



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE EDAFOLOGÍA

EVALUACIÓN DE MÉTODOS PARA LA CARTOGRAFÍA DIGITAL DE CLASES DE TIERRA CAMPESINAS

GUSTAVO CRUZ CÁRDENAS

T E S I S

PRESENTADA COMO REQUISITO PARCIAL

PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO

2009

La presente tesis, titulada: Evaluación de métodos para la cartografía digital de clases de tierra campesinas, realizada por el alumno: Gustavo Cruz Cárdenas, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTOR EN CIENCIAS
EDAFOLOGÍA

CONSEJO PARTICULAR

CONSEJERO: _____
DR. CARLOS ALBERTO ORTIZ SOLORIO

ASESOR: _____
DR. ENRIQUE OJEDA TREJO

ASESOR: _____
DR. JUAN FELIPE MARTÍNEZ MONTOYA

ASESOR: _____
DR. ATENÓGENES LEOBARDO LICONA VARGAS

ASESOR: _____
DR. ERASTO DOMINGO SOTELO RUIZ

Montecillo, Texcoco, México, noviembre de 2009

EVALUACIÓN DE MÉTODOS PARA LA CARTOGRAFÍA DIGITAL DE CLASES DE TIERRA CAMPESINAS

Gustavo Cruz Cárdenas, Dr.

Colegio de Postgraduados, 2009

El mapeo digital de suelos consiste en emplear algoritmos computacionales y predictores que representan las variables para la generación de mapas de suelos. Existen evidencias que estos mapas son confiables. Sin embargo, para la elaboración de mapas de clases de tierra campesinas, utilizando técnicas digitales se tiene poca información y los mapas producidos son de baja calidad porque han utilizado sólo los valores de reflectancias de las clases de tierras como predictores y algoritmos limitados en cuanto a su configuración. Por lo anterior, en esta investigación se evaluó la calidad de los mapas de clases de tierra campesinas generados en México, en condiciones ambientales contrastantes (árida, templada y tropical), a partir de técnicas empleadas en la cartografía digital de suelos con el objetivo de generar una metodología aplicable en las diferentes regiones. Se emplearon seis clasificadores: árboles de decisión, redes neuronales artificiales, mínima distancia, paralepipedo, máxima verosimilitud y inverso de la distancia. Así, como imágenes de satélite y atributos topográficos como predictores. Los resultados mostraron que la variable que más influyó en aumentar la precisión y la exactitud de los mapas fue la elevación. En cuanto a los algoritmos, el inverso de la distancia fue el mejor en comparación con máxima verosimilitud, redes neuronales artificiales y árboles de decisión. Además, es menos complejo para configurar, no requiere de los predictores y es más eficiente por necesitar menos puntos de muestreo para su configuración adecuada. Este último factor es también muy importante en la cartografía digital de clases de tierra campesinas porque requiere que la información sea lo más confiable posible para que el entrenamiento del algoritmo sea preciso, lo cual se logra con la toma de esta información en campo y a través de entrevistas con los campesinos, este paso es muy relevante y puede establecerse que el muestreo en campo no se debe omitir. No obstante, si se puede emplear un esquema que permita maximizar el muestreo, recomendándose el diseño espacial sistemático.

Palabras clave: Precisión y exactitud de mapas, imágenes de satélite, atributos topográficos, inteligencia artificial, tamaño y diseño de muestro espacial.

EVALUATION OF METHODS FOR DIGITAL MAPPING OF FARMLAND CLASSES

Gustavo Cruz Cárdenas, Dr.

Colegio de Postgraduados, 2009

The digital mapping of soils is to use computer algorithms and predictive variables representing the generation of soil maps. There is evidence that these maps are reliable. However for the mapping of farmland classes using digital techniques, with slight information and maps produced are of low quality because they have used only the reflectance values of the farmland classes and algorithms to predict limited in its configuration. For these reasons, this research evaluated the quality of the maps of farmland land classes generated in Mexico in contrasting environmental conditions (arid, temperate and tropical), from techniques used in digital mapping of soils with the aim of generate a methodology applicable in different regions. Six classifications were used: decision trees, artificial neural networks, minimum distance, parallelepiped, maximum likelihood and inverse distance. Thus, as remote sensing data and topographic attributes as predictors. The results showed that the most influential variable to enhance the precision and accuracy of the maps was the elevation. As for the algorithms, the inverse of the distance was the best compared to maximum likelihood, artificial neural networks and decision trees. It is also a reduced amount of complex to configure, does not require the predictors and is more efficient by requiring a reduced amount of sampling points for suitable configuration. This last factor is also very important in the digital mapping of farmland classes because it requires that the information is as reliable as possible for the training of the algorithm is accurate, which is achieved by making this information in the field and through interviews with farmers, this step is very relevant and can be established that the field sampling should not be ignored. However, if we can use a scheme that maximizes the sample, recommending systematic spatial design.

Keywords: Precision and accuracy of maps, remote sensing data, topographic attributes, artificial intelligence, size and spatial sampling design.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología por la beca otorgada para la realización de mis estudios de Doctorado.

Al Colegio de Postgraduados y a los profesores que ahí laboran por mi formación académica.

Al Dr. Carlos Alberto Ortiz Solorio por su asesoramiento, disponibilidad, dedicación y apoyo constante durante la realización del presente trabajo.

A la Dr. Enrique Ojeda Trejo por sus consejos, orientación y acertadas sugerencias en la realización de ésta investigación.

Al Dr. Juan Felipe Martínez Montoya, Dr. Erasto Sotelo Ruiz y Dr. Atenógenes Leobardo Licona Vargas por la disposición de su tiempo para la revisión de la tesis.

A la señorita Carmen Bojorges Bautista por su apoyo durante toda esta etapa del Doctorado.

A todos mis amigos que de una u otra forma han contribuido en el proceso de mi superación.

DEDICATORIA

A mis padres, Silvana Cárdenas Hernández y Jerónimo Cruz López por su motivación para concluir esta etapa de mi vida. A Tavito y Silvia que palabras faltarían para expresar mi sentir. A mi hermana Sandra y su familia (Jesús y Chuchito) por su cariño y amistad. Por último, a toda mi familia que siempre me han alentado para seguir superándome.

CONTENIDO

	Página
INTRODUCCIÓN GENERAL.....	1
CAPITULO I. COMPUTER ASSISTED CARTOGRAPHY USING TOPOGRAPHICAL ATTRIBUTES: PRECISION AND ACCURACY OF LOCAL SOIL MAPS IN CENTRAL MEXICO.....	22
CAPITULO II. EVALUATION OF FOUR DIGITAL CLASSIFIERS FOR AUTOMATED CARTOGRAPHY OF LOCAL SOIL CLASSES BASED ON REFLECTANCE AND ELEVATION IN MEXICO.....	39
CAPITULO III. DIGITAL MAPPING OF FARMLAND CLASSES.....	62
DISCUSIÓN Y CONCLUSIÓN GENERAL.....	81

INTRODUCCIÓN GENERAL

1. Planteamiento del problema

Los levantamientos de suelos son metodologías para estudiar y describir sistemáticamente al recurso suelo para poder darle un uso óptimo (Ortiz y Cuanalo 1981). Uno de los productos de los levantamientos de suelos son mapas que se hacen a través de la cartografía donde se consideran los factores de formación de suelos para predecir su distribución. Estos mapas se hacen a través de la interpretación de fotografías aéreas e información auxiliar, como mapas de uso de suelo y vegetación, mapas de geología e información de campo (Hudson 1992). Sin embargo, este procedimiento es caro, subjetivo y requiere demasiado tiempo, pero existe otra alternativa que se conoce como el mapeo digital de suelos que consiste en la producción asistida por computadora de mapas de clases o propiedades de suelos (Scull *et al.* 2003).

En el presente siglo se ha incrementado exponencialmente varios estudios sobre el mapeo digital de suelos, los cuales se basan en la inferencia de sus clases o propiedades a través de información espacial de variables ambientales (Carré *et al.* 2007). Existen modelos básicos para inferir la distribución de las clases de suelos basados en la ecuación de formación de suelo de Jenny (1961), donde el suelo está en función del material parental, clima, tiempo, organismos y topografía.

