumenta en la misma dirección de 1077 a 1400 mm, aunque no de manera homogénea, ya que los valores mayores se encuentran hacia la parte centro-oriental de la cuenca (Parra y Puche, 2004; Pineda, 2012).

El principal uso de la tierra en la cuenca del río Caramacate es la ganadería extensiva,

Al Doctor José Guerrero, por su contribución en el apoyo acerca de la aplicación de las metodologías para el análisis químico de los suelos.

Al Ing. Alvaro Vilorio, cuyo interés por las técnicas de Inteligencia Artificial dejó el camino abierto para la profundización de actividades científicas en el campo de la Ciencia del Suelo.

Al equipo CIESA: Ing. Cándido Sumoza, Ing. Willian Tovar, Ing. Elsa Gutiérrez, Ing. María Tovar, Ing. Fernando Pulido e Ing. Carlos Vásquez, por su colaboración y apoyo en las actividades de campo y en la realización de los análisis químicos y físicos de los suelos.

A los compañeros del Postgrado en Ciencia del Suelo: Bestalia Flores, Efraín Visconti, Juan Carlos Rey, Víctor Sevilla, Luis De Armas, Jairo Ferrer y Gabriel Machado, quienes con su intercambio de ideas contribuyeron como fuente de apoyo en las diferentes fases de la investigación.

A la señora Carmen García (Lleida, España) por su gran apoyo durante las pasantías en esas tierras lejanas.

A la compañera Zaida Cortez, asistente de laboratorio del CIESA UNERG, por su participación en el acondicionamiento y preparación de las muestras de suelo y en la determinación de las diferentes fracciones de agregados.

A todos..., agradecido por siempre.

ocupando el 54,5% de la superficie, donde el manejo inadecuado como el sobrepastoreo y las quemadas anuales en zonas de altas pendientes, ha ocasionado severos daños ambientales que se manifiestan en forma de erosión hídrica laminar y concentrada de variada intensidad (Ochoa, 2003).

### **3.2.1.3. Área piloto**

La investigación se realizó en un área muestra de la región central de la cuenca del río Caramacate, representativa de una superficie de 6.760 ha; sin embargo, para el estudio de los factores relacionados con el suelo y el paisaje, se consideró una sección geográfica rectangular de 12.324 ha centrada en la zona de interés, denominada “área de referencia espacial”, a fin de facilitar el procesamiento digital y la aplicación de algoritmos.

Este sector se localiza en la zona de vida Bosque Seco Tropical, en la que predomina el tipo de cobertura vegetal de herbazal con bosques de galería a lo largo de los cursos de agua, y en una zona de vida de Bosque Húmedo Premontano, dominado por un bosque siempreverde (sector El Altar) rodeado de bosque semidecíduo en áreas de mayor altitud (Pineda, 2012), cuyas características topográficas y microclimáticas le confieren condiciones diferentes al resto de la zona.

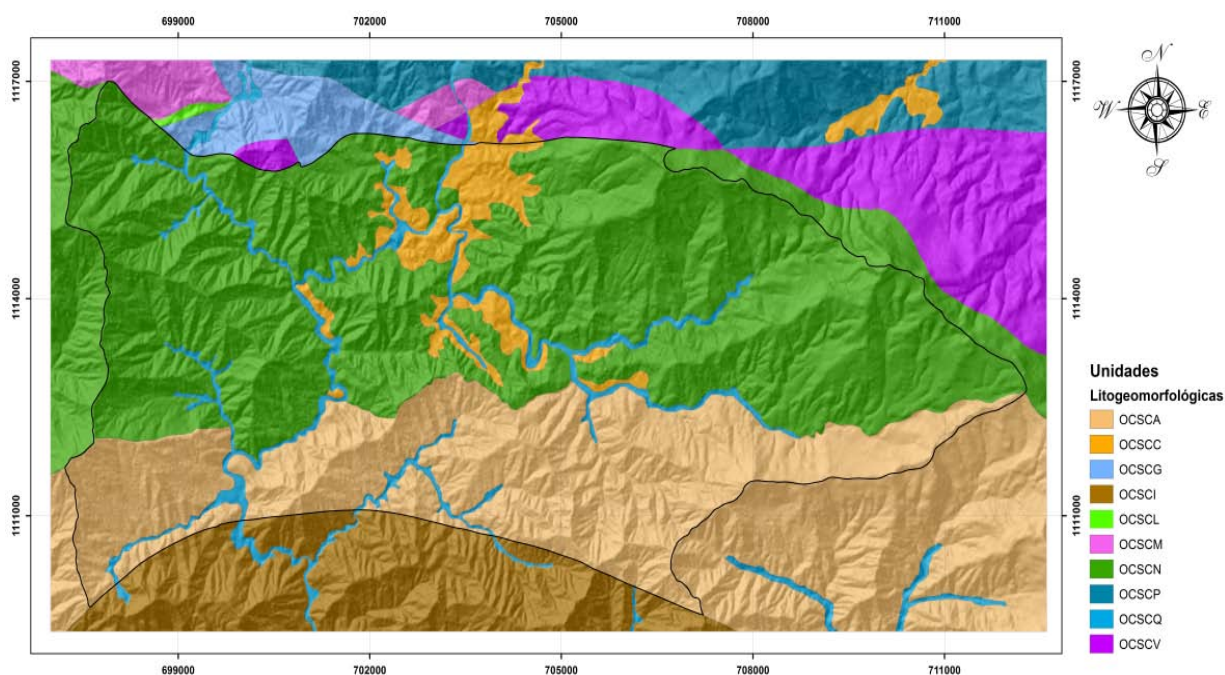
De acuerdo con la aplicación del sistema de clasificación de paisajes propuesto por Elizalde (2012), en el área de referencia espacial se obtuvo un total de 10 unidades litogeomorfológicas, donde afloran materiales dominantes de las formaciones geológicas El Caño-El Chino y El Carmen, constituidas por asociaciones meta-volcano-sedimentarias del Grupo Villa de Cura - VC (Urbani y Rodríguez, 2004). A estos materiales geológicos se le suman los materiales superficiales, provenientes de los sedimentos aluviales y de mantos coluviales o coluvio-aluviales cuaternarios delimitados a escala 1:25.000 (Pineda, 2008; Pineda *et al.*, 2011). El Cuadro 3-1 simplifica la clasificación jerárquica realizada hasta el nivel 5, y en la Figura 3-2 se observa la distribución geográfica de las unidades litogromorfológicas presentes en la zona de estudio.

En el área domina el paisaje de montaña, con un relieve muy accidentado con variaciones de altitud comprendidas entre 433 y 1420 msnm y pendientes promedios de 30 a 50%. Los suelos en su mayoría son entisoles, inceptisoles y alfisoles, con régimen de humedad ústico. Algunos suelos son muy pedregosos ó excesivamente pedregosos en superficie, pero en general, son bien drenados, de permeabilidad y drenabilidad media. La variabilidad de suelos en las laderas ha sido incrementada por la incidencia y superposición de movimientos en masa, que determinan la coexistencia de zonas con cicatrices de diferentes edades y sectores donde existe acumulación de los sedimentos superficiales (Prada *et al.*, 2004; Pineda *et al.*, 2011).

**Cuadro 3-1.** Principales unidades litogemorfológicas presentes en la cuenca del río Caramacate.

Nivel	Categoría	Designación	Nombre
1	Megarregión	O	Orogénica
2	Región	OC	Cordillera de la Costa Central
3	Provincia	OCS	Serranía del Interior
4	Subprovincia	OCSC	Complejo Montañoso de la Serranía del Interior
5	Unidad Litogemorfológica	OCSCA	Metalavas de El Carmen (VCca)
		OCSCC	Coluviones superficiales: El Caño-El Chino
		OCSCG	Rocas volcano-sedimentarias del río Guaire (LHrg)
		OCSCI	Granofels de Santa Isabel (VCSI)
		OC_SCL	Peridotitas de Lomas de Níquel (LHcp)
		OCSCM	Gabros de Mesia (LHcg)
		OCSCN	Metatobas de El Caño-El Chino (VCñc)
		OCSCP	Filitas de Paracotos (LHP)
		OCSCQ	Sedimentos del Cuaternario (Qal)
		OCSCV	Metalavas de Tiara (LHT)

**Fuente:** Atlas geológico de la Cordillera de la Costa, Venezuela. Mapas a escala 1:25.000. Edic. Fundación Geos, UCV. Caracas. Urbani y Rodríguez (2004); Pineda (2008).



**Figura 3-2.** Unidades litogemorfológicas presentes en el área de referencia espacial de la cuenca del río Caramacate.

### 3.2.2. Procesamiento de la información cartográfica

Para la subdivisión del área en unidades litológicas se utilizaron mapas geológicos digitales a escala 1:25.000 (6745-I-NO, 6746-II-NE, 6746-II-NO, 6746-II-SE, 6746-II-SO) (Urbani y Rodríguez, 2004), para una superficie de 12.324 ha. También se utilizaron hojas cartográficas y ortofotomapas a la misma escala, corregidas geoméricamente en el sistema de proyección UTM (*Datum* Sirgas-Regven, Huso 19). En la Figura 3-3 se presenta la

cartografía básica, una imagen satelital corregida geoméricamente y las capas vectoriales empleadas en la generación de MDE.

Las curvas de nivel se vectorizaron con separaciones de 20 m de altitud, al igual que la red hidrográfica, el límite de la cuenca y los puntos de control, y con estas coberturas vectoriales se realizó la interpolación de un MDE base de 20 m de resolución espacial (modelo raster) por medio del comando *Topogrid* de ArcGis 9.2® (ESRI, 2006), el cual utiliza el método *Anudem* (Hutchinson, 1989; Hutchinson *et al.*, 2011). (Figura 3-4). Posteriormente se realizó la interpolación de otros MDE con resoluciones espaciales de 10, 15 y 30 m, en el marco referencial de la escala de trabajo (1:50.000).

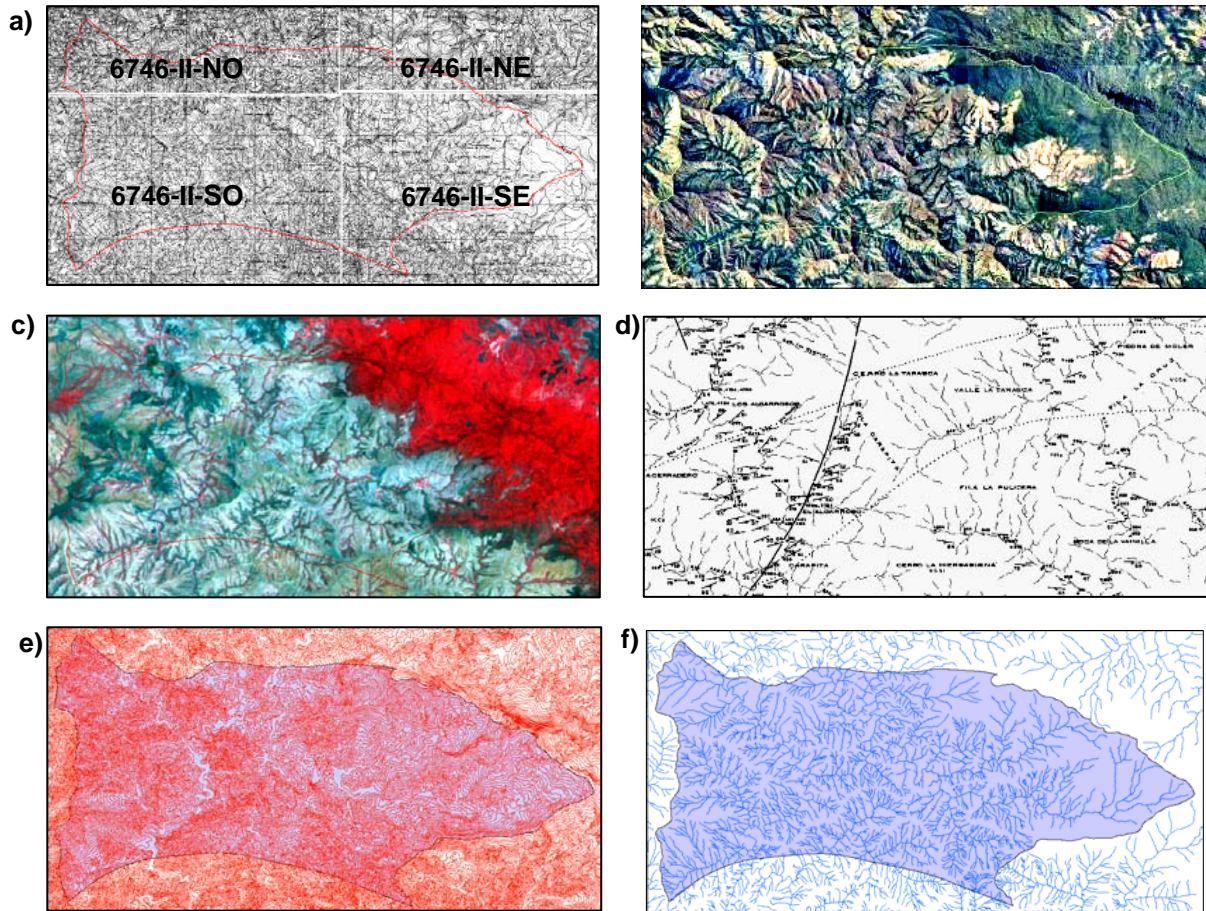
### 3.2.3. Parámetros descriptivos de la superficie de terreno

Los MDE de diferente resolución espacial se utilizaron para derivar mapas de diversos parámetros topográficos empleados como covariables ambientales. Estos incluyeron: altitud (Alt), grado de pendiente (Pend), orientación de la pendiente (Asp), posición relativa (PR), curvatura vertical (Perfil C), curvatura horizontal (Plano C), área de captación (Area C), índice topográfico de humedad (ITH), índice de transporte de sedimentos (ITS) y el índice del potencial de escorrentía (IPE).