En el primero de ellos es de Zhu *et al.* (2001) y Qi y Zhu (2003) quienes proponen el método SoLIM (modelo de inferencia suelo-paisaje por sus siglas en ingles). Cuya técnica combina el conocimiento local de suelos de los científicos con las herramientas de los sistemas de información geográfica, teoría de conjuntos difusos, redes neuronales artificiales y árboles de decisión. Esta técnica se fundamenta en la ecuación de formación del suelo y el modelo paisajista (Hudson 1992), la cual consiste en tres principales componentes: 1) un modelo de similaridad que representa a los suelos como un continuum; 2) un grupo de técnicas automatizadas para la inferencia de suelos usando el modelo de similaridad; 3) un grupo de procedimientos para derivar los productos de la información de suelos del modelo de similaridad. El modelo requiere como datos de entrada atributos topográficos, información de geología y tipo de vegetación. Los autores mencionan que el SoLIM trabaja en forma adecuada en áreas donde hay diferencias significativas de pendiente.

El segundo modelo es el Prototipo de Teoría Categórica que es una variante del SoLIM (Qi *et al.* 2006), la cual se basa en que una categoría o clase de suelo puede ser representada por su prototipo. A partir de las propiedades del suelo se genera un grupo compuesto de propiedades que ejemplifican la clase, las cuales después servirán como puntos referencia para la identificación de diferentes clases de suelos. Nuevas categorías o clases se identificaran y determinaran la membrecía para establecer el grupo al que pertenecen. El modelo consiste de dos pasos; el primero involucra la adquisición del conocimiento sobre las categorías de suelos establecidas por los científicos y representar el conocimiento en términos de prototipos y grados de membrecía; y el segundo es el prototipo-base que infiere sobre el suelo.

El tercer modelo es el SCORPAN de McBratney *et al.* (2003), el cual se basa en siete factores de predicción: atributos medidos del suelo (*S*), clima (*C*), organismos (*O*), topografía (*R*), material parental (*P*), edad (*A*); espacio y posición geográfica (*N*). Los autores mencionan que su técnica es la formalización y generalización de varios artículos revisados donde se usó el mapeo digital de suelos. La inferencia se hace al igual que en los métodos anteriores a través de la extrapolación de datos de entrenamiento y con algoritmos supervisados el cual está representado por *f* en la Ecuación (1).

$$S_c = f(S, C, O, R, P, A, N) \quad \text{o} \quad S_a = f(S, C, O, R, P, A, N) \quad (1)$$

Con el modelo se pueden inferir clases de suelos (S_c) o variables categóricas, como tipo de drenajes de suelos, clases de fertilidad, series de suelos, entre otros. También, se pueden inferir atributos individuales del suelo (S_a) como contenido de materia orgánica, pH, capacidad de intercambio catiónico y conductividad eléctrica, que en general se refiere a variables continuas.

Otro enfoque está orientado a la cartografía de clases de tierras campesinas, cuya inferencia se obtiene por dos vertientes. La primera es a través del método de fotointerpretación tradicional, desarrollado por Licona *et al.* (1993) que se basa en el hecho de que cada clase de tierra tiene una ubicación espacial sobre el terreno y por lo tanto se expresan sus rasgos característicos en las fotografías aéreas. Estos autores mencionan que es un método inductivo (basado en las técnicas de Buring 1960, Frost 1960 y Peña 1965); que permite identificar clases de tierras en zonas no estudiadas a partir de las marcas características que, las mismas clases de tierra, producen sobre fotografías aéreas de áreas piloto.. La segunda alternativa es el empleo de algoritmos computacionales como mínima distancia, paralepipedo y máxima verosimilitud,

utilizando como variables de entrada o predictores sólo a las bandas de las imágenes de satélite (Cajuste 1991, Martínez 1993, Segura *et al.* 2003: 2004). El procedimiento consiste en tomar puntos de muestreo que representen la distribución de las clases de tierra campesinas. Estas clases se relacionan con los valores de reflectancia de cada una de las bandas. A su vez con los valores de reflectancia se entrenan a los algoritmos y con esto se obtiene un mapa de clases de tierra campesinas a través de procedimientos digitales o conocido como *Cartografía Digital de Clases de Tierra Campesinas*. Con esta última técnica los resultados no han sido del todo satisfactorios ya que la calidad de los mapas que se han generado es menor al 50% y en algunos casos no se ha podido realizar la cartografía digital.

Algunos de los problemas detectados para la Cartografía Digital de Clases de Tierra Campesinas se deben a las limitaciones del software (número mínimo de píxeles para clasificar) y la resolución espacial de las imágenes (60 m) que se han utilizado, lo que no ha permitido una buena representación de la complejidad y el tamaño de las unidades cartográficas de cada clase de tierra.

Por lo anterior, el presente trabajo se desarrolló bajo el supuesto de que si se exploran procedimientos y predictores utilizados en la actualidad para la cartografía digital de clases de suelos, se puede generar mapas de clases de tierra campesina con precisión (porcentaje de puntos de muestreo correctos sobre algún mapa que corresponde a la leyenda del mismo) y exactitud (posición correcta de linderos entre clases) mayor al de los trabajos previos. Algunos de esos procedimientos son, por ejemplo, el empleo de algoritmos de inteligencia artificial como clasificadores, la combinación de la reflectancia con atributos topográficos como predictores y una resolución espacial con tamaño de pixel menor a 30 m. Adicionalmente se establecieron criterios para seleccionar el mejor esquema de muestreo y tamaño de muestra que represente la heterogeneidad de la distribución de las clases de tierras.

2. Objetivos

General

Generar una metodología para la cartografía digital de clases de tierra campesinas en diferentes ambientes.

Específicos

1. Determinar la combinación de atributos topográficos que generen mapas de clases de tierra campesinas de mayor calidad.
2. Evaluar la precisión y exactitud de los mapas de clases de tierra campesinas generados con mínima distancia, paralelepípedo, máxima verosimilitud y redes neuronales artificiales, utilizando como predictores a la reflectancia y elevación.
3. Establecer la mejor configuración de diseño espacial y tamaño de muestra de los sitios de entrenamiento o puntos de muestreo para el mapeo digital de clases de tierra campesinas.

3. Hipótesis

1. Con el uso del modelo digital de elevación como predictor se generan mapas de calidad.
2. Las redes neuronales artificiales y la combinación de la reflectancia y elevación, generan los mapas de clases de tierra campesinas con mayor precisión y exactitud, en comparación con los métodos del paralelepípedo, mínima distancia, máxima verosimilitud y el uso individual de la reflectancia.
3. Con el 30% o más de tamaño de muestra y un diseño sistemático se generan mapas de clases de tierra campesina con precisión y exactitud adecuadas.

4. Revisión de Literatura

Levantamiento de suelos

En los levantamientos de suelos, se describen las características de los suelos en una determinada área, se clasifican los suelos de acuerdo a los estándares de un sistema de clasificación, se marcan sus linderos sobre un mapa y se hacen predicciones acerca de su comportamiento (Soil Survey Division Staff 1993).

El levantamiento de suelos comprende dos etapas principales: la planeación y ejecución (Ortiz y Cuanalo 1981). La planeación involucra la discusión de los propósitos del levantamiento y la naturaleza de la información requerida para suplir la información. Se recomienda que sea un trabajo interdisciplinario y en coordinación con agencias científicas y técnicas que puedan proporcionar información sobre el área. Además, se deben tomar decisiones provisionales

como la escala e intensidad de la cartografía de suelos y los tipos de mapas. La etapa de ejecución se divide en tres: 1) *Pre-campo*, donde se realiza la fotointerpretación, si no se tiene mucha experiencia se sigue que primero se haga un recorrido de campo llevando consigo las fotografías, para relacionar los rasgos fotográficos con el paisaje. El reconocimiento general también se usa para obtener información de la accesibilidad del área, que ayuda a planear el trabajo de campo, principalmente en la selección de las rutas para la cartografía de suelos y para seleccionar sitios de muestreo; 2) *Campo*, se comienza con un segundo reconocimiento del área, el cual es más profundo, y es entonces cuando realmente se pueden diseñar en forma detallada las especificaciones del levantamiento y la planeación inicial puede modificarse. En esta etapa las principales actividades son la cartografía de suelos y obtención directa de datos para evaluar las tierras de acuerdo con el o los propósitos para los cuales el levantamiento fue planeado; y 3) *Post-campo*, se corrige la fotointerpretación, las muestras de los suelos colectadas se analizan en laboratorio y se hace una interpretación integral de los resultados con los de campo.

Los levantamientos de suelos tienen la finalidad de darle un uso inmediato a cada clase, de acuerdo a las necesidades específicas de los usuarios. Por ejemplo, se determina la capacidad de uso o uso potencial del suelo: para terrenos agrícolas, forestales o de agostadero; o la clasificación por aptitud al riego, que es una interpretación de las cualidades que presentan los suelos para su uso bajo irrigación.

McBratney *et al.* (2000) mencionan que los levantamientos suelos convencionales han sido criticados por emplear demasiado el carácter cualitativo. Al respecto, Zhu (1997:2000) menciona que los levantamientos de suelos están limitados por la habilidad para delinear las unidades cartográficas. La delimitación de estas unidades se basa sobre un modelo conceptual discreto, en el cual el fotointerprete produce un mapa tan exacto como sea su propia habilidad en la fotointerpretación. Bajo este modelo los suelos identificados en campo son representados como unidades cartográficas (polígonos). Uno de los problemas asociados con este método es que suelos con poca superficie puede ser ignorado o varios tipos de suelos pueden incluirse en una unidad cartográfica, lo cual obliga a los fotointerpretes a realizar múltiples delineaciones de unidades cartográficas y recorridos de campo para garantizar que se incluyan todos los suelos del área y que no se tenga más de un tipo de suelo en una unidad cartográfica y que los linderos entre clases sean exactos. Sin embargo, por la escala de generalización y la

complejidad en la distribución de los suelos para generar mapas con un alta precisión y exactitud se requieren mapas escalas grandes (1:10 000) y con costos altos para la verificación de cada clase de suelos y la exactitud de los linderos.

Al realizar la interpretación de clases de suelos en fotografías aéreas en la etapa de pre-campo los errores pueden ser los siguientes (Zhu *et al.* 2001): Cambios sutiles en microambientes que no pueden ser fácilmente observados debido a la limitación de percepción visual, especialmente cuando se tratan varios procesos simultáneamente, lo que puede resultar en la identificación de pocos linderos de suelos. Por otra parte, la interpretación visual consume mucho tiempo y se pueden generar errores humanos. Por ejemplo, después de pasar cierto tiempo con el estereoscopio se cometen errores como delimitación equívoca de linderos. Otro error es cuando se trasladan los linderos de la fotografía área al mapa base con lo que disminuye la calidad del mapa, además, del tiempo consumido. Este proceso obliga a los científicos de suelos a pasar más tiempo fotointerpretando y corrigiendo la cartografía a través de trabajo de campo, lo que hace que le dediquen menos tiempo a la identificación y descripción de perfiles de suelos. Finalmente, debido a la formación única de los paisajes y la técnica empleada en cada nuevo levantamiento debe de repetirse todo el proceso. Por lo cual, los costos de los levantamientos son altos y para la mayoría de los países con recursos económicos limitados no son económicamente viables.

Pedometría

En respuesta a estas críticas, tiempo de aprendizaje, complejidad y subjetividad de los procesos de fotointerpretación, errores humanos, altos costos de levantamientos, etc. en los últimos 30 años se han propuesto modelos cuantitativos (McBratney *et al.* 2000). Basados en los avances en la tecnología de adquisición de información imágenes de satélite, disponibilidad de modelos digitales de elevación, creación de bases de datos a nivel mundial, nacional y regional, y avances en software del SIG. Las técnicas que se han utilizado para describir, clasificar y estudiar la distribución espacial de los patrones de suelos en forma más objetiva se han agrupado en un campo emergente de la ciencia del suelo conocido como la Pedometría. De acuerdo a la sociedad de Pedometría (www.pedometrics.org) esta es definida como: “*La aplicación de métodos estadísticos y matemáticos para el modelado cuantitativo de los suelos, con el propósito de analizar su distribución, propiedades y comportamiento* (Figura 1).

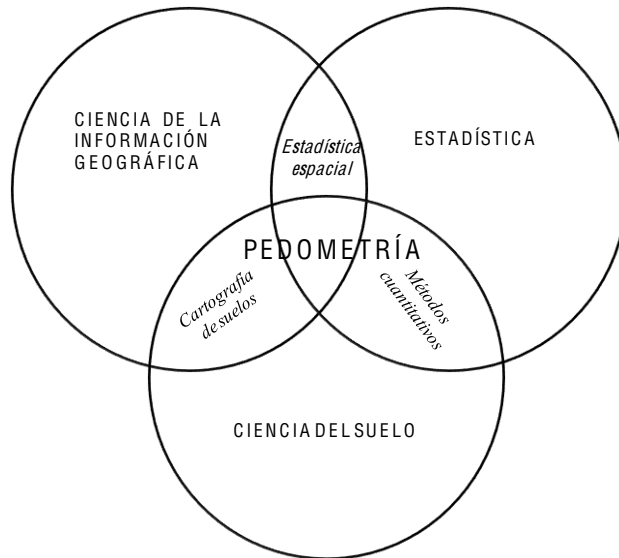


Figura 1. Componentes de la Pedometría.

La Pedometría no es nueva, como métodos matemáticos y estadísticos ha sido aplicada en los estudios de suelos desde 1960. Sin embargo, fue formalmente reconocida como una diferente rama de la pedología en la década pasada. La Figura 2 muestra el incremento en el tiempo en los estudios cualitativos (levantamiento de suelos convencionales), cuantitativos (Pedometría) y la intersección reciente entre las dos áreas del conocimiento por los avances mencionados y la disponibilidad de bases de datos en los últimos 10 años para realizar estudios con información cualitativa y cuantitativa (McBratney *et al.* 2000).

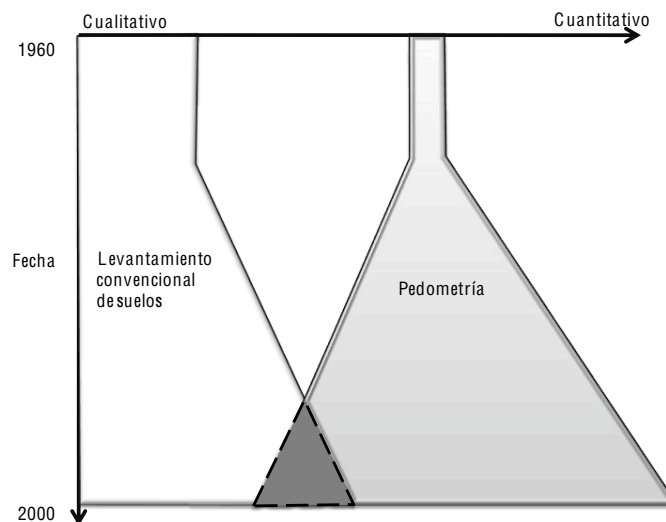


Figura 2. Línea del tiempo del crecimiento del levantamiento de suelos tradicional y Pedometría.

Dentro del campo de la Pedometría se encuentra el Mapeo Digital de Suelos (MDS), donde se utilizan datos de entrada o predictores, atributos topográficos e hidrológicos, imágenes de

satélite, mapas de usos de suelo, índices de vegetación, propiedades de suelo y atributos geológicos principalmente. Además, de puntos de muestreo que relacionan estas variables con los tipos de suelos o sus propiedades. Esta información se configura y se utiliza algún algoritmo computacional o clasificador para generar un mapa de suelos (Figura 3).

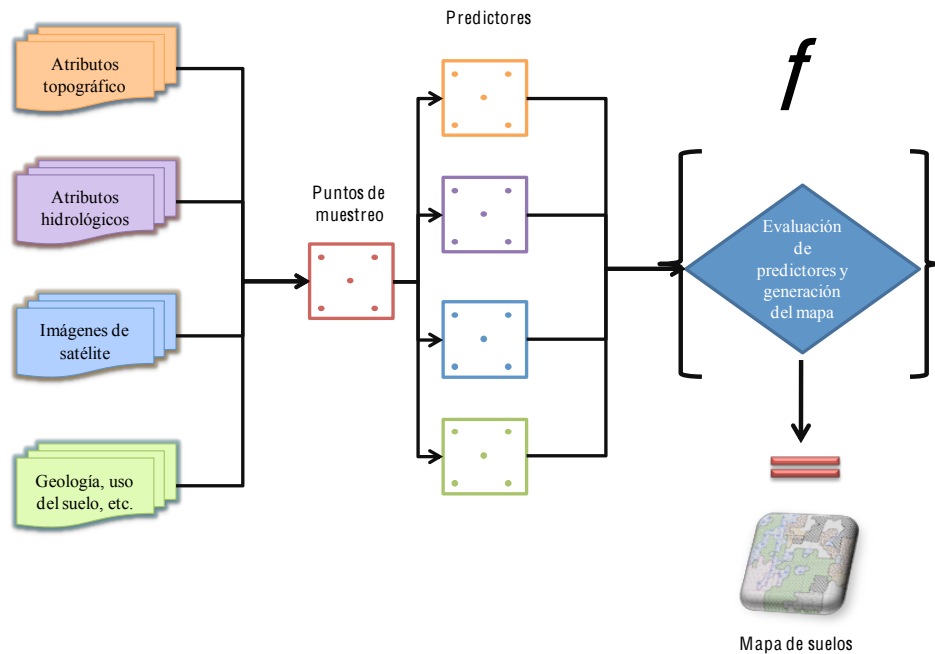


Figura 3. Modelo típico para el mapeo digital de suelos.