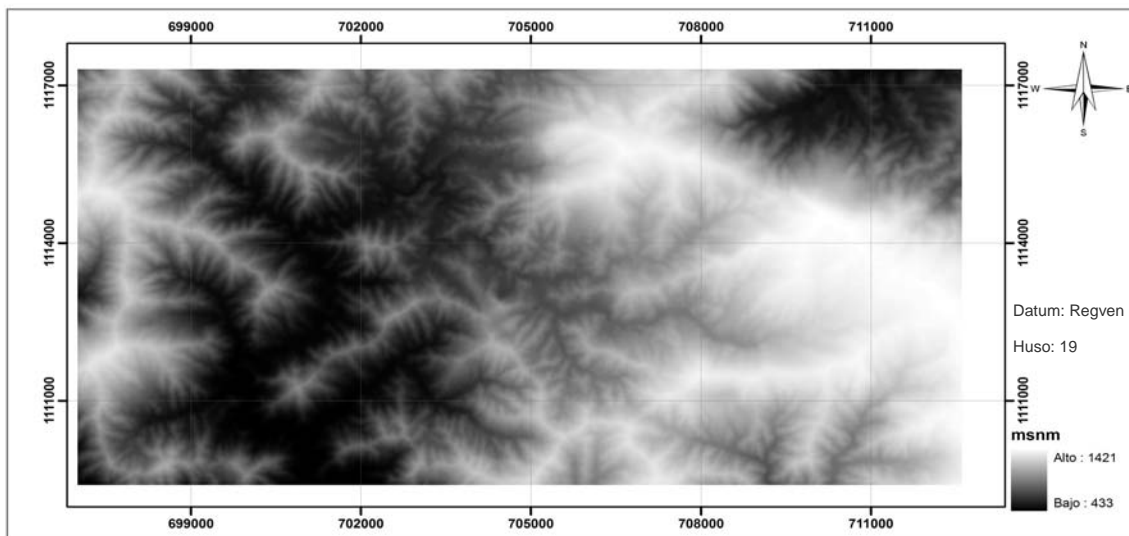
Los parámetros topográficos fueron calculados con apoyo del programa SAGA GIS® (SAGA, 2011). El grado de la pendiente y la orientación (aspecto) fueron determinados empleando la ecuación de Zevenbergen y Thorne (1987), quienes utilizan un algoritmo de diferencia finita de segundo orden, ajustado a los cuatro (4) vecinos más cercanos en la ventana de píxeles.

Los parámetros de la curvatura (plano y perfil) también se basaron en el algoritmo de Zevenbergen y Thorne (1987) sobre ventanas de 3x3 celdas. El área de captación se obtuvo mediante el algoritmo de Tarboton *et al.* (1991), el cual utiliza un polinomio cuadrático con 9 términos. El índice de humedad (ITH) fue determinado según la ecuación:  $ITH = \ln(\text{Area\_C} / \text{Pend}^\circ)$  (Gessler *et al.*, 1995; Wilson y Gallant, 2000). El índice ITS se determinó a partir de la ecuación  $ITS = (n+1) (\text{Area\_C} / 22,13)^n (\text{sen } \text{Pend}^\circ / 0,0896)^m$ , donde  $n=0,4$ ;  $m=1,3$ . El IPE se obtuvo según la ecuación de Moore *et al.* (1993) ( $IPE = \ln(\text{Area C} \times \text{Pend}^\circ)$ ).

La posición relativa (PR) se determinó mediante la fórmula:  $PR = ([MDE] - [MDE_{\text{Min}}] / ([MDE_{\text{Max}}] - [MDE_{\text{Min}}]))$ , donde *Max* y *Min* corresponden a las estadísticas máximas y mínimas derivadas del MDE original, en una ventana de 3x3 celdas. En el Cuadro 3-2 se presenta un resumen conceptual del significado físico de cada uno de los atributos topográficos mencionados, con sus respectivas fuentes bibliográficas.



**Figura 3-3.** Hojas cartográficas del área de referencia espacial-ARE (a), ortofotomapas (b), imagen satelital SPOT (c), mapas geológicos (d) y las coberturas de curvas de nivel (e) e hidrografía (f) del área de estudio.

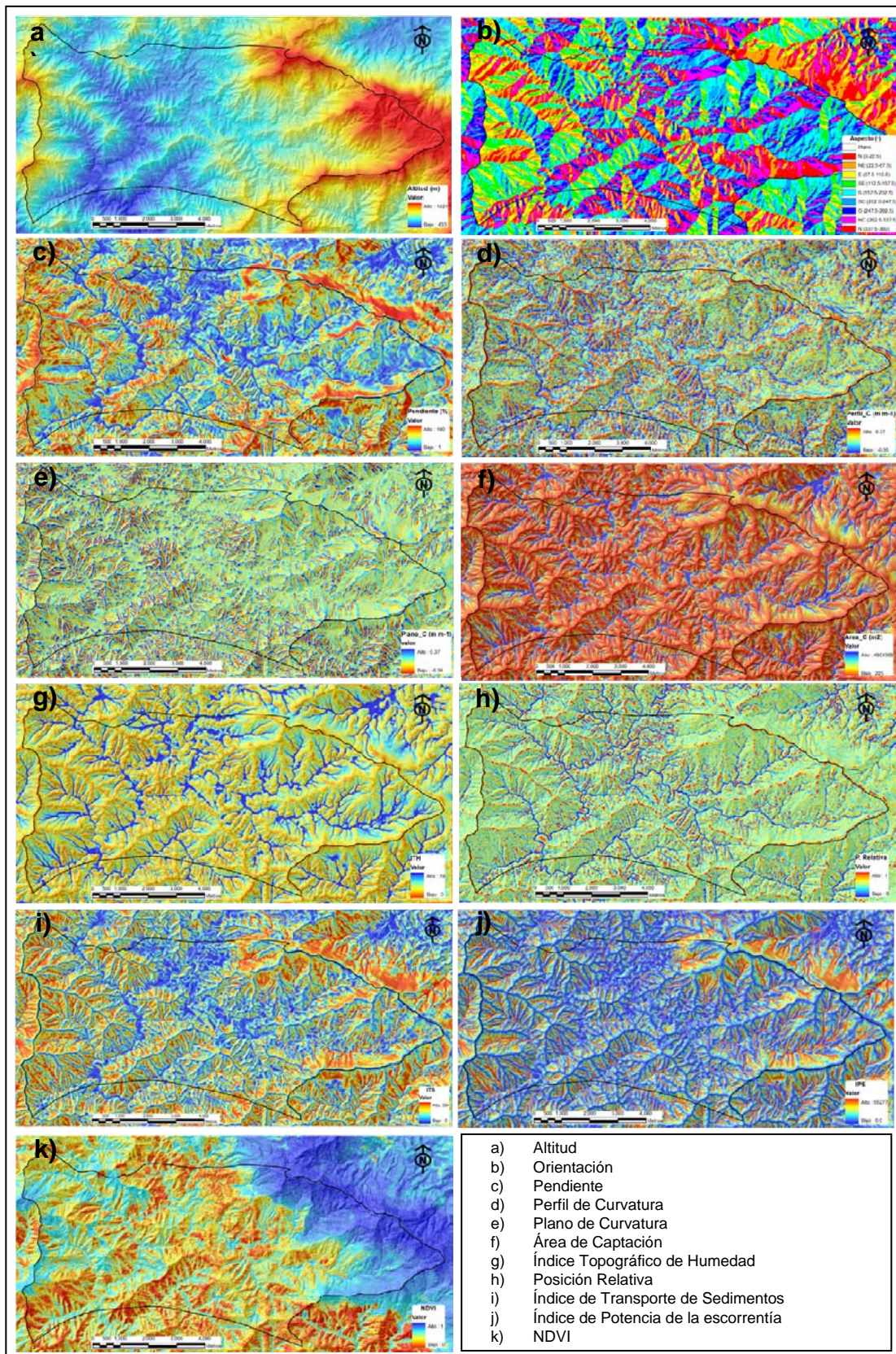


**Figura 3-4.** Modelo digital de elevación del área de referencia espacial de la cuenca del río Caramacate.



**Cuadro 3-2.** Atributos topográficos utilizados en la clasificación digital del terreno de la cuenca del río Caramacate.

<b>Atributo</b>	<b>Definición</b>	<b>Referencia</b>
Altitud (Alt)	Se conoce también como gradiente topográfico, y está relacionado con el valor representado por cada píxel en un modelo digital de elevación (MDE). Constituye la primera derivada del modelo de elevación respecto a los ejes de coordenadas.	Burrough y McDonnell (1998) Zevenbergen y Thorne (1987)
Pendiente (Pend)	Es la tasa de cambio de la altitud en dirección descendente, y está referida a la inclinación que presenta la superficie del terreno con respecto a un plano imaginario horizontal. Se corresponde con la derivada de primer orden de la altitud.	Moore <i>et al.</i> , (1993) Burrough y McDonnell (1998)
Orientación (Asp)	También llamado aspecto o dirección de la pendiente. Es el ángulo medido en la dirección horaria desde el norte (azimut) donde se produce la máxima pendiente.	Moore <i>et al.</i> (1993)
Plano de curvatura (Plano C)	Curvatura en dirección transversal a la pendiente o cambio de la pendiente con la distancia en dirección transversal (curvatura horizontal). Es una medida de la divergencia o convergencia topográfica, y de la concentración del agua en el paisaje. Adopta valores positivos en las formas convexas y negativos en las cóncavas	Zevenbergen y Thorne (1987)
Perfil de curvatura (Perfil C)	Curvatura en dirección de la máxima pendiente o tasa de cambio de la pendiente en dirección vertical. Es un atributo que determina el camino del agua y la deposición de materiales, y es utilizada para caracterizar los cambios en la velocidad de la escorrentía y en el transporte de sedimentos.	Moore <i>et al.</i> (1993)
Área de captación (Área C)	Es el número de celdas ladera arriba que drenan a través de cada celda del modelo. Es proporcional a la escorrentía superficial o subsuperficial poco profunda en un punto dado del paisaje e integra los efectos del área de contribución ladera arriba y la convergencia y divergencia de la cuenca para la escorrentía.	Tarboton <i>et al.</i> 1991 Moore <i>et al.</i> (1993) Wilson y Gallant (2000)
Índice topográfico de humedad (ITH)	Es un predictor de zonas de saturación del suelo y relaciona el área de captación específica con el grado de la pendiente.	Gessler <i>et al.</i> (1995); Wilson y Gallant (2000)
Índice de potencia de la escorrentía (IPE)	Es una medida de la erosividad de la escorrentía. Se expresa matemáticamente como el logaritmo natural del producto del área de captación específica y el grado de la pendiente, y refleja el potencial erosivo del flujo superficial	Moore <i>et al.</i> (1993) Hengl y Rossiter (2003)
Índice de transporte de sedimentos (ITS)	Representa los efectos del relieve sobre los procesos de transporte de los sedimentos, es una medida de la capacidad de transporte de sedimentos del flujo de agua en un punto dado del paisaje. Predice localizaciones de áreas de erosión y deposición neta.	Wilson y Gallant, (2000) Hengl y Rossiter, (2003)
Posición relativa (PR)	Se refiere a la relación entre la posición de un punto elevado con respecto a otro más bajo en una ventana de 3x3 celdas de un MDE. Las posiciones relativas altas se acercan a la unidad (1) y las posiciones relativas bajas tienden a cero (0).	Verbrugge (2006)

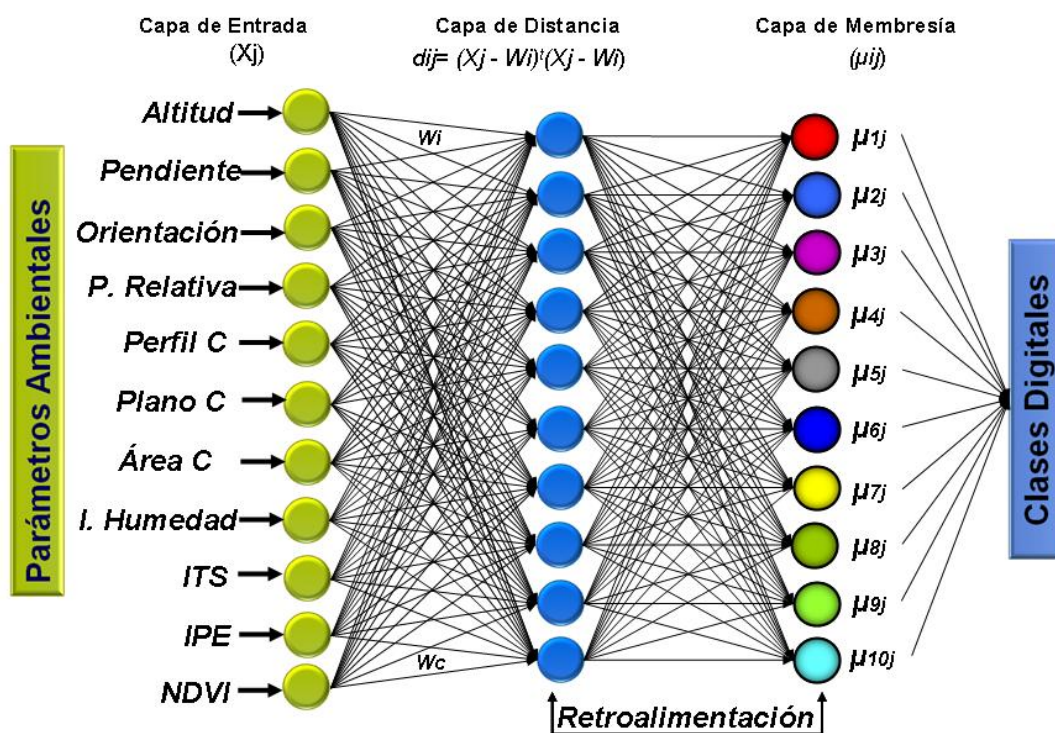


**Figura 3-5.** Variables auxiliares derivadas del MDE (atributos e índices topográficos) y de la imagen satelital multispectral en el área de la cuenca del río Caramacate estudiada.