Clasificadores

Los clasificadores que se utilizan en el MDS se encuentran en software libre y comerciales los más comúnmente, y utilizados son mínima distancia, paralelepípedo (Ver pág. 38), máxima verosimilitud, árboles de decisiones y redes neuronales artificiales.

El clasificador de *MÁXIMA VEROSIMILITUD* calcula la probabilidad de un píxel de pertenecer al grupo de píxeles que conforman una clase, el píxel es asignado a la clase con más alta probabilidad. Este clasificador se basa en la fórmula de probabilidad bayesiana (Tso y Mather 2009) y se puede reducir a la siguiente Ecuación (2):

$$P(w_j | x_i) \propto P(x_i | w_j) \quad (2)$$

Si se asigna el píxel i a la clase k que maximice la Ecuación (2), el resultado es la solución de máxima verosimilitud. Normalmente, la probabilidad condicional de $P(x_i/w_j)$ se asume que sigue una distribución gaussiana o normal. $P(x_i/w_j)$ se puede expresar como:

$$P(\mathbf{x}_i|w_j) = \frac{1}{\sqrt{2\pi}^\rho \sqrt{|C_j|}} \exp\left(-\frac{1}{2} \times (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \times C_j^{-1} \times (\mathbf{x}_i - \boldsymbol{\mu}_j)\right) \quad (3)$$

Donde C_j es la matriz de covarianza de la clase w_j con dimensiones ρ , $\boldsymbol{\mu}_j$ es el vector de medias de la clase w_j y $|X|$ denota la determinante. El vector de medias y la matriz de covarianza se utilizan para caracterizar cada clase. La importancia de escoger un tamaño de muestra que sea adecuado para proveer un buen estimador de estos parámetros es importante.

El clasificador de *ÁRBOL DE DECISIONES* esta dado por la siguiente ecuación (Quinlan 1988):

$$gain\ ratio(M) = gain(M) / split\ info(M) \quad (4)$$

Donde $gain(M)$ es la información de la ganancia después de que un atributo M se escoge como una prueba para las muestras de entrenamiento, y $split\ info(M)$ es la información que se genera cuando X muestras de entrenamiento son particionadas en n subgrupos. Supongamos que se tiene un grupo de muestras de entrenamiento X y cada muestra X pertenece a la clase k_j donde $j=1,2,\dots,m$; $info(X)$ es el promedio de la cantidad de la información medida en bits necesarios para clasificar una muestra, esto es definido como:

$$Info(X) = - \sum_{j=1}^m \frac{freq(k_j, X)}{|X|} \times \log_2 \left(\frac{freq(k_j, X)}{|X|} \right) \quad (5)$$

Donde $freq(k_j, X)$ es el numero de muestras de el grupo X que pertenece a la clase k_j y $|X|$ es el numero de muestras de entrenamiento en el grupo X . Si S es cualquier grupo de la muestra de entrenamiento, $info(X)$ es también conocida con la entropía del grupo S . Considere que el grupo X es particionado en n subgrupos X_i , ($i=1, 2, \dots, n$) acorde al grupo de prueba M . Entonces la suma ponderada de los subgrupos será el requerimiento de la información esperada $info_M(X)$ para clasificar una muestra. Este es definido como:

$$info_M(X) = \sum_{i=1}^n \frac{|X_i|}{|X|} \times info(X_i) \quad (6)$$

La ganancia de información $gain\ info(X)$ después de aplicar la prueba M es igual a $info(X) - info_M(X)$. Similar a la definición de $info(X)$, pero como en este caso cualquier grupo de muestra S , $split\ info(M)$ puede ser determinado por:

$$split\ info(M) = - \sum_{i=1}^n \frac{|X_i|}{|X|} \times \log_2 \left(\frac{|X_i|}{|X|} \right) \quad (7)$$

El clasificador PERCEPTRÓN MULTICAPA (PML) es una importante *RED NEURONAL ARTIFICIAL* (RNA), aplicado en diversos campos. Un PML consiste de un grupo de unidades de entrada, una o más unidades ocultas y una capa de salida. El PML se ha aplicado junto con

el algoritmo de propagación inversa, la cual consiste en dos fases: paso de hacia adelante y otro hacia atrás. Los valores de salida son comparados con los valores esperados (error), los pesos de la RNA son ajustados de acuerdo al error. El procedimiento se detalla a continuación (Mas y Flores 2008):

1. Inicialización. Se asignan números aleatorios a los pesos sinópticos, usando una distribución uniforme aleatoria.
2. Cálculo hacia adelante. Por cada ejemplo $(x(n), d(n))$ en el grupo de entrenamiento se propaga las señales de entrada a través de la red, capa por capa, como se indica en la siguiente ecuación:

$$v_j^{(l)}(n) = \sum_{i=0}^{m_0} w_{ji}^{(l)} y_i^{(l-1)}(n) \quad y_j^{(l)} = \varphi_j(v_j(n)) \quad (8)$$

3. Propagación inversa. Calcula el error de las señales correspondientes a cada neurona.

$$e_j(n) = d_j(n) - y_j(n) \quad (9)$$

Donde $d_j(n)$ es la salida deseada de la neurona j y $y_j(n)$ es la salida actual de la neurona j . Calcula los gradientes locales de la red, definido por la Ecuación (10):

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(l)} \varphi_j'(v_j^{(l)}(n)) \\ \varphi_j'(v_j^{(l)}(n)) \sum_k \delta_k^{(l+1)}(n) w_{kj}^{(l+1)}(n) \end{cases} \quad (10)$$

neurona j en capa de salida L

neurona j en capa oculta L

Donde $\varphi_j(\cdot)$ denota diferenciación con respecto al argumento. Ajusta los pesos sinópticos de la capa l siguiendo la regla delta generalizada.

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha \left[w_{ji}^{(l)}(n-1) \right] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n) \quad (11)$$

Donde η es el rango de aprendizaje que se usa para actualizar los pesos del internodo y α es la constante del momento. El término momento usa la configuración de pesos previos para determinar la dirección de la búsqueda para el error mínimo global y permitir al algoritmo de propagación inversa a “pick up speed” si un número de pasos consecutivos cambia los pesos en la misma dirección.

Se repiten los pasos 2 y 3 por cada ejemplo de n en el grupo de entrenamiento y por cada época (iteración completa de los grupos de entrenamiento), hasta que la condición de terminación se logra. Típicamente, la condición de la terminación es dado por un número de iteraciones o cuando se da la convergencia, se para la iteración. La configuración de una RNA

entrenada es evaluada por el cálculo de la raíz del error cuadrado medio entre los valores esperados y los valores de activación de los nodos de salida. En el caso de la clasificación, se calcula el porcentaje de muestras correctamente clasificadas del grupo de validación.

Predictores o datos de entrada utilizados en los modelos de inferencia de suelos

Los predictores que se utilizan para alimentar los modelos e inferir las clases de suelos son, entre otros, el modelo digital de elevación, pendiente, aspecto, curvatura del perfil, datos de sensores remotos, los índices de vegetación y de humedad topográfico, y dirección del flujo.

El predictor Modelo Digital de Elevación (MDE) representa la distribución espacial de la elevación del paisaje (Bishop y Minasny 2006). A partir de un MDE se calculan las siguientes variables topográficas (Shary *et al.* 2002, Bishop y Minasny 2006); la Pendiente que es una medición cuantitativa del valor máximo del incremento de elevación, con rango de 0 a 90° y geoméricamente, es el ángulo entre el plano horizontal y la tangencial de la superficie; el Aspecto (o dirección de la pendiente) que es la medida del norte con respecto a las manecillas del reloj y es el azimut de la elevación; la Curvatura del Perfil que es la segunda derivada de la elevación en una dirección perpendicular; el Área de Retención de Humedad es definida por el área que contribuye al flujo a través de un pixel y; el Índice de Humedad Topográfico (IHT) se calculado con la siguiente fórmula:

$$IHT = \ln\left(\frac{A_s}{\tan\beta}\right) \quad (12)$$

Donde A_s es el área de retención de humedad y β es la pendiente. Los valores altos de IHT indican un incremento de condiciones de saturación y estos valores se localizan comúnmente en las partes bajas de las cuencas

La Reflectancia de las clases de suelos es otro predictor que se genera a partir de la percepción remota. Esta se puede definir como un proceso donde la información de un objeto es explorada sin tener contacto físico con él (Johnston 1998). Los sensores, que hacen posible la percepción remota, pueden dividirse en pasivos y activos. Los primeros recogen energía electromagnética de la superficie terrestre, ya sea por reflejo de la luz solar o emitida en el IR térmico según su nivel calorífico. Por su parte, los sensores activos tienen la capacidad de emitir su propia energía para iluminar los rasgos de interés (Soria et al. 1998). Los componentes básicos de un sistema de percepción remota son: la escena, el sensor y el tratamiento de la información, con los siguientes elementos (Lira 1995): fuente de energía, cubierta terrestre, sistema sensor, sistema de recepción-comercialización, interprete y usuario final.

La energía electromagnética, viaja por el espacio como ondas electromagnéticas. Las ondas se diferencian por su longitud, que es la distancia entre dos crestas sucesivas de una onda. El espectro electromagnético consiste en todas las longitudes de onda de la energía electromagnética. Se subdivide en rayos gamma, rayos X, rayos ultravioleta (UV), visible, infrarrojo (IR), microondas y ondas de radio (Fontal *et al.* 2005).

La banda del visible se extiende de 0.4 a 0.7 micrómetros, como se observa en la Figura 4a. Contiguo a la región del rojo se encuentra la banda del infrarrojo cercano (IRC). Esta región aunque es indetectable por el ojo humano, puede ser detectada por sensores artificiales y es muy importante en la percepción remota (Lillesand y Kiefer 1987). Cuando la energía electromagnética, como la del sol, llega a un objeto, hay tres cosas que pueden pasar con ella. Esta puede ser Reflejada, Transmitida o Absorbida por el objeto (Tso y Mather 2009).

Las características de reflectividad del material superficial se pueden cuantificar por la reflectancia espectral, en porcentaje, la cual se obtiene de la división de la energía reflejada en una longitud de onda dada por la energía incidente. La cuantificación de la energía reflejada depende principalmente de tres factores: la magnitud de la energía de incidencia, la rugosidad y el tipo del material (Tso y Mather 2009). Normalmente, los dos primeros factores se toman como constantes. Por lo tanto, sólo se considera el tercer factor.

En la Figura 4b se muestra la reflectancia promedio en la región óptica del espectro para los materiales superficiales ideales: suelo seco desnudo, agua clara y vegetación verde. Se observa cómo los materiales superficiales se pueden separar en términos de su firma espectral. La reflectancia de la vegetación varía considerablemente en todo el espectro electromagnético. La más baja reflectancia se presenta en 0.4 μm (banda azul), mientras que la más alta reflectancia se encuentra alrededor de las bandas IRC y parte IRM. La firma espectral del suelo desnudo, en contraste, muestra que la reflectancia se incrementa ligeramente con la longitud de onda. Esta reflectancia en la banda visible es más grande que la de vegetación, mientras que en las bandas del IR y parte del IFM es menor que la de la vegetación. En las longitudes de onda de más de 1.4 μm es superior nuevamente. La alta reflectancia de la vegetación en la banda del IRC, combinado con su baja reflectancia, banda roja, se usa para las construcciones de índices, como el de vegetación normalizada (IDVN).

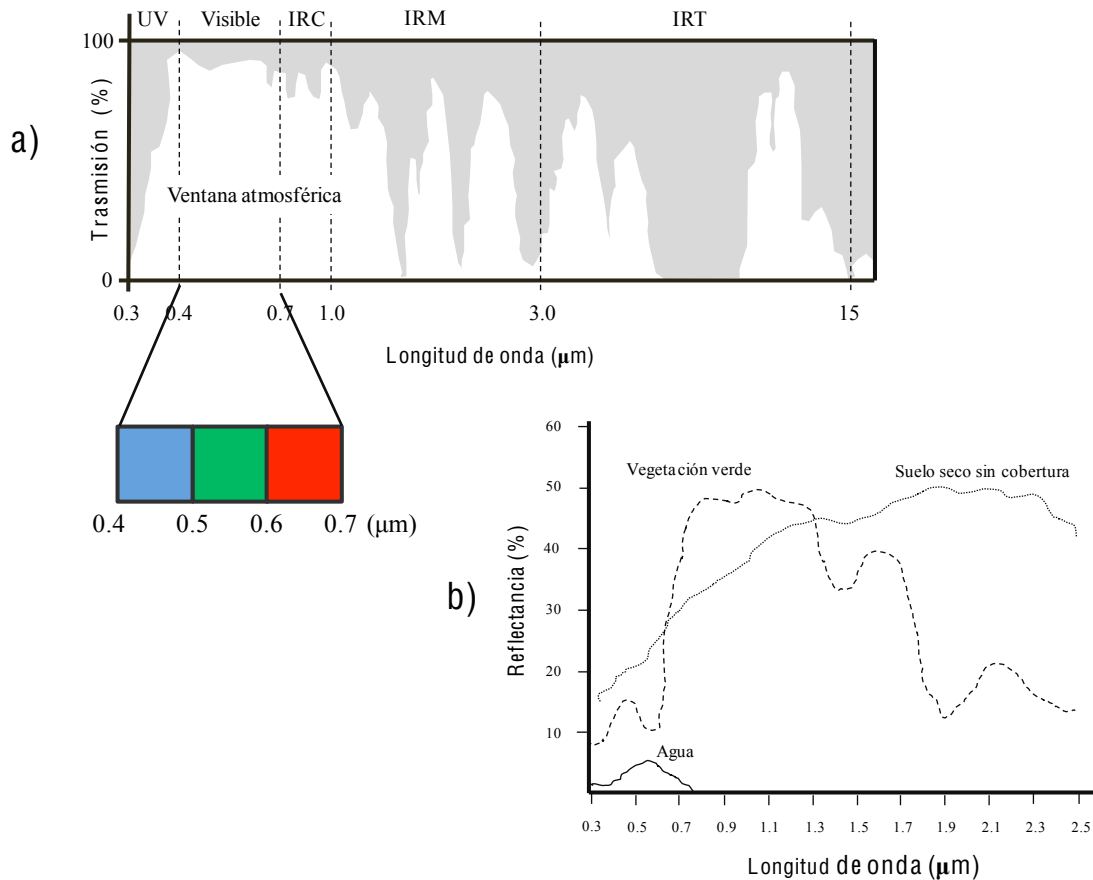


Figura 4. Ventana atmosférica (a). Reflectancia típica de tres materiales (b). IRC= Infrarrojo cercano medio; IRM=Infrarrojo medio; IRT=Infrarrojo térmico.