Para la obtención del índice de vegetación de diferencia normalizada (*Normalized Difference Vegetation Index*, NDVI) (Rouse *et al.*, 1974), se empleó una imagen de satélite multispectral *SPOT 5* multispectral (2003) de 20 m de resolución espacial, tomando en cuenta la banda roja visible (*Red*) y la banda infrarroja cercana (*NIR*) del espectro electromagnético ( $NDVI = \frac{NIR - Red}{NIR + Red}$ ), para lo cual se utilizó el programa ERDAS IMAGINE® (2005). Las variables descriptivas de la superficie del terreno se representan en la Figura 3-5.

### 3.2.4. Sistema clasificador del terreno: Red de agrupamiento borroso de Kohonen (FKCN)

La red neuronal FKCN utilizada en el análisis morfométrico consta de tres capas. La capa de entrada contiene los valores normalizados de las 11 variables ambientales procedentes del MDE y de la imagen de satélite; la capa de distancias tiene una cantidad de neuronas equivalente al número preestablecido de clases DST, y la tercera capa calcula la función de pertenencia de cada celda para cada una de las clases de terreno, basada en las distancias computarizadas en la capa anterior y los valores preestablecidos del coeficiente de borrosidad ( $\phi$ ). La Figura 3-6 esquematiza un modelo de la arquitectura de la red FKCN con once (11) neuronas en la capa de entrada, diez (10) neuronas en la capa de distancia y la capa de salida con las diez (10) clases digitales de superficie de terreno.



**Figura 3-6.** Estructura de la RNA utilizada en el análisis de atributos morfométricos en el área de la cuenca del río Caramacate estudiada.

Los atributos topográficos y la variable NDVI fueron agrupados en una matriz de datos bajo formato *ASCII* (*American Standard Code for Information Interchange*, Código Americano Normalizado para el Intercambio de Información), para la aplicación del sistema de clasificación FKCN en cada uno de los MDE con sus respectivas resoluciones. Se evaluó el agrupamiento de píxeles con diversos números de clases (6 a 12) y diferentes coeficientes de borrosidad ( $\phi = 1,1$  a  $1,6$ ), ya que la magnitud de este exponente determina el grado de borrosidad del modelo.

En la RNA cada neurona es descrita por un vector  $n$ -dimensional de pesos, donde  $n$  es igual al número de variables en los datos de entrada. Los vectores de pesos se inicializan con valores aleatorios y mediante un proceso iterativo, se lleva a cabo la adaptación de la red de manera sistemática y secuencial. Cuando un vector de entrada se presenta a la red, esta capa calcula la distancia euclidiana entre el vector de entrada y cada uno de los vectores de pesos. La neurona con la distancia mínima a partir del vector de entrada es elegida como el nodo ganador y los vectores de pesos se actualizan para acercarlos al vector de entrada, proceso que se repite de manera iterativa (Kohonen 1991; Ehsani y Quiel, 2007). El aprendizaje de la red se detiene cuando se alcanza el número máximo de iteraciones o cuando la diferencia entre una iteración y la anterior es menor o igual al coeficiente de convergencia del modelo. Ambos parámetros son asignados por el usuario.

Para la obtención de las clases DST se realizó el siguiente procedimiento: a) *ingreso de datos de entrada* en formato *ASCII* para generar una matriz de variables; b) *entrenamiento de la red neuronal*, especificando los parámetros de aprendizaje (*número de clases: 6-12; exponente de borrosidad: 1,1-1,6; error de convergencia: 0,0001-0,001; número de iteraciones: 20-50*); c) obtención de mapas de valores de similitud de las clases de terreno; y d) generación del modelo final.

La red entrenada genera cinco archivos de salida: a) *Topology.txt*, el cual guarda la estructura de la red entrenada; b) *Statistics.txt*, contiene las estadísticas del modelo generado, tales como el índice de rendimiento borroso (*Fuzziness Performance Index*, FPI), el error final de convergencia y los valores promedios, máximos y mínimos de cada variable de entrada con el centro de cada clase, entre otros; c) *Convergence.txt*, donde se almacena el error de convergencia por cada época de generación del modelo; d) *Membership map.asc*, el cual contiene los valores de membresía (grados de pertenencia o valores de similitud) de cada unidad de resolución espacial (celda) a las clases DST, con valores graduales comprendidos entre 0,1 y 1,0; y e) *Fkcn\_map.asc*, este archivo contiene el mapa final en formato *ASCII*, de acuerdo con la topología generada. Consiste en la integración de los archivos de los valores de similitud representativos de cada clase ( $> 0,5$ ). El mapa se puede visualizar en el sistema FKCN o en cualquier programa SIG, previa

conversión del formato (e.g. *ASCII* a raster) (Viloria, 2007).

Se debe destacar que las variables introducidas al sistema FKCN cuantifican diferentes propiedades físicas del terreno, y por lo tanto, utilizan diferentes unidades de representación. Esto hace que las variables tengan varianzas muy dispares entre sí, hecho que no responde a la variabilidad interna real, sino a la distinta magnitud que toman las unidades con la que se representa cada parámetro morfométrico. Esto hace necesario escalar o estandarizar los valores de los atributos, por lo cual el FKCN utiliza un proceso de estandarización de las variables originales a variables normalizadas con media 0 y varianza 1.

### 3.2.5. Número óptimo de clases digitales de terreno

Para la obtención del mejor modelo generado por el sistema FKCN se utilizó un enfoque inductivo, basado en el procedimiento de Odeh *et al.* (1992), el cual relaciona el índice de borrosidad alcanzado (FPI, *Fuzziness Performance Index*) con el número de clases. Estos parámetros se obtienen utilizando el algoritmo *Fuzzy C-means* (FCM) (Bezdek, 1981) del programa *Fuzme 3.5* (Minasny y McBratney, 2002).

En el agrupamiento, de acuerdo con los autores mencionados, el algoritmo FCM utiliza una función objetiva (J) para minimizar los errores de la suma de cuadrados dentro de las clases. Esta función se define mediante la siguiente fórmula:

$$J = \sum_{j=1}^n \sum_{k=1}^c (\mu_{ik})^{\phi} D^2(x_i c_k) \quad (3-1)$$

donde,  $n$  es el número de datos;  $c$  el número de clases;  $c_k$  es el vector que representa el centroide de la clase  $k$ ;  $x_i$  es el vector que representa los datos individuales  $i$ ;  $D^2(x_i c_k)$  es la distancia al cuadrado entre  $x_i c_k$  de acuerdo con la distancia seleccionada (e.g. Mahalanobis, Euclidiana, Diagonal) y  $\phi$  es el exponente que determina el grado de borrosidad. El término  $\mu_{ik}$  se refiere a la función de membresía para  $n$  objetos  $i$  agrupados en  $c$  clases. La derivada de los valores de la función objetiva (OFV) con respecto a los coeficientes de borrosidad ( $-dJ/d\phi$ ) para distintos números de clases, genera el mejor valor de  $\phi$ .

La selección del número óptimo de clases en FCM se realizó por repetición de la clasificación para un rango de número de clases (e.g. entre 6 y 12). En cada agrupamiento obtenido se generan dos parámetros de clasificación tales como el FPI y la entropía de partición modificada (MPE, *modified partition entropy*). El FPI estima el grado de borrosidad generada por cada número específico de clases. Matemáticamente, se define como:

$$FPI = 1 - [(cF - 1)/(c - 1)] \quad (3-2)$$

Donde  $c$  es el número de clases y  $F$  es el coeficiente de partición calculado como:

$$F = (1/n) \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^2 \quad (3-3)$$

$F$  es conceptualmente comparable a la relación del conjunto de varianzas dentro de las clases y la varianza entre clases y es cercana a 1 para los agrupamientos más significativos. Cuando  $F$  es igual a 1, el FPI será igual a 0 y las clases serán discretas (no borrosas) ya que las subestructuras estarán bien diferenciadas entre sí. Si el FPI es igual a 1 ( $F=0$ ), la borrosidad en el conjunto de datos será máxima. De esta manera, la minimización de FPI indica un número óptimo de clases borrosas que reflejan adecuadamente el comportamiento del conjunto de datos.

La MPE determina el grado de desorden creado por cada número específico de clases. Se calcula mediante la ecuación:

$$MPE = H / (\log c) \quad (3-4)$$

Donde  $H$  es una función de entropía, matemáticamente similar a la entropía termodinámica, cuyos valores se aproximan a cero cuando las agrupaciones son altamente significativas. Se define de la siguiente manera:

$$H = (1/n) \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik}) \log(\mu_{ik}) \quad (3-5)$$

La combinación del número de clases y los exponentes borrosos con la variación del FPI, presenta un punto de inflexión indicativo del número de clases más adecuado, caracterizado por el menor grado de borrosidad y la mayor organización interna. Esto es indicativo de que un adecuado agrupamiento tendrá una combinación de grandes valores relativos de  $F$  y pequeños valores de  $H$  (Burrough *et al.*, 2000).

### 3.2.6. Selección del modelo representativo de las clases digitales

Para la selección de un modelo representativo de las clases DST se evaluó la dependencia de la resolución de los MDE, considerando la persistencia y la forma (morfofometría) de dichas clases. La resolución del MDE forma parte del control de las clases digitales dominantes, por lo que se considera que una clase es dependiente de la resolución si no está representada igualmente en todas las resoluciones consideradas (Arrel *et al.*, 2007).

Para que las clases sean reconocidas como persistentes deben estar presentes en la misma área del espacio geográfico a diferentes resoluciones espaciales. Los cambios de las clases DST con la resolución proporcionan información acerca de las características de dichas unidades en el paisaje, cuando se visualizan a diferentes resoluciones (Arrell *et al.*, 2007).

## RESUMEN

El desarrollo sustentable del país requiere con urgencia información de suelos a escalas apropiadas. Sin embargo, los métodos convencionales de levantamiento de suelos son lentos, costosos y limitan la producción de esta información con la prontitud requerida, especialmente en sistemas dinámicos, complejos y de difícil acceso como los ambientes montañosos. La evolución de las tecnologías geoespaciales como los sistemas de información geográfica (SIG) y la teledetección, el desarrollo de las técnicas geomorfométricas y el avance de los métodos estadísticos multivariados y las tecnologías de inteligencia artificial, tales como las redes neuronales artificiales y la lógica borrosa, ofrecen nuevas oportunidades para superar esta limitación y producir información edáfica de manera más eficiente, con adecuada precisión y exactitud, en el menor tiempo posible. Con la finalidad de proponer un enfoque de cartografía digital, alternativo al levantamiento de suelos convencional para la producción de información confiable de suelos y paisajes en cuencas altas a nivel semidetallado, se seleccionó un área de referencia en la cuenca del río Caramacate (tributario del río Guárico), en un sector de la Serranía del Interior de la Cordillera de la Costa, en el estado Aragua, Venezuela. Para la obtención de las unidades de paisaje se realizó una clasificación cuantitativa no supervisada, basada en una red de agrupamiento borroso de Kohonen (FKCN). Se realizaron evaluaciones de modelos digitales de elevación (MDE) de diferentes resoluciones espaciales (8, 10, 15 y 20 m), se seleccionó el modelo más representativo y se obtuvo un total de 10 clases digitales de superficie de terreno que fueron correlacionadas a nivel de tipos de relieve, cuyo significado geomorfológico se interpretó a través de su distribución espacial, de las descripciones de los centros de clases y las funciones de pertenencia generadas en la capa de salida de la red. Se realizó además un muestreo anidado de 5 niveles jerárquicos, para analizar la estructura de variación espacial de los atributos edáficos a varias escalas: Unidades litogeomorfológicas, tipo de paisaje, tipo de relieve o clases digitales de superficie de terreno, ventana  $3 \times 3$  celdas y en una celda (15m) de resolución espacial, lo cual generó un esquema balanceado de 64 observaciones ( $n=2 \times 2 \times 2 \times 4 \times 2$ ). Se determinó que el nivel que contribuyó con mayor proporción a la varianza total de los suelos es el tipo de relieve (laderas de montaña), indicando que las clases de superficie de terreno representan un nivel apropiado para realizar posteriores muestreos, interpolaciones y la generalización de atributos, con énfasis en la cartografía digital de suelos a escalas comprendidas entre 1:20.000 y 1:100.000. Se aplicaron técnicas de cartografía digital de propiedades del suelo tomando en cuenta las clases del modelo digital de superficie de terreno, la geomorfología y la estructura de la variabilidad espacial del suelo. Para ello, se utilizaron cuatro enfoques de inferencia para la representación cartográfica de propiedades de los suelos basada en sitios de muestreo. El primero, es un modelo de inferencia espacial suelo-paisaje combinado con teorías de conjuntos borrosos (FCM), el segundo se basa en un análisis de regresión lineal múltiple (RLM); el tercero es un enfoque que combina el análisis de regresión lineal múltiple con técnicas de interpolación (RK, *Kriging de Regresión*), y el cuarto, se refiere a la combinación de técnicas neuro-borrosas FKCN con la interpolación de residuos (FKR). La evaluación de la confiabilidad de los modelos de predicción permitió determinar que la exactitud está relacionada con la estructura de variación de las

En este sentido, una clase digital puede visualizarse como isométrica si su forma no cambia con la escala.