Ventajas del Mapeo Digital de Suelos

De acuerdo con Zhu *et al.* (2001) y McBratney *et al.* (2003), las ventajas del MDS, con base en los anteriores clasificadores y predictores, en comparación con las técnicas tradicionales de levantamientos de suelos son: a) *Mapeo consistente*, la automatización permite aplicar en forma consistente el modelo de suelo-paisaje en toda el área (Hengl y Rossiter 2003); b) *Actualización rápida de los levantamientos de suelos*, bajo este enfoque se pueden producir nuevas versiones de mapas de suelos en forma rápida, ya que pueden ser hechos en horas o días en vez de meses o años. La habilidad de actualizar rápidamente bases de datos espaciales de los mapas de suelos permite a los edafólogos procesar la información espacial con la tecnología actual y entender mejor los suelos (Behrens y Scholten 2006); c) *Reducir costos y tiempo*, se libera a los edafólogos de la demandante tarea de realizar la cartografía tradicional permitiendo que se enfoquen a estudiar y descubrir las relaciones de suelo-ambiente (Kozlova

y Konyushkova 2009); d) *Continuidad del conocimiento*, cuando los edafólogos se retiran, su experiencia también se va con ellos. Obviamente, se desea que su experiencia pueda ser retenida para mantener la continuidad del conocimiento sobre la relación suelo-ambiente entre las generaciones de edafólogos. En la automatización del mapeo este conocimiento puede ser guardado por medio de algoritmos o modelos que puedan ser reproducidos posteriormente y puedan ser mejorados, lo cual permite que nuevos edafólogos acorten el tiempo de su aprendizaje (Zhu y Band 1994). e) *Generación de productos digitales*, los productos del MDS pueden ser usados directamente en los sistemas de información geográfica sin tener que pasar por procesos de digitalización. Este proceso es caro y puede reducir la calidad de los productos finales, ya que es posible introducir el error en la digitalización y en el llenado de la base de datos (Zhu 1997).

En el presente trabajo se considera que todas estas ventajas del MDS pueden ser aplicadas al mapeo digital de clases de tierras campesinas, ya que en cada clase de tierra se manifiesta la relación suelo-ambiente, percibida por los productores locales y estudiada por la etnopedología.

Cartografía de clases de tierra campesinas

Williams y Ortiz (1981) mencionan que la etnopedología comprende los siguientes tópicos: la percepción popular de los procesos y propiedades de los suelos; la cartografía local y su taxonomía; las explicaciones y dinámica de sus propiedades; el manejo local del suelo; la percepción popular de la relación entre los dominios de plantas y suelos; y la valoración de la percepción popular de este recurso en prácticas agrícolas y otras actividades.

En lo que respecta a la cartografía de las clases de tierras Ortiz-Solorio *et al.* (2005) señalaron que existe evidencia en documentos prehispánicos, como en el código de Vergara y Santa María, que pueden considerarse como los primeros mapas de suelos en el mundo. La conservación de tal conocimiento, entre generaciones de campesinos, ha hecho posible generar mapas de clases de tierra campesinas en comunidades y ejidos en la actualidad. Además, de la existencia de una terminología tradicional que hace referencia a unidades cartográficas y su distribución sobre una parcela o grupo de parcelas. Se ha demostrado la capacidad de los campesinos para localizar los linderos de las clases de tierras sobre materiales cartográficos como mapas parcelarios, mapas topográficos o fotografías aéreas (lo cual constituye el mapa base del método) debido a que poseen una exactitud mental para diferenciar la variación

espacial de esos linderos acorde con el relieve, la vegetación, el uso del suelo y mesoclima (Barrera-Bassols *et al.* 2006). Con base a los puntos anteriores se pueden generar mapas detallados de escala 1:10 000. En México el primer mapa de clases de tierra campesinas fue generado por Pájaro y Ortiz (1987) en el ejido de San Salvador Atenco, México. Desde esa fecha, se han generado mapas de clases de tierras en 33 ejidos, como parte de la investigación de Tesis de estudiantes (14), por organizaciones agrícolas (12) y como prácticas de cursos de etnopedología (7) (Ortiz-Solorio y Gutiérrez-Castorena 2001).

Los productos que se generan con esta técnica no sólo son mapas donde se representa la distribución y comportamiento de las clases de tierra campesinas, sino también se obtiene otro tipo de información ya que el conocimiento del campesino sobre su entorno es amplio. Por ejemplo, Cruz *et al.* (1998) mencionan que la clasificación campesina de tierras se considera como una alternativa viable para generar, transferir y adoptar tecnología agrícola para los productores. Debido a la estrecha relación que existe entre las clases de tierras y la tecnología de producción de los cultivos, es decir, con el rendimiento. Sánchez *et al.* (2002) señalaron que con la clasificación local de tierras también es posible identificar a los factores limitantes para la producción de la caña de azúcar y a partir de ellos tomar iniciativas para resolverlas, debido a la estrecha relación que existe entre las clases de tierras, la tecnología de producción de los cultivos y el rendimiento. Por su parte, Cruz *et al.* (2008) encontraron claras evidencias de que el conocimiento local que se tiene sobre el recurso suelo es detallado, porque no sólo se reconoce la capa superficial, como en otros estudios sino también la capa subsuperficial, lo que le permite al citricultor del ejido Pueblillo, Papantla, Ver., distinguir entre las tres clases de tierras cuál es la más apta para la producción de naranja.

Por otra parte, Ryder (2003) menciona que el conocimiento local de las tierras sirve para calibrar las técnicas de evaluación de recursos naturales y relacionarlos con conceptos científicos como erosión de suelo, taxonomía de suelos, factores edáficos y climáticos que influyen en la producción agrícola y la sustentabilidad de la tierra. Al respecto, Cools *et al.* (2003) señalan que el conocimiento de sustentabilidad de tierras que tienen los campesinos y el que tienen los científicos es diferente, ya que partir del conocimiento campesino estos investigadores obtuvieron un mapa de 63 unidades de tierra. Mientras con el procedimiento técnico se obtuvieron nueve.

Por lo tanto, si se toma como punto de partida esta información y se almacena en un sistema de información geográfica para alimentar a los clasificadores, ayudaría a delimitar aéreas homogéneas entre sí y diferentes entre ellas. Los mapas que se generarían serían de mayor calidad en cuanto a la predicción de las unidades cartográficas y a la información que ellas representan en comparación con los mapas de suelos, tal como lo demostró Lleverino *et al.* (2000) al evaluar la precisión y exactitud de un mapa de clases de tierra campesinas y un mapa de suelos generado en forma tradicional. El primer mapa tuvo 72% y 73% más de precisión y exactitud respectivamente, en comparación con el segundo mapa. Sin embargo, los mapas de clases de tierra campesinas que se han generado mediante el mapeo digital han sido de baja calidad (Ortiz 1999), algunos problemas que se han presentado se describen a continuación. Martínez (1993) menciona que no pudo identificar las clases de tierra campesinas de un ambiente árido porque la resolución de los píxeles de las imágenes de satélite que empleó fue baja (60 m), lo que dificultó la identificación de las clases de tierras, debido a que las parcelas de los productores son menores al tamaño del píxel. Además, el clasificador que empleó tuvo deficiencias debido a que utilizaba un número mínimo de píxeles para realizar la discriminación de las clases de tierra campesinas. Por su parte, Cajuste (1991) en un ambiente templado pudo generar un mapa de clases de tierra campesinas, aunque en su trabajo no se menciona la calidad del mapa. Sin embargo, esta se determinó comparando el mapa real (de campo) con el que él generó digitalmente en su investigación, la precisión de su mapa es de 52%.

A partir de lo anterior se observa que el mapeo digital de las clases de tierra campesinas (MDCTC) no ha sido del todo eficiente, en trabajos pasados, porque sólo se empleó la reflectancia como único predictor y con algoritmos poco eficaces. Al respecto, Ortiz (1999) señala que las clases de tierra reconocidas por los campesinos en campo constituyen patrones, es decir, no se identifican con un solo atributo, sino más bien están constituidas por un grupo de atributos. Al utilizar sólo los valores de reflectancia provocan que varias clases de tierras se confundan, es decir, se agrupan varias clases de tierra en una sola. Sin embargo, la metodología puede mejorarse si se emplean los predictores, diseño espacial de puntos de muestreo y algoritmos que se utilizan en el mapeo digital de suelos y con esto crear mapas de clases de tierra campesinas de buena calidad. Por ejemplo, con el uso de clasificadores de inteligencia artificial como redes neuronales artificiales y árboles de decisión en

investigaciones de mapeo digital de suelos, la confiabilidad de los mapas que se obtienen es mayor al 80%. En cuanto a las variables de entrada a parte de la reflectancia, el modelo digital de elevación y los atributos topográficos que se derivan de este mejoran la inferencia de las clases de suelos con lo cual se aumenta la precisión de los mapas en más del 20%. Sin embargo, la calidad de los resultados que se han obtenido depende fundamentalmente de los puntos de muestreo que se emplean para entrenar al clasificador, es decir, que también estos puntos puedan representar la heterogeneidad de las clases a inferir. Por lo tanto, una parte fundamental en el mapeo digital es la visita a campo para recolectar información que represente lo más cercano posible la distribución de las clases.

Con base en lo anterior, en la presente investigación se aplicaron los principios del MDS para el MDCTC en tres aspectos fundamentales que son: utilizar mayor número de predictores que se relacionen con la distribución de las clases de tierra campesinas, clasificadores sofisticados y configuración de un tamaño y diseño de puntos de muestreo ideal para el entrenamiento de los algoritmos. Los resultados se reportan en tres artículos que corresponden a cada uno de los temas tratados.

5. Referencias

- BARRERA-BASSOLS, N., J.A. ZINCK, J. A. and VAN RANST, E., 2006, Local soil classification and comparison of indigenous and technical soil maps in a Mesoamerican community using spatial analysis. *Geoderma*, 135, pp. 140-162.
- BEHRENS, T. and SCHOLTEN, T. 2006, Digital soil mapping in Germany-a review. *Journal of Plant Nutrition and Soil Science*, 169, pp. 434-443.
- BISHOP, A.T.F. and MINASNY, B., 2006, Digital Soil-Terrain Modeling: The Predictive Potential and Uncertainty, In: S. Grunwald (Ed) *Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics*, Taylor and Francis FL, 185-213.
- BUI, N.E., 2004, Soil survey as a knowledge system. *Geoderma*, 120, pp. 17-26.
- CAJUSTE, B.L., 1991, *Procedimiento para la Cartografía de Clases de Tierras Campesinas mediante Sensores Remotos*. Tesis de Maestría. Colegio de Postgraduados. Montecillo, México.

- CARRÉ, F, MCBRATNEY, A., MAYR, T. and MONTANARELLA, L., 2007, Digital soil assessments: Beyond DSM. *Geoderma*, 142, pp. 69-79.
- COOLS, N., DE PAUW, E. and DECKERS, J. 2003. Towards an integration of conventional land evaluation methods and farmers' soil suitability assessment: a case study in northwestern Syria. *Agriculture, Ecosystems and Environment*, 95, pp. 327–342.
- CRUZ, B.R., VOLKE, H.V., TURRENT, F.A. y PÁJARO, H.D. 1998, Clasificación de Tierras Campesina para la generación y transferencia de tecnología agrícola entre pequeños productores: caso del maíz en la Región Central de Veracruz. *Terra*, 16, pp. 1-10.
- CRUZ, C.G., ORTIZ, S.C.A., GUTIÉRREZ, C. MA. DEL .C. y VILLEGAS, M.A., 2008, Las clases de tierras citricolas del ejido Pueblillo, Papantla, Veracruz. *Terra Latinoamericana*, 26, pp. 11-19.
- FONTAL B., SUAREZ, T., REYES, M.F., CONTRERAS R. y ROMERO, I., 2005, *El Espectro Electromagnético y sus aplicaciones*. Universidad de Los Andes. Mérida, Venezuela.
- HENGL, T. and ROSSIETER, G.D. 2003, Supervised landform to enhance and replace photo-interpretation in semi-detailed soil survey. *Soil Science Society of America Journal*, 67, pp. 1810-1822.
- HUDSON, D.B., 1992, The soil survey as paradigm-based science. *Soil Science Society of America Journal*, 56, pp. 836-841.
- JENNY, H. 1961. *E.W. Hilgard and the birth of modern soil science*. Farallo Publication, Berkeley, CA.
- JOHNSTON, A.C., 1998, *Geographic Information Systems in Ecology*. Blackwell Science Ltd. USA.
- Kozlova, N.D. and Konyushkova, V. M. 2009, State of the Art and Prospects of Digital Soil Mapping: The Results of the Third Global Workshop (Logan, USA, 2008). *Eurasian Soil Science*, 42, pp. 699–702.
- LICONA, V.A., ORTIZ, S.C.A. y PÁJARO, H.D., 1993, El uso de la fotointerpretación en la cartografía de clases de tierra campesina. *Revista Geografía Agrícola*, 18, pp. 85-93.
- LILLEN SAND, M.F. and KIEFFER, W.F., 1987, *Remote sensing and image ineterpretation*. Second Ed. John Wiley and Sons. University of Wisconsin, Madison, USA.
- LIRA, J., 1995, *La Percepción Remota: nuestros ojos desde el espacio*. Cuarta reimpresión. Fondo de Cultura Económica. México, D. F.

- LLEVERINO, G. E., ORTIZ, S.C.A. y GUTIÉRREZ C., Ma. del C., 2000, Calidad de los -mapas de suelos en el ejido de Atenco, estado de México. *Terra*, 18, pp. 103-113.
- MARTÍNEZ, M.J.F., 1993, *El uso de fotografías aéreas e imágenes de satélite en la cartografía de tierras*. Tesis de Maestría. Colegio de Postgraduados. Montecillo, México.
- MAS, F.J. and FLORES, J.J., 2008, The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, 29, pp. 617-663.
- MCBRATNEY, B.A., MENDONÇA-SANTOS, L. M. and MINASNY, B., 2003, On digital soil mapping. *Geoderma*, 117, pp. 3-52.
- MCBRATNEY, B.A., ODEH, A.I. O., BISHOP, A.T.F., DUNBAR, S.M. and SHATAR, M.T., 2000, An overview of pedometric techniques for use in soil survey. *Geoderma*, 97, pp. 293–327.
- ORTIZ, S.C.A. 1999, *Los levantamientos etnoedafológicos*. Tesis de Doctorado. Colegio de Postgraduados, Montecillo, México.
- ORTIZ, S.C.A. y CUANALO, DE LA C.H., 1981, *Introducción a los levantamientos de suelos*. Colegio de Postgraduados. Chapingo, México.
- ORTIZ-SOLORIO, C. A. y GUTIÉRREZ-CASTORENA, MA. DEL C., 2001, Mexican Ethnopedology: the Experience in Soil Mapping. Science and Tradition: *Roots and Wings for Development*, pp. 107-136.
- ORTIZ-SOLORIO, C.A., GUTIÉRREZ-CASTORENA, MA. DEL C., LICONA-VARGAS, A.L. and SÁNCHEZ-GUZMÁN, P., 2005, Contemporary influence of indigenous soil (land) classification in Mexico. *Eurasian Soil Science*, 38, pp. S89-S94.
- PÁJARO, H. D. y ORTIZ, S. C. A., 1987, *El levantamiento de suelos y su relación con cartografía y clasificación de clases de tierras campesinas*. Centro de edafología, Colegio de Postgraduados.
- QI, F. and ZHU, X.A., 2003, Knowledge discovery form soil maps using inductive learning. *International Journal of Geographical Information Science*, 17, pp. 771-795.
- QI, F., XING-ZHU, A., HARROWER, M. and BURT, J., 2006, Fuzzy soil mapping based on prototype category theory. *Geoderma*, 136, pp. 774-787.
- QUINLAN, R. J., 1988, Simplifying decision trees, In: B Gaines and J Boose (Eds) *Knowledge Acquisition for Knowledge-Based System*, Academic Press CA, 241-254.

- RYDER, R., 2003, Local soil knowledge and site suitability evaluation in the Dominican Republic. *Geoderma*, 111, pp. 289–305.
- SÁNCHEZ, G.P., ORTIZ, S.C.A., GUTIÉRREZ, C. MA. DEL C. y GÓMEZ, D.J.D. 2002, Clasificación campesina de tierras y su relación con la producción de caña de azúcar en el sur de Veracruz. *Terra*, 20, pp. 359- 369.
- SCULL, P., FRANKLIN, J., CHADWICK, O.A. and MCARTHUR, D., 2003, Predictive soil mapping: a review. *Progress in Physical Geography*, 27, pp. 171-197.
- SEGURA, C.M.A., ORTIZ, S.C.A. y GUTIÉRREZ, C. MA. DEL C., 2003, Localización de suelos de humedad residual a partir de imágenes de satélite: Clasificación automática supervisada de la imagen. *Terra*, 21, pp. 149-156.
- SEGURA, C.M.A., ORTIZ, S.C.A. y GUTIÉRREZ, C. MA. DEL C., 2004, Localización de suelos de humedad residual a partir de imágenes de satélite: Parte 2. Factores que influyen en su reflectancia y clasificación supervisada con los procedimientos: Mínima distancia y máxima verisimilitud. *Terra*, 22, pp. 135-142.
- SHARY, A.P., SHARAYA, S.L. and MITUSOV, V.A., 2002, Fundamental quantitative methods of land surface analysis. *Geoderma*, 107, pp. 1-32.
- SOIL SURVEY DIVISION STAFF., 1993, *Soil survey manual*. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.
- SORIA, R.J., ORTIZ, S.C.A., ISLAS, G.F. y VOLKE, H.V. 1998. *Sensores Remotos; principios y aplicaciones en la evaluación de los recursos naturales, experiencia en México*. S.M.S.S. Publicación especial No. 7.
- TSO, B. and MATHER, M.P., 2009, *Classification methods for remotely sensed data*. Second Edition, CRC Press, Boca Raton, FL.
- WILLIAMS, B.J. and ORTIZ, S.C.A., 1981, Middle American Folks Soil Taxonomy. *Annals Association of American Geographers*, 71, pp. 335-358.
- ZHU, A.X. 1997, A similarity model for representing soil spatial information. *Geoderma*, 77, pp. 217-242.
- ZHU, A.X. and BAND, E.L. 1994, A knowledge-based approach to data integration for soil mapping. *Canadian Journal of Remote Sensing*, 20, pp. 408-418.
- ZHU, A.X., 1997, A similarity model for representing soil spatial information. *Geoderma*, 77, pp. 217–242.