Para la evaluación se utilizó el índice de desviación de clases (IDC) basado en la diferencia de la proporción relativa de píxeles de cada clase (%Px) y los valores medios de las clases DST ( $X_{CD}$ ) a distintas resoluciones espaciales, para discriminar el modelo de mayor consistencia. Los valores altamente positivos y negativos expresan sobreestimación o subestimación de una clase a una resolución dada, y valores iguales a cero indican tanto isometría como persistencia.

### **3.2.7. Significado geomorfológico de las clases digitales de superficie de terreno**

A las clases DST obtenidas se les asignó el significado geomorfológico a través de una interpretación de las descripciones de sus centroides (concepto central de la clase) y su distribución espacial. También se consideraron los mapas en formato *raster* de los valores de membresías obtenidos con el algoritmo FKCEN, conjuntamente con la información litogeomorfológica y la zona de vida dominante en el sector de estudio.

### **3.2.8. Correlación entre clases digitales de superficie de terreno y el sistema jerárquico de clasificación de paisajes**

Las clases DST fueron correlacionadas con los niveles 5 (unidad litogeomorfológica), 6 (tipo de paisaje) y 7 (tipo de relieve) de la clasificación jerárquica de paisajes propuesta por Elizalde (2012). Para este propósito, se procedió inicialmente a identificar y ajustar al mapa de clases DST, los límites representados en el mapa geológico (Urbani y Rodríguez, 2004; Pineda, 2008). Posteriormente, con la información de los centroides de las clases se separan los diferentes tipos de paisaje (*e.g.* montañas altas, medias y bajas, y valles en Bosque Seco Tropical o Bosque Húmedo Premontano) en cada unidad litogeomorfológica. Finalmente, se separaron las laderas de acuerdo a su orientación dominante, los vallecitos y las crestas dentro de cada paisaje, para obtener los tipos de relieve.

La correlación con el sistema de clasificación del paisaje permitió la integración de varios atributos topográficos de carácter regional (*e.g.* altitud, pendiente media, índice topográfico de humedad) y local (*e.g.* orientación de la pendiente, curvatura), con el efecto de otros factores formadores del paisaje, como el material parental, el clima y la vegetación, obteniéndose una descripción sencilla y objetiva para cada nivel categórico de clasificación.

### **3.2.9. Evaluación del modelo digital de superficie de terreno**

El modelo representativo de las clases DST se evaluó mediante la aplicación de dos métodos de comparación de resultados, utilizando el mapa de clasificación geomorfológica de paisajes de la zona (Pineda 2008). El primer método consistió en una adaptación del

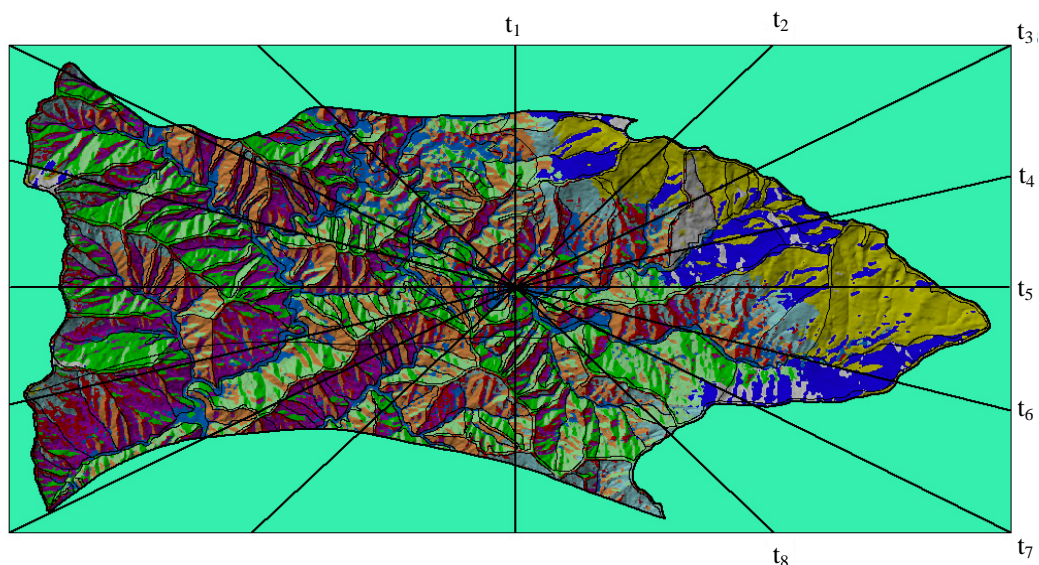


procedimiento de transectas, aplicado por Steers y Hajeck (1979) como método de estimación de la composición taxonómica de unidades cartográficas. Cada transecta consistió en una línea recta que atraviesa el área de estudio, de un extremo a otro, en diversas direcciones (Figura 3-7).

A lo largo de cada transecta, se contó el número de veces que la línea cruzó un límite entre unidades del mapa geomorfológico y el número de veces que el límite cruzado coincidía con un cambio de clase del agrupamiento digital de terreno. Con base en la distribución binomial, se determinó la proporción ( $p$ ) de límites coincidentes como:

$$p = \sum a_i / \sum m_i \quad (3-6)$$

donde  $a_i$  es el número de límites entre unidades de paisaje que coinciden con un cambio de clase DST en la transecta  $t_i$ , y  $m_i$  es el número total de límites entre unidades de paisaje observados en la transecta  $t_i$ .



**Figura 3-7.** Diseño de la aplicación de transectas ( $t_i$ ) para la determinación de la proporción de límites de coincidencia entre modelos en el área de la cuenca del río Caramacate estudiada.

La desviación típica ( $s$ ) de la distribución binomial es igual a:

$$s = p(1 - p) / nt - 1 \quad (3-7)$$

donde  $nt$  es el número de transectas.

Una vez conocida la desviación típica se procedió a calcular el intervalo de confianza (IC) de la proporción de límites coincidentes, para un nivel de probabilidad de 90%, con la siguiente ecuación:

$$IC = P \pm t(s / nt) \quad (3-8)$$

donde  $t$  es el valor tabulado de la distribución de *Students*, para ese nivel de probabilidad y  $nt-1$  grados de libertad.

El segundo método utilizado consistió en la evaluación cuantitativa mediante matrices de confusión y el estadístico *Kappa* (Cohen, 1960, 1968; Hudson y Ramn, 1987). El índice de concordancia *Kappa* se utiliza como un método para evaluar clasificaciones multi-categorías, permitiendo determinar hasta qué punto la concordancia observada es superior a la que se espera obtener por puro azar, y se define de la siguiente manera:

$$k = (\Sigma fo - \Sigma fe) / (n - \Sigma fe) \quad (3-9)$$

donde  $\Sigma fo$  es la suma de las frecuencias observadas en la diagonal principal de una tabulación cruzada,  $\Sigma fe$  es la suma de las frecuencias esperadas en esa diagonal, y  $n$  es el número total de casos <sup>2</sup> (celdas o píxeles).

En caso de un grado de concordancia perfecta el valor de kappa es igual a 1, si la concordancia observada es igual a la esperada kappa es igual a 0; y en el caso de que el acuerdo observado sea inferior al esperado, el índice kappa es menor que cero.

La sumatoria de las frecuencias observadas en la diagonal principal permite la obtención del número de píxeles bien asignados frente al total de píxeles ( $n$ ), mediante la exactitud global del modelo (EG), de la siguiente manera:

$$EG = \Sigma fo / n \quad (3-10)$$

El error de muestreo (EM) para estas validaciones se generó a partir de la fórmula del muestreo aleatorio simple (Chuvieco, 2008):

$$EM = \sqrt{EG \times \Sigma fe / n} \quad (3-11)$$

Finalmente, se obtuvo la fiabilidad real del modelo, determinando el intervalo de confianza para un nivel de probabilidad del 95%:

$$IC = EG \pm t (EM) \quad (3-12)$$

donde  $t$  es el valor tabulado de la distribución de *Students*, para el nivel de probabilidad considerado en la estimación.

---

<sup>2</sup> Para el número total de casos se realizó un muestreo aleatorio basado en 10 000 celdas distribuidas en las diferentes clases digitales de terreno, para ser comparadas con las unidades de paisaje obtenidas por métodos tradicionales de clasificación sistemática.

### 3.3. RESULTADOS Y DISCUSIÓN

#### 3.3.1. Características de los modelos digitales de elevación

El Cuadro 3-3 presenta las características básicas de los diferentes MDE generados en esta parte del estudio, para la posterior selección del modelo representativo de las unidades de paisaje de la cuenca del río Caramacate. Los MDE generados poseen resoluciones espaciales variables entre 10 y 30m, y presentan formas geométricas rectangulares con la finalidad de facilitar la aplicación de algoritmos matemáticos y el procesamiento de datos de las covariables ambientales, las cuales fueron utilizadas como parámetros de entrada para el posterior agrupamiento de píxeles en clases digitales de superficie de terreno (Clases DST).

**Cuadro 3-3.** Características de los MDE generados para el área espacial atributiva en la zona de estudio.

Características	Descripción			
	MDE-10 m	MDE-15 m	MDE-20 m	MDE-30 m
Coordenadas (UTM)	N:1.172.98,3 m	S: 1.109.398,3 m	O: 697.000,9 m	E: 712.600,9 m
Número de filas	790	527	395	258
Número de columnas	1.560	1.040	780	510
Resolución MDE (m)	10	15	20	30
Área de la celda (m <sup>2</sup> )	100	225	400	900
Número de celdas	1.232.400	548.080	308.100	131.580
Superficie (ha)	12.050	12.332	12.324	11.842

Se evidencia que la relación entre el grado de resolución espacial y el número de píxeles de cada modelo es inversamente proporcional, con una tendencia exponencial. El aumento del número de datos a medida que se reduce el tamaño de los píxeles exige mayor esfuerzo computacional.

#### 3.3.2. Número óptimo de clases digitales de superficie de terreno

Para la estimación de los modelos de las clases DST mediante el sistema de clasificación neuro-borroso (FKCN), se utilizaron diez (10) variables derivadas del MDE, y una (1) variable derivada de la imagen satelital multiespectral (NDVI).

El Cuadro 3-4 muestra los estadísticos descriptivos de estas variables, calculados a partir del MDE base de 20 m de resolución. La variación del índice de rendimiento borroso (FPI) en función de los valores de exponente borroso ( $\phi$ ) y número de clases DST, se muestra en la Figura 3-8.

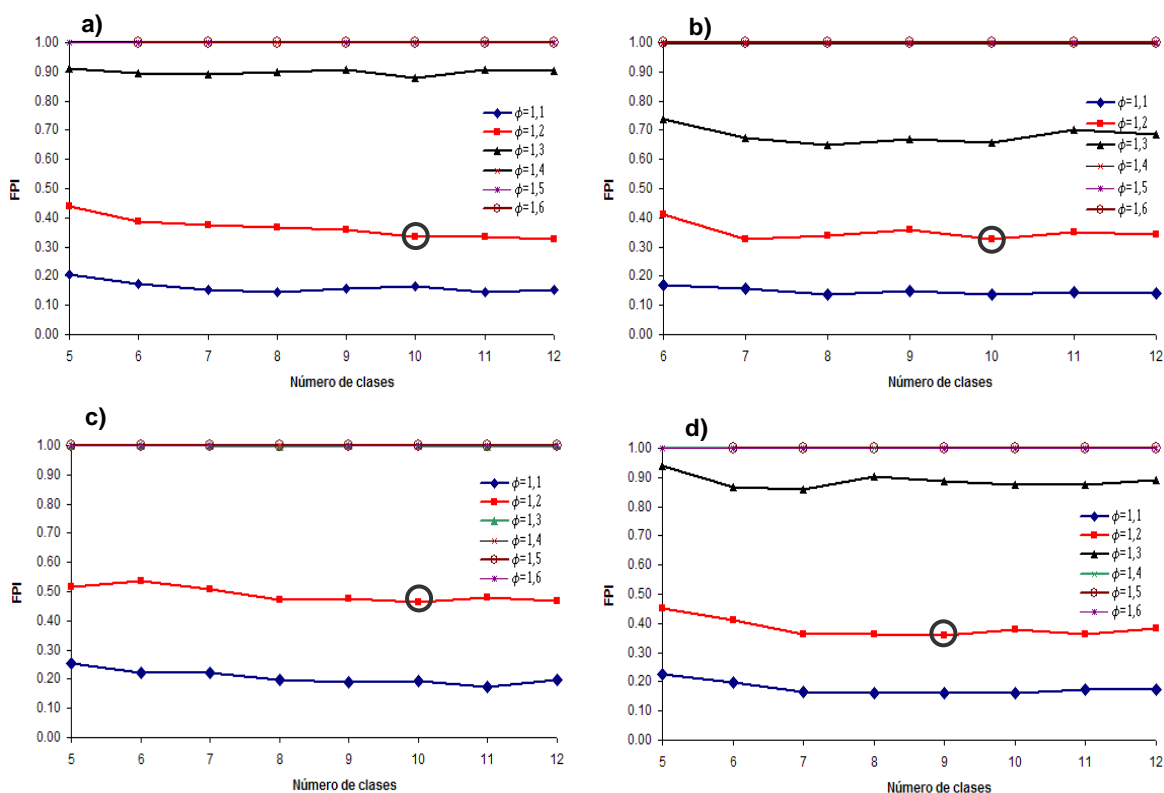
En la figura indicada, el FPI varía de 0 a 1 y es una medida del grado de separación entre las clases borrosas. Un valor igual a 0 indica que no hay borrosidad en los datos y las clases son

discretas. Por el contrario, cuando FPI es igual a 1, la borrosidad es máxima, y cada individuo carece de una afiliación clara a cualquier clase, ya que todos los centros de clase tienen valores similares (Odeh *et al.*, 1992; Morales y Vilorio, 2006).

**Cuadro 3-4.** Parámetros estadísticos descriptivos del MDE de la cuenca del río Caramacate.

Atributo	Mínimo	Máximo	Media	CV (%)	Varianza
Altitud (m)	433,5	1421,0	802,7	193,7	37530
Aspecto (rad)	2,0E-06	6,28	3,23	1,79	3,22
PR	0,0	1,0	0,49	0,12	0,01
Area_C (m <sup>2</sup> )	225	5489271	9593	360,3	2,9E+09
ITH	5,5	19,2	8,8	1,81	3,28
Pendiente (rad)	0,00	1,20	0,55	0,21	0,04
Perfil_C (m.m <sup>-2</sup> )	-0,3528	0,3713	-0,0016	0,06	0,0037
Plano_C (m.m <sup>-2</sup> )	-0,3384	0,3731	0,0017	0,07	0,0045
ITS	0,0	389,0	30,8	21,2	447,9
IPE	0,0	55278	242,2	915,0	837312
NDVI	0,28	0,65	0,047	0,15	0,02

PR: Posición Relativa; Area\_C: Área de Captación; ITH: Índice Topográfico de Humedad; ITS: índice de Transporte de Sedimentos; IPE: Índice de Potencia de la Escorrentía; NDVI: Índice de Vegetación de Diferencia Normalizada. <sup>1</sup>MDE 20 m de resolución espacial.



**Figura 3-8.** Variación del índice de rendimiento borroso (FPI) en función del número de clases para once variables ambientales: a) MDE 10 m, b) MDE 15 m; c) MDE 20 m y d) MDE 30 m.

En todas las situaciones, independientemente de la resolución espacial, el FPI se acerca al valor de la unidad (1) para todas las clases con valores  $\phi$  comprendidos entre 1,3 y 1,6. Esto indica que la borrosidad alcanzada es máxima y no hay distinción entre clases, por lo que estos resultados son descartados como posibles modelos. También se observa que cuando el FPI se aproxima a cero ( $\phi=1,1$ ) se pierde la borrosidad del modelo, y los límites de las clases obtenidos tienden a ser más bruscos, lo cual tampoco es adecuado como modelo representativo.

Para todos los modelos obtenidos se destaca un punto de inflexión en la cual existe una minimización del FPI con valores comprendidos entre 0,34 y 0,43 (señalados con un pequeño círculo), que contribuye a la determinación del número óptimo de clases caracterizadas por ser menos borrosas y menos desorganizadas internamente (mayor F: coeficiente de partición y menor MPE). Por estas razones, se seleccionó como número óptimo de clases el valor del punto de inflexión en las curvas correspondientes a  $\phi= 1,2$ , obteniéndose 10 clases DST para los modelos generados con resoluciones de 10, 15 y 20 m, y 9 clases para el modelo de 30 m.

### **3.3.3. Selección del modelo representativo de las clases digitales**

Para la selección del modelo representativo de las clases digitales del área de estudio se evaluó la dependencia de la resolución de las clases DST a través de la persistencia, la forma en el espacio geográfico (morfometría) y el índice de desviación de las clases (IDC). En el Cuadro 3-5 se presentan los resultados de la dependencia de la resolución de las diferentes clases DST (utilizando 11 variables auxiliares), donde se indica la proporción de píxeles clasificados en cada clase de terreno.

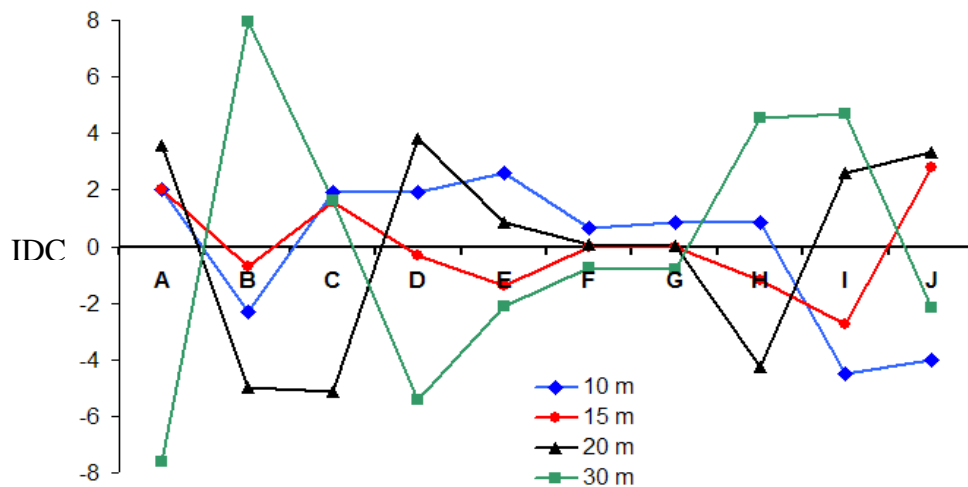
En el mismo Cuadro 3-5, también se señala el promedio del conjunto de píxeles para cada clase (en términos relativos), y en la Figura 3-9 se representa el valor del IDC, basado en la diferencia entre la proporción relativa de píxeles y la media de las clases a distintas resoluciones espaciales. De igual manera, los modelos digitales de superficie de terreno obtenidos a diferentes resoluciones se presentan en la Figura 3-10, donde se visualiza la distribución espacial y la forma de cada una de las clases con relación a la escala.

Los resultados evidencian que la mayoría de las clases DST son dependientes de la resolución, ya que no están representadas idénticamente en todas las escalas consideradas. Las clases más dependientes de la resolución, son A (crestas) y D (Laderas), debido a que no son reconocidas en todos los niveles de resolución y desaparecen con los tamaños de píxeles más gruesos (30 m). El resto de las clases son persistentes en todos los niveles, ya que están presentes en el mismo espacio geográfico a diferentes resoluciones espaciales.

**Cuadro 3-5.** Variación de la proporción (%) de las clases del modelo digital de superficie de terreno con la resolución del MDE, en el área de estudio.

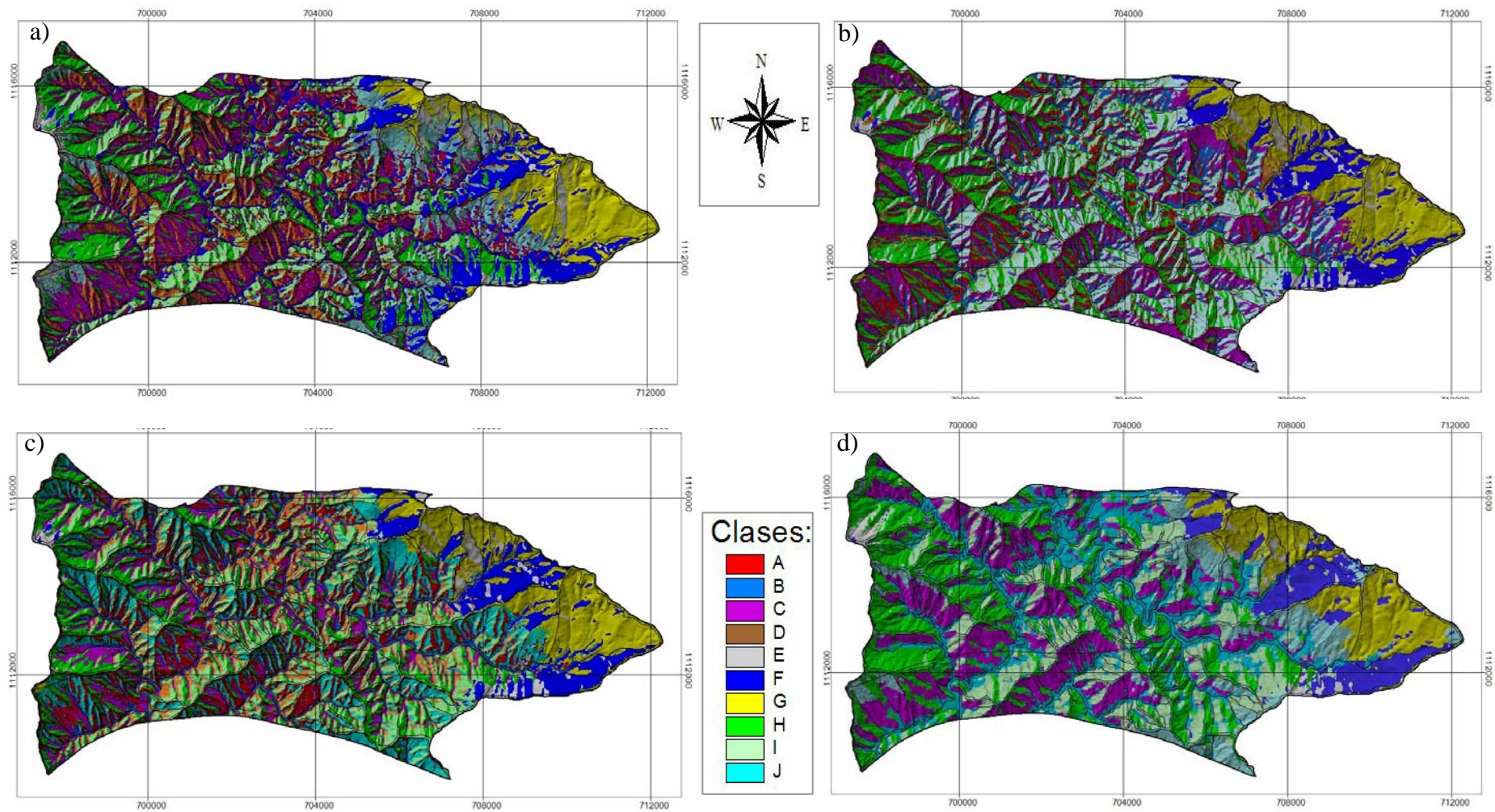
Clase	Fisiografía	Resolución del MDE <sup>1</sup>				Media
		10 m	15 m	20 m	30 m	
A	Crestas y Vigas	9,6	9,7	11,2	0,0	7,6
B	Vegas	7,4	9,0	4,8	17,7	9,7
C	Laderas	16,2	15,8	9,1	15,9	14,3
D	Laderas	7,3	5,1	9,3	0,0	5,4
E	Laderas	9,5	5,5	7,7	4,8	6,9
F	Laderas	8,9	8,2	8,3	7,5	8,2
G	Laderas	9,8	8,9	9,0	8,1	8,9
H	Laderas	17,1	15,1	12,1	20,9	16,3
I	Laderas	6,3	8,1	13,4	15,5	10,8
J	Laderas	7,8	14,6	15,1	9,6	11,8

<sup>1</sup>Valores relativos de píxeles, equivalentes a la superficie relativa ocupada por cada clase.



**Figura 3-9.** Variación de las diferencias entre la proporción relativa de píxeles y la media de las clases digitales de superficie de terreno a distintas resoluciones espaciales en el área estudiada.

De igual manera, las clases de mayor similitud son F y G entre 10 y 20m de resolución, las cuales se consideran isométricas debido a que sus formas varían poco con la escala. Estas apreciaciones indicarían que el MDE 30 m no permite reconocer algunos tipos de relieve que ocupan poca superficie en el área de estudio, como las crestas (clase A) y las laderas correspondientes a la clase D. En el modelo digital de superficie de terreno a 20 m de resolución se confunden vegas con crestas. Las crestas aparecen sobrestimadas porque se encuentran formando parte de otras unidades (*e.g.* clase C), y las vegas son subestimadas en comparación con los modelos de mayor resolución; por lo tanto, este modelo también dejaría de ser el más apropiado para representar las clases en la zona estudiada.



**Figura 3-10.** Modelos digitales de superficie de terreno a diferente resolución espacial: a) MDE 10 m, b) MDE 15 m, c) MDE 20 m, y d) MDE 30 m en la zona estudiada.

Los modelos de superficie de terreno de mayor resolución espacial (10 y 15 m) permitieron identificar todas las clases digitales, basados en la coherencia espacial existente dentro de la zona de estudio. Las unidades obtenidas con estos MDE presentan visualmente mayor claridad, mejor distinción y menor borrosidad dentro de las unidades, siendo diferentes a las obtenidas con resoluciones más gruesas (20 y 30 m). El MDE de 15 m representa el mayor número de unidades digitales con menor desviación a los valores medios de cada clase (menor IDC). Como su resolución es menos exigente que la del MDE de 10 m desde el punto de vista de esfuerzo computacional, se escogió como el modelo más apropiado de las clases DST en el sector de interés de la cuenca del río Caramacate a la escala de trabajo.

La resolución espacial de 15 m se aproxima al valor empleado como “regla de oro” en el ámbito cartográfico, equivalente a una tercera parte de la escala de trabajo utilizada en la investigación (1:50.000). Por ello, constituyó una solución óptima acerca del modelo más representativo de las unidades de paisaje de la cuenca en estudio.

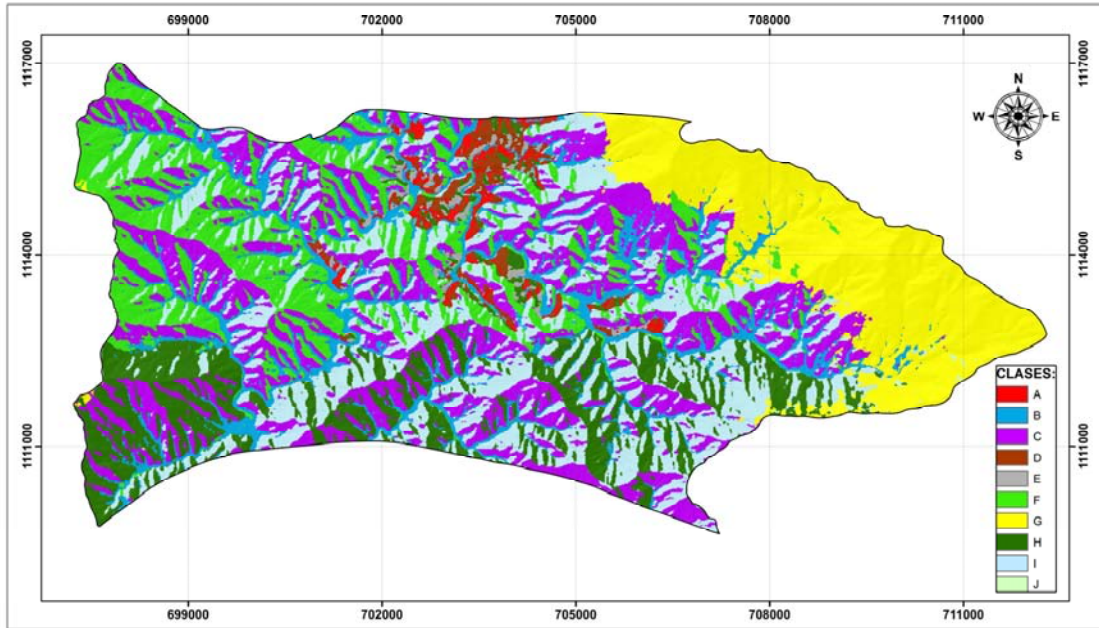
#### **3.3.4. Influencia de la litología como variable discreta en la clasificación digital de la superficie de terreno**

En la mayoría de los sistemas de clasificación jerárquica de paisajes, las unidades litológicas (UL) son consideradas como un criterio de clasificación geomorfológica a diferentes niveles categóricos (Elizalde, 2012; Zinck, 2012). El sistema clasificador neuro-borroso permite la utilización de variables discretas como capa de entrada, por lo que se realizó una evaluación tomando en cuenta las UL dominantes en el sector estudiado.

Para esta situación, la red FKCN fue entrenada con 10 atributos morfométricos a 15 m de resolución espacial, más el NDVI y las UL como parámetros de entrada. Con base en el índice de borrosidad alcanzada (FPI), se seleccionó la combinación de 10 clases DST y exponente borroso ( $\phi$ ) 1,2 como la más adecuada para esta nueva clasificación.

En la Figura 3-11 se visualiza la separación entre las formaciones geológicas dominantes en la cuenca (El Caño-El Chino en la parte superior y El Carmen en la parte inferior), y dentro de ellas las partes más elevadas dominada por paisajes de montaña (clases C, F-J), las zonas de transición correspondiente al sistema de piedemonte (clases A, D, E) y las áreas bajas dominadas por paisajes de valles intramontanos (clase B). Sin embargo, en el modelo obtenido no fue posible discriminar los tipos de relieve relacionados con las crestas y vigas de las unidades de paisajes de montaña dominantes en el sector, así como tampoco algunas unidades de paisaje del sector Nordeste del área de estudio, debido a la alta discretización entre categorías.





**Figura 3-11.** Modelo de clases digitales de superficie de terreno incluyendo la variable litología en el proceso de clasificación en el área estudiada.

Del análisis realizado se infiere que, aunque las RNA son capaces de discriminar clases diferentes basadas en aspectos litológicos, las unidades generadas son más discretas (con menos borrosidad interna), lo que parece indicar que dicha variable tiene alta influencia en el proceso de clasificación, ya que las categorías obtenidas incluyen varias de las unidades que podrían ser discriminadas cuando se utilizan solamente las variables continuas. Es decir, las variables discretas tienen un enorme peso y ejercen una gran influencia en la organización, distribución y definición de las clases digitales finales.

Es importante tomar en cuenta que las variables discretas requieren un trabajo manual para la previa delineación de las unidades, lo que implica inversión de tiempo y hace que la operación no sea considerada automatizada del todo. Por tales razones, en la definición de clases geomorfométricas, se vislumbran al menos tres opciones: a) incluir las diferentes unidades litológicas como una variable discreta; b) estratificar la zona de estudio por la litología existente (dejando un área de solapamiento entre unidades adyacentes) y clasificar por separado el área correspondiente a cada unidad litológica; c) clasificar toda el área de estudio e introducir posteriormente la información litológica en la interpretación de los resultados.

En este sentido, la opción (a) parece ser la menos apropiada porque la variable discreta puede introducir artefactos o sesgos en la clasificación del terreno, como se ha demostrado en esta investigación. La escogencia entre las opciones (b) y (c) dependerá del tamaño del área ocupada por cada unidad litológica, la capacidad de procesamiento

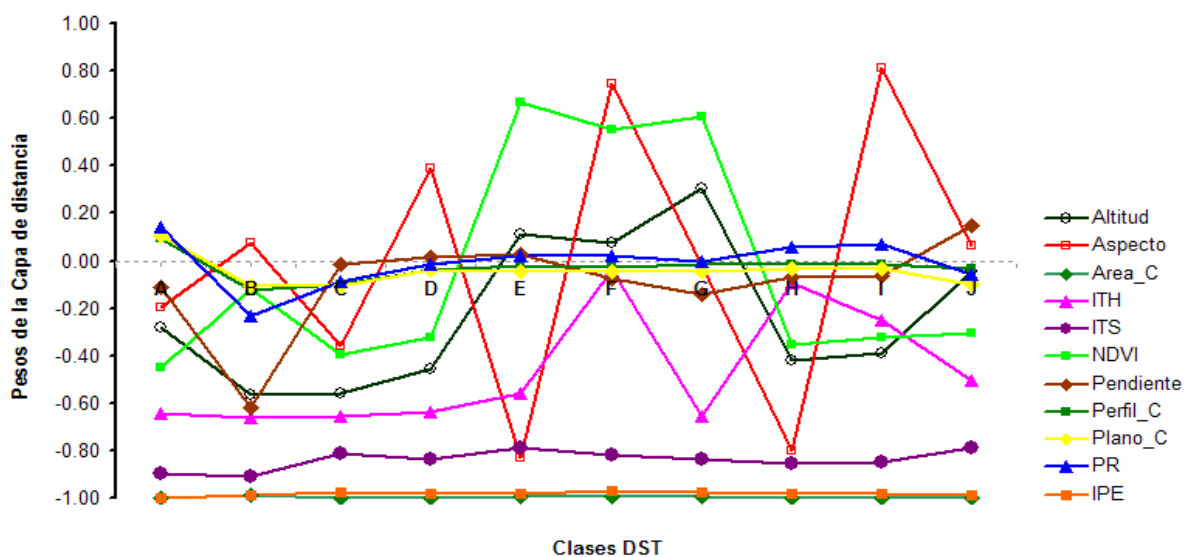
propiedades de los suelos, ya que es alta cuando la mayor variación ocurre a nivel del tipo de relieve y paisaje, y es baja cuando existe una alta variación de corto alcance. Los mapas de cada propiedad del suelo se integraron utilizando dos técnicas de inferencia espacial fundamentadas en tecnologías de inteligencia artificial. La primera consistió en la aplicación del algoritmo de agrupamiento neuro-borroso *FKCN*, empleando once conjuntos de neuronas artificiales como capas de entrada al sistema, para la obtención final de un modelo digital con nueve clases para el enfoque *FCM-FKCN*, y diez clases de suelo locales para los enfoques *RK-FKCN* y *FKR-FKCN*. La segunda técnica, se basó en la aplicación de un modelo suelo-paisaje combinado con teorías de conjuntos borrosos para un grupo de categorías taxonómicas preestablecidas. Finalmente, los resultados obtenidos se sintetizaron como una propuesta de cartografía digital para la generación de información de suelos y paisajes en áreas montañosas.

del *hardware*, y de la naturaleza del límite (gradual o abrupto) entre las unidades litológicas. En el caso de estudio de esta investigación, la opción (c) parece más adecuada. Consecuentemente, se utilizarán las unidades litológicas como información base para la interpretación de las clases DST, en vez de incluirlas como parte de la RNA. De esta manera, se evitará el efecto distorsionante de la clasificación cuando se introduce la litología a la red neuro-borrosa, como una variable discreta.

### 3.3.5. Evaluación del número de variables ambientales en la definición de unidades geomorfométricas

El número de variables utilizadas como entradas al sistema de clasificación FKCN es equivalente al número de neuronas de la red, por lo que fue necesario realizar pruebas de sensibilidad basadas en las evaluaciones con diferentes grupos de neuronas artificiales. Esto se hizo a fin de determinar el grado de influencia de cada una de ellas en la definición y reproducibilidad de unidades geomorfométricas. En tal sentido, se utilizó el MDE de 15 m de resolución espacial, tomando en cuenta las once (11) variables ambientales continuas empleadas en el análisis del terreno.

Para la evaluación de la sensibilidad de los parámetros de superficie de terreno se emplearon los valores normalizados de las variables de entrada en los centroides de cada una de las clases DST generadas por FKCN. En este sentido, en la Figura 3-12 se presentan los valores de las variables auxiliares, representadas por los pesos de la capa de distancia a cada una de las neuronas de entrada a la red FKCN y la capa de salida.



**Figura 3-12.** Valores de los pesos de la capa de distancia de cada uno de los atributos en las clases DST generadas por FKCN en el área estudiada.

Los pesos de la capa de distancia son equivalentes a los valores de las conexiones entre

neuronas, y son generados durante el proceso de entrenamiento de la red, tomando en cuenta las relaciones topológicas entre los parámetros de entrada y la salida gráfica.

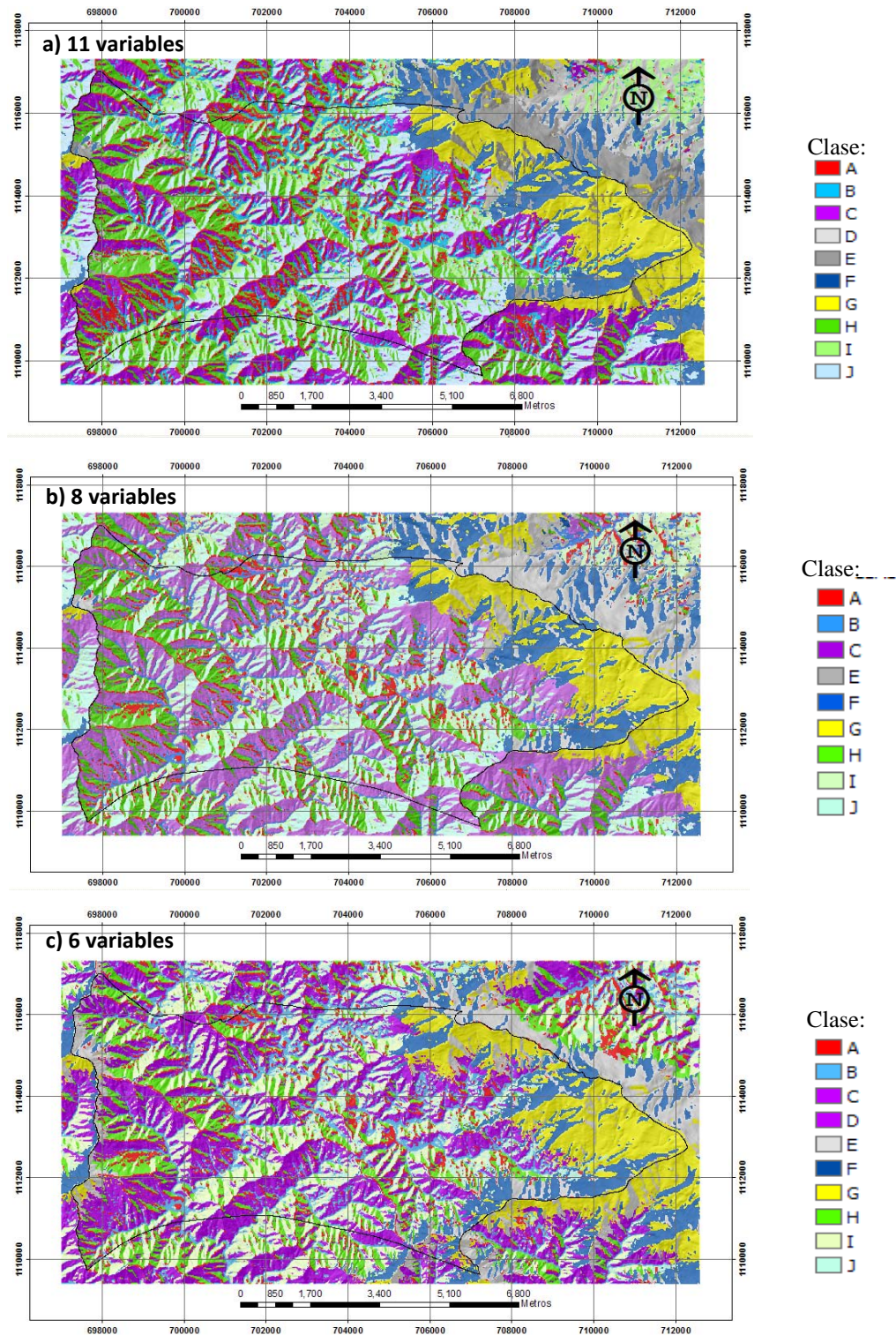
Las variaciones de las variables en el rango -1 a 1 reflejan la importancia de los atributos primarios en la definición de las diez (10) clases digitales del terreno. Se puede visualizar la poca influencia en la discriminación de las clases DST aportada por el área de captación y el IPE, cuyos pesos contribuyen muy poco a la diferenciación de las clases digitales.

Para fortalecer los argumentos anteriores se realizaron comparaciones gráficas y numéricas entre todo el conjunto de variables continuas (11) y las de mayor influencia en la discriminación geomorfométrica, tales como: altitud, pendiente, aspecto, ITH, PR, Perfil C y Plano C y NDVI (8 variables de entrada a la red). Posteriormente se eliminaron los atributos ITH y NDVI para evaluar el comportamiento de seis (6) variables de entrada.

En todos los casos el entrenamiento de la red se realizó para diez (10) clases y un valor de  $\phi = 1,2$  a fin de realizar las comparaciones pertinentes. La salida gráfica de estos tres conjuntos de variables auxiliares se presenta en la Figura 3-13, en la cual se visualiza una gran similitud en la distribución geográfica de las clases DST, corroborándose la influencia de los pesos de las variables en la discriminación morfométrica.

Para las comparaciones numéricas entre los diferentes conjuntos de variables se realizaron otras evaluaciones empleando un mínimo de atributos topográficos. En primer término, del conjunto de seis (6) variables se prescindió de la curvatura del terreno, quedando cuatro (4) variables, y posteriormente se excluyó la posición relativa (PR) para un mínimo de tres (3) variables auxiliares, a fin de evaluar el comportamiento de los atributos topográficos primarios (altitud, grado y orientación de la pendiente), los cuales son calculados directamente de los datos de elevación.

Los atributos primarios suministran características discriminantes para la diferenciación de unidades geomorfométricas, y para la descripción y distinción entre dichas clases en áreas de montaña (Adediran *et al.*, 2004; Chagas, 2006; Valera, 2012). La orientación de la pendiente (Aspecto) se utilizó en todas las combinaciones, ya que produce una zonificación adicional o una definición *a priori* que representa claramente las diferentes unidades geomorfométricas (Drägut y Blanshke, 2006). Además, es una variable muy importante para la representación de la superficie del terreno y el mapeo de unidades homogéneas desde el punto de vista geomorfológico, inclusive en los niveles detallados de la clasificación de paisajes propuesta por Elizalde (2012).



**Figura 3-13.** Modelos resultantes de las pruebas de sensibilidad al número de variables utilizadas como entrada al sistema FKCN en el área estudiada.

Los resultados de la variación de la cantidad relativa de píxeles o celdas correspondientes a las clases DST para las agrupaciones de 3 a 11 variables se resumen en el Cuadro 3-6, la cual complementa la descripción de la variación espacial del agrupamiento de píxeles.

**Cuadro 3-6.** Variación de la proporción de las clases digitales con el número de variables ambientales en el área de estudio.

Clase DST	Fisiografía	Número de variables utilizadas <sup>1</sup>					Media
		11	8	6	4	3	
A	Crestas, Vigas	9,7	6,7	6,9	7,2	-	7,6
B	Vegas	9,0	7,7	8,7	13,9	15,6	11,0
C	Ladera	15,8	24,1	11,8	14,8	14,0	16,1
D	Ladera	5,1	-	13,8	9,3	10,2	9,6
E	Ladera	5,5	8,0	7,8	10,4	7,5	7,8
F	Ladera	8,2	8,4	9,3	7,1	9,3	8,5
G	Ladera	8,9	8,4	9,9	14,2	10,1	10,3
H	Ladera	15,1	13,1	11,3	8,8	21,5	14,0
I	Ladera	8,1	9,6	12,7	14,2	11,8	11,3
J	Ladera	14,6	14,1	7,8	-	-	12,2

<sup>1</sup> Valores expresados en % de píxeles clasificados

De los datos atributivos y espaciales se obtuvo la siguiente información: a) la mayoría de las clases correspondientes a laderas de paisajes de montaña con diferentes exposiciones y la clase B (vegas) son las más persistentes, ya que aparecen en todos los modelos obtenidos; b) las clases menos persistentes son A (crestas), D y J (laderas de montaña); c) la clase A está presente en todos los grupos en los que juega papel importante la posición relativa; d) en los conjuntos de pocos atributos de superficie de terreno (3 y 4 variables) se evidencia una sobrestimación de las vegas; d) la mayor sobrestimación está presente en la clase C del grupo de 8 variables, ya que la clase faltante forma parte de ella, y e) en general, existen pocas desviaciones de la proporción de píxeles de las clases con 11, 6 y 3 variables, con respecto a la media del conjunto de datos.

El análisis anterior indica que con los centroides de los atributos primarios (la altitud, el grado y la orientación de la pendiente) más la posición relativa, existen definiciones concretas de las unidades digitales del terreno. En el sector de la cuenca estudiada queda demostrado que seis (6) atributos de superficie de terreno parecen ser suficientes para describir sus potencialidades morfométricas. Sin embargo, en esta investigación se consideró recomendable incluir un mayor número de variables ambientales, a fin de obtener valores de centroides de clases que aportarán mayor información a la interpretación geomorfológica de las clases DST. Por esta razón, de las variables ambientales iniciales se excluyeron solo el área de captación y el IPE, por su escasa

contribución a la diferenciación de las clases (Figura 3-13).

### 3.3.6. Significado geomorfológico de las clases digitales de superficie de terreno

La salida del sistema de clasificación FKCEN permitió identificar unidades de paisaje, que pueden ser correlacionadas y descritas hasta el nivel de tipos de relieve del sistema de clasificación de paisaje propuesto por Elizalde (2012). El valor de los centroides obtenidos con los parámetros de entrenamiento de la red en cada una de las clases DST, a 15 m de resolución espacial se presenta en el Cuadro 3-7.

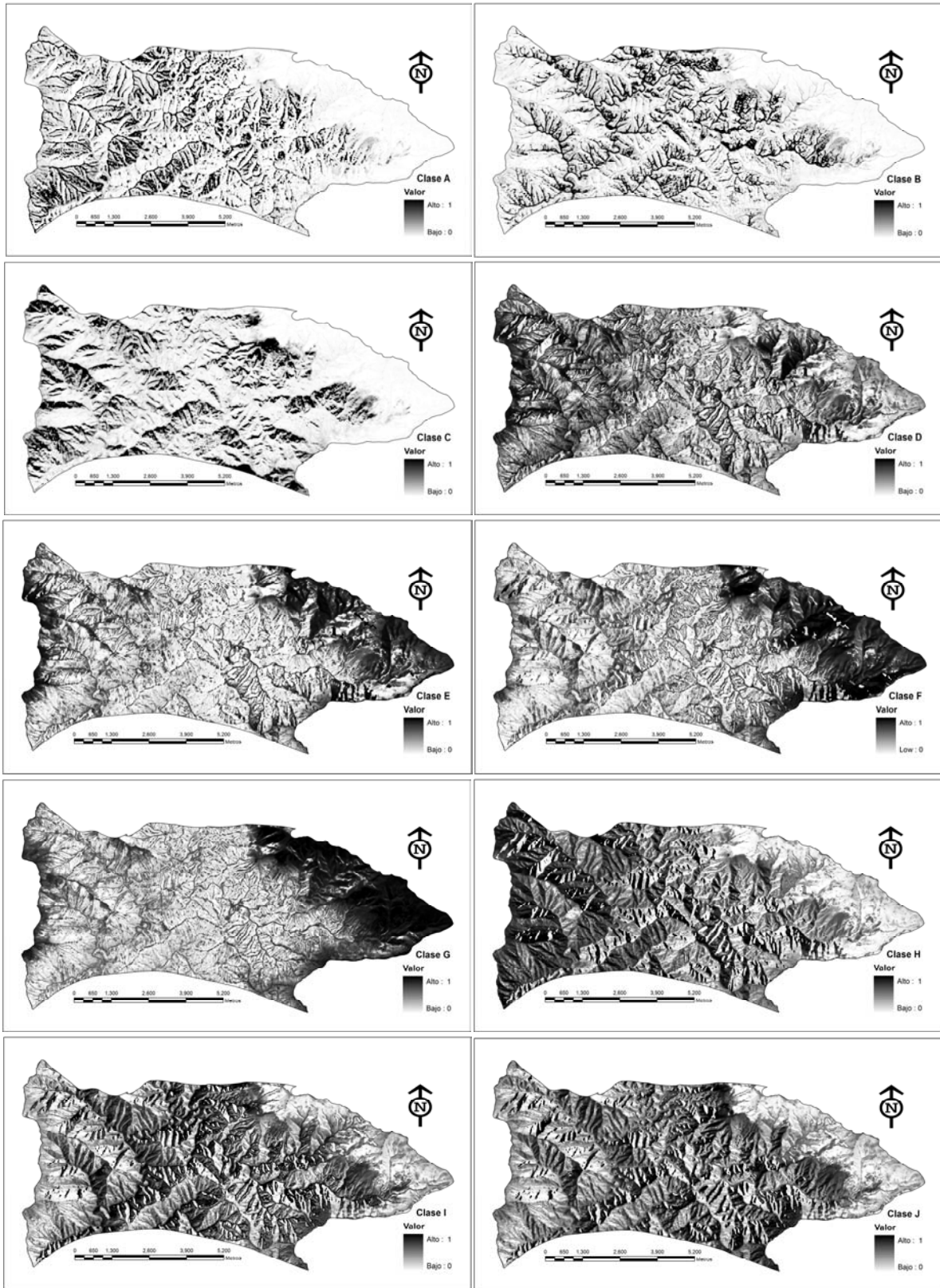
**Cuadro 3-7.** Valores del concepto central de cada clase digital de superficie de terreno identificada en la zona de estudio de la cuenca del río Caramacate.

Clase DST	Unidad Fisiográfica	Variables Auxiliares								
		Alt (m)	PR	Asp (°)	Pend (%)	ITH	Perfil C (m.m <sup>-2</sup> )	Plano C (m.m <sup>-2</sup> )	ITS	NDVI
A	Crestas SE	707	0,61	127	30	7,9	0,0451	0,0479	20	0,19
B	Vegas S	648	0,38	207	12	12,7	-0,0501	-0,0362	18	0,36
C	Ladera S, MM	764	0,50	183	56	7,8	-0,0012	0,0001	36	0,19
D	Ladera E, MM	743	0,45	77	34	7,9	-0,0267	-0,0090	32	0,28
E	Ladera NE, MA	1055	0,51	47	53	8,6	0,0048	0,0037	42	0,54
F	Ladera NO, MA	988	0,50	312	45	8,7	-0,0010	0,0010	35	0,50
G	Ladera S, MA	1122	0,51	195	37	8,9	0,0063	0,0042	31	0,52
H	Ladera NE, MB	722	0,48	50	50	10,8	-0,0123	-0,0067	28	0,19
I	Ladera NO, MB	688	0,44	310	31	8,2	-0,0349	-0,0169	29	0,27
J	Ladera NO, MM	769	0,54	300	54	7,7	0,0208	0,0219	41	0,21

MA: Montañas altas, MM: Montañas medias, MB: Montañas bajas, ALT: altitud, PR: Posición relativa, Asp: Orientación, Pend: Pendiente del terreno, ITH: Índice topográfico de humedad, Perfil C: Perfil de curvatura, Plano C: Plano de curvatura, ITS: Índice de transporte de sedimentos, NDVI: Índice de vegetación de diferencia normalizada.

La información atributiva del concepto central de las clases DST está integrada a la información geoespacial de los valores de las funciones de pertenencia (FP) o grado de similitud a dichas clases. Los valores FP de cada píxel a las clases DST se esquematizan en la Figura 3-14, por medio de una escala gradual que varía entre 0 y 1, donde los colores oscuros indican un mayor grado de similitud del píxel a la clase DST correspondiente (FP >0,5), mientras que los colores claros revelan un menor grado de membresía a dicha clase.

La red FKCEN produce un modelo espacial integrado de las diferentes clases DST asignando a cada píxel la clase con el valor más alto de la función de pertenencia. En este modelo la clase B se relaciona con las vegas de valles intramontanos, la clase A corresponde a crestas y vigas de montañas y las clases restantes correlacionan con laderas de montañas, con diferencias en altura relativa, exposición, curvatura del terreno y cobertura vegetal.



**Figura 3-14.** Modelos de la distribución de los valores de las funciones de pertenencia a las diferentes clases digitales de superficie de terreno.



De la interpretación de los centroides de clase (Cuadro 3-7), su distribución espacial y los valores de FP se extrae el siguiente significado geomorfológico de las clases DST: la *clase B* se relaciona con las zonas más bajas de la cuenca, y corresponde a las vegas de los valles intramontanos. Presenta pendientes moderadamente inclinadas (< 12%), orientadas hacia el Sur, con forma topográfica convergente de alta concavidad, por lo que tiene el mayor ITH y la mayor área de captación de la cuenca. También se caracteriza por presentar una cobertura vegetal dominada por bosques de galería y actúa como depósito de materiales provenientes de las partes más altas de la cuenca. Esta clase se relaciona con las áreas de sedimentos del cuaternario (Qa1), acarreados principalmente por los ríos Caramacate, El Trompillo y El Altar, y la quebrada El Onoto.

La *clase A* corresponde a crestas y vigas de montañas con una altitud media de 707 msnm, orientadas hacia el SE, pendientes del 30% y forma topográfica divergente altamente convexa, lo que favorece el transporte y la pérdida de materiales. Estas áreas presentan los menores valores del índice de humedad y del área de captación, con una escasa cobertura vegetal dominada principalmente por gramíneas.

La *clase G* se relaciona con laderas de montañas altas con alturas superiores a los 1122 msnm, con orientación al Sur y pendientes superiores al 37%, con formas topográficas convexas-convexas. Presenta los mayores valores de NDVI en la cuenca, con una cobertura vegetal dominada por bosques siempreverdes y semidecíduos, que permite amortiguar la escorrentía evitando el transporte de sedimentos. Esta clase tiene mayor presencia en la unidad VCñc, específicamente en las montañas del Topo Bramador (NE de la cuenca) y en la montaña Valladolid, correspondiente a la zona boscosa del sector El Altar.

Las *clases E y F*, también son laderas de montañas altas con alturas comprendidas entre 988 y 1055 msnm, con pendientes muy escarpadas que oscilan entre 45 y 53%, orientadas en sentido NE y NO, respectivamente. Presentan formas topográficas predominantemente convexas-convexas y lineales, lo que influye en los valores del área de captación y el índice de humedad. Sin embargo, los valores de NDVI son indicativos de una cobertura vegetal boscosa, lo cual aminora la influencia de la escorrentía superficial y la pérdida de sedimentos. La clase E se distribuye en la unidad litomorfológica VCñc, en la parte alta de la montaña del Topo Bramador al Norte de la Fila Valladolid, mientras que la clase F tiene presencia en la Fila El Altar de la unidad VCca.

Las *clases C y D* corresponden a laderas de montañas medias, con una altitud ubicada entre 743 y 764 msnm, con diferentes exposiciones, pendientes que varían entre 34 y 56% y formas topográficas cóncava-lineal y cóncava-cóncava, respectivamente. Ambas clases difieren en cuanto a valores de área de captación, índice de humedad y NDVI, siendo inferiores en la clase C por su ubicación relativa, lo que influye en una mayor

susceptibilidad a los procesos erosivos. La clase C está distribuida geográficamente en sectores de El Onoto, Cerro Monte Alto y El Cacao de la unidad VCñc, y en la Montaña La Cruz y La Tarasca de la unidad VCca. La clase D se distribuye principalmente en la formación geológica VCñc, en la parte baja del Topo Bramador al Este de la zona estudiada, asociada con la clase E.

Las *clases I y J*, son laderas de montañas bajas y medias asociadas, con alturas comprendidas entre 688 y 769 msnm, con pendientes escarpadas que oscilan entre 31 y 54%, orientadas en sentido NO. Presentan formas topográficas cóncava-cóncava y convexa-convexa, y tienen valores contrastantes de área de captación e índice de humedad, con valores de NDVI indicativos de una cobertura vegetal integrada por gramíneas y arbustales a nivel de los interfluvios. La clase J es más susceptible a la escorrentía y pérdida de sedimentos, debido que presenta pendientes más escarpadas. Ambas clases están distribuidas geográficamente en las partes bajas (adyacente a las vegas y canales naturales de drenaje) y altas de las laderas de las montañas El Cacao y Los Ranchos de la unidad VCñc, y en la Montaña La Cruz y La Tarasca de la unidad VCca.

La *clase H* se relaciona con laderas de montañas bajas con alturas medias de 722 msnm, con orientación NE y pendientes escarpadas de 50%, con formas topográficas convexas-convexas. Presenta valores de NDVI con una cobertura dominada por vegetación herbácea, que favorece la escorrentía y el transporte de sedimentos. Esta clase está distribuida geográficamente en sectores de El Cacao, y parte de La Palmita y La Cruz de la unidad VCñc, y en la Montaña Pulidera, Bachaquero y Carapita de la unidad VCca.

Las clases digitales de terreno producidas por la red neuro-borrosa fueron divididas en función de su geología subyacente para obtener clases con mayor significado, para el análisis geomorfológico.

### **3.3.7. Correlación entre las clases digitales de superficie de terreno y el sistema jerárquico de clasificación de paisajes**

En el enfoque aplicado en este estudio se procedió de manera opuesta a los sistemas convencionales de clasificación jerárquica de paisajes, ya que se empleó un enfoque inductivo donde se organizan las unidades desde lo particular hasta lo general. En este proceso, las clases digitales de terreno determinaron el marco cartográfico de las unidades geomorfológicas en el nivel correspondiente al tipo de relieve. Estas clases fueron integradas en clases de las categorías superiores del sistema de clasificación, tales como los tipos de paisaje y las unidades litogeomorfológicas, con mayor heterogeneidad. Esto permitió acelerar el proceso de identificación y definición de las categorías de paisaje.

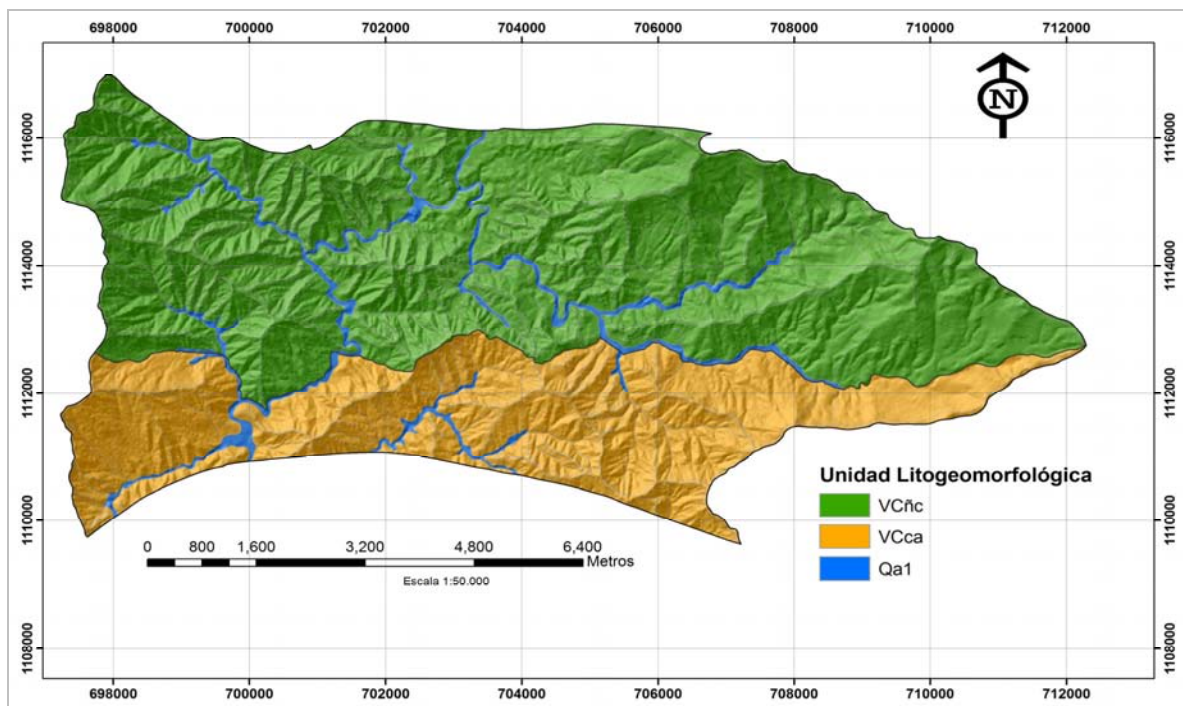
### 3.3.7.1. Unidades Litogeomorfológicas

Para obtener el mapa de unidades litogeomorfológicas, se comparó el patrón de las clases DST del mapa FKC� con el mapa geológico de Urbani y Rodríguez (2004) y el mapa geomorfológico de Pineda (2008). De esta manera, los límites entre las unidades geológicas fueron ajustados a los cambios observados en el mapa de clases digitales, diferenciando las dos formaciones geológicas dominantes en el sector. También se logró distinguir la unidad conformada por aluviones y sedimentos recientes arrastrados por el río Caramacate (Qa1), la cual atraviesa el sector de estudio en sentido Norte-Sur. Un resumen de estas unidades se presenta en el Cuadro 3-8 y en la Figura 3-15, donde se visualiza la línea geológica superficial vectorizada y ajustada a los límites de las clases DST.

**Cuadro 3-8.** Unidades litogeomorfológicas dominantes en el sector de estudio de la cuenca del río Caramacate.

Categoría	Denominación	Descripción	Área (ha)
Unidad Litogeomorfológica	VCñc	Formación Metatobas de El Caño-El Chino	4340
	VCca	Formación Metalavas de El Carmen	2128
	Qa1	Sedimentos del Cuaternario	292
			6760

Fuente: Urbani y Rodríguez (2004); Pineda (2008)



**Figura 3-15.** Principales unidades litogeomorfológicas discriminadas en la zona de estudio.

### 3.3.7.2. Tipo de paisaje

El mapa de unidades de paisaje se obtuvo a través de la integración de las unidades del mapa de clases DST, dentro de cada unidad litogeomorfológica. En este nivel de clasificación se delimitaron áreas más homogéneas que las unidades litogeomorfológicas, en montañas (altas, medias y bajas) y valles intramontanos, para lo cual se emplearon varios atributos de carácter regional, tales como: altitud, pendiente general, índice topográfico de humedad, índice de transporte de sedimentos y las condiciones bioclimáticas de la zona, utilizando como guía la red de drenaje y las curvas de nivel.

El Cuadro 3-9 resume la descripción de los atributos considerados en las 23 unidades identificadas en el sector de estudio, y la Figura 3-16 esquematiza los tipos de paisaje identificados en cada unidad litogeomorfológica, cuya leyenda fue definida con apoyo de SIG, con base en la toponimia de las hojas cartográficas a escala 1:25 000.

**Cuadro 3-9.** Unidades de paisaje presentes en el sector de estudio de la cuenca del río Caramacate.

UL	Paisaje	UG	Clases DST	Altitud (m)			Pend (%)	ITH	ITS	Zona de Vida	Área (ha)	
				Min	Media	Max						
VCñc	MA	1	A, C, E, F, G, H, I, J	671	1028	1413	40	9,1	28	BHP	1007	
		2	A, C, D, E, F, G, I, J	676	949	1282	37	8,8	32	BHP	731	
		3	A, C, D, H, I	530	756	1019	44	8,5	36	BST	182	
		4	A, C, E, H	559	794	1138	46	8,3	37	BST	306	
	MM	5	A, C, H, I, J	495	668	1041	41	8,4	31	BST	619	
		6	A, H, I	559	689	901	28	9,1	22	BST	434	
		7	A, C, H, I, J	509	648	803	38	8,4	28	BST	146	
		8	A, C, H, I, J	522	649	843	38	8,5	28	BST	343	
		9	A, C, H, I, J	598	688	822	28	8,9	19	BST	148	
		10	A, C, H, I, J	506	620	792	35	8,6	24	BST	294	
		MB	11	A, C, H, I	555	622	725	27	9,0	19	BST	97
			12	A, C, H	599	638	702	23	9,5	16	BST	34
Qal	Valle	13	B	505	623	839	14	11,9	15	BST	205	
VCca	MA	14	K, M, O, P, Q, R, S,	679	931	1385	37	8,9	31	BHP	518	
		15	K, M, R, T	575	811	1017	41	8,3	30	BST	195	
		16	K, M, R, S, T	718	897	1110	41	8,6	31	BST	185	
	MM	17	K, M, N, R, S, T	476	734	1083	40	8,8	36	BST	403	
		18	K, M, R, S, T	471	624	821	41	8,4	31	BST	381	
		19	K, M, R, S, T	537	722	885	34	8,5	25	BST	298	
		20	K, R, M	629	721	887	37	8,5	28	BST	43	
	MB	21	K, R, S, T	471	576	682	35	8,4	26	BST	47	
		22	K, R, S, T	490	577	662	32	8,5	23	BST	57	
	Qal	Valle	23	L	469	536	710	12	11,8	17	BST	87

UL: Unidad litogeomorfológica; UG: Identificación de la unidad geomorfológica; VCñc: Metatobas de El Caño-El Chino, VCca: Metalavas de El Carmen, MA: Montaña alta; MM: Montaña media; MB: Montaña baja; Pend: Pendiente del terreno; ITH: Índice topográfico de humedad; ITS: Índice de transporte de sedimentos; BST: Bosque Seco Tropical; BHP: Bosque Húmedo Premontano.