- ZHU, A.X., 2000, Mapping soil landscape as spatial continua: The neural network approach. *Water Resources Research*, 36, pp. 663–677.
- ZHU, X.A., HUDSON, B., BURT, J., LUBICH, K. and SIMONSON, D., 2001, Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal*, 65, pp. 1463-1472.

CAPITULO I
COMPUTER ASSISTED CARTOGRAPHY USING TOPOGRAPHICAL
ATTRIBUTES: PRECISION AND ACCURACY OF LOCAL SOIL MAPS IN
CENTRAL MEXICO

COMPUTER ASSISTED CARTOGRAPHY USING TOPOGRAPHICAL
ATTRIBUTES: PRECISION AND ACCURACY OF LOCAL SOIL MAPS IN
CENTRAL MEXICO¹

G. Cruz-Cárdenas^a, C.A. Ortiz-Solorio^a, E. Ojeda-Trejo^a, J.F. Martínez-Montoya^b, E.D.
Sotelo-Ruiz^c, and A.L. Licona-Vargas^d

^a*Colegio de Postgraduados, Campus Montecillo, Carretera México-Texcoco km 36.5, 56230, Texcoco, México, México.* ^b*Colegio de Postgraduados, Campus San Luis Potosí, Iturbide 73, 78620, Salinas de Hidalgo, Salinas, SLP, México.* ^c*Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Campo Experimental Zinacantepec, Carretera Zacatepec-Galeana km 0.5, 62780, Zinacantepec, México, México.* ^d*Universidad Autónoma Chapingo, CRUO, km 3 Carretera Huatusco-Jalapa, 94100, Huatusco, Veracruz, México.*

Resumen

El objetivo del presente estudio fue realizar la cartografía asistida por computadora (CAC) de clases de suelos locales, empleando varias combinaciones de atributos topográficos producidos en SIG (modelo numérico del terreno, aspecto, pendiente y curvatura del perfil). Se usó un árbol de decisiones para determinar el número de atributos topográficos necesarios para producir la cartografía digital de clases de suelos locales. Se evaluó la calidad de los mapas producidos por CAC con base a su precisión y exactitud. La evaluación se realizó en la zona central de México, se utilizaron tres mapas de clases de suelos locales en condiciones de clima y paisaje contrastantes: desierto, templado y tropical. Los resultados muestran que en las tres áreas de estudio, los mapas producidos por CAC, la elevación es el rasgo topográfico con el que generaron los mapas de mayor precisión (72%). Sin embargo, la exactitud de los mapas (posición de linderos) fue menor a 33%. Esto significa que la exactitud de los mapas producidos aun es baja y estudios posteriores se debe enfocarse en este parámetro para mejorarlo.

Palabras clave: Árbol de decisiones, modelo numérico del terreno, calidad de mapas.

¹ Submitted in *Computers and Geosciences*, in review.

Abstract

The aim of this study was to carry out computer assisted cartography (CAC) of local soil classes, using several combinations of topographical attributes produced in GIS (terrain numerical model, aspect, slope, and profile curvature). Decision tree was used to find the number of topographical attributes necessary to produce digital cartography of local soil classes. The maps produced by CAC were evaluated using the parameter of quality of the map defined as its precision and exactitude. The evaluation was carried out in Central Mexico using three maps of local soil classes with contrasting landscape and climatic conditions: desert, temperate, and tropical. The results found showed that for the three areas studied, the maps produced by CAC using the parameter of elevation as topographical feature had the higher precision (72%). However the accuracy of the maps (location of boundaries) was low 33%. This means that the accuracy of the maps produced still remains low and further research is necessary to improve this indicator.

Key words: Decision tree, terrain numerical model, map quality.

1. Introduction

Digital Soil Mapping (DSM) in soil science (McBratney *et al.* 2003) or predictive soil mapping (Scull *et al.* 2003) is the computer-assisted production of digital maps of soil type and soil properties. DSM can rely upon, but is distinct from, soil mapping involving manual delineation of soil boundaries by field soil scientists and Farmers. Digitized and georeferenced soil survey information does not become DSM until the GIS layer is used to derive other soil related information within a GIS or similar information software application. DSM make use of technological advances, including GPS receivers, field scanners, and remote sensing, and computational advances, including geostatistical interpolation and inference algorithms, GIS, digital elevation model, and data mining. Semi-automated techniques and technologies are used to acquire process and visualize information on soils and auxiliary information, so that the end result is obtained at cheaper costs (Hengl and Rossiter 2003). Products are commonly assessed for the precision and uncertainty and can be more easily updated when new information comes available (McBratney *et al.* 2003, Scull *et al.* 2003).

Recently several methods for DSM have been developed for example: SoLIM (Zhu *et al.* 2001, Qi and Zhu 2003) or the scorpan-SSPFe (soil spatial prediction function with spatially

autocorrelated errors) (McBratney *et al.* 2003). These model are based the soil variation is associated with the five formation factors: climate, organisms, topography, parent material, and time, and rely mainly upon topographical features.

The most used topographical features in soil computer assisted cartography (CAC) are elevation, slope, aspect, and profile curvature. The precision range of the generated maps goes from 50% to 84% (Lagacherie and Holmes 1997, Dobos *et al.* 2001, Moran and Bui, 2002, Hengl and Rossiter 2003, Peng *et al.* 2003, Qi and Zhu 2003, Schmidt and Hewitt 2004, Giasson *et al.* 2006, Smith *et al.* 2006, Qi *et al.* 2006, Ziadat 2007). These maps produced by DSM are generated from scientific data by the extrapolation of soil properties, but do not take in consideration soil local knowledge. Alternatively Local soil classification has been the central focus of studies undertaken worldwide to understand farmer local knowledge about their soils and the majority of studies focus on the cartography of local soil classes (Barrera-Bassols and Zink 2000, Niemeijer and Mazzucato 2003, Ortiz-Solorio *et al.* 2005). The Local soil maps have been generated based on local knowledge from farmer there are more soil-landscape relationship than scientific procedures (Lleverino *et al.* 2002, Krasilnikov and Tabor 2003). The local soil classes are built by the identification of cartographic units. These units are related by direct observations of the farmers of the relation soil-landscape moreover the maps have a high correlation with features considered important for the farmers such as working, yield, fruit quality, vegetation, and others (Ortiz-Solorio *et al.* 2001, Cruz *et al.* 2008).

The quality of the soil maps are estimated taking in consideration its precision and accuracy (Brown 1988). Precision refers to the dispersion presented by the properties of the soil on the central concept or typical profile in the case of the cartographic unit. Accuracy represents the correct location of the soil boundaries.

The objective of this study was to evaluate the precision and accuracy of the maps produced with CAC in three areas of Mexico with contrasting climate, and landscapes. Maps for these three areas were produced by CAC using the following topographical attributes: elevation slope, aspect, and profile curvature, commonly these features have been produced combining all of them, in this research thee features were considered also individually.

2. Materials and methods

2.1 *Study regions*

In this research three contrasting regions with previous soil local study were selected. One is the arid region or San Luis Potosi, which is located between the meridians 101° and 101°20' West longitude (WL) and the parallels 22°16' and 22°34' of North latitude (NL); altitude oscillates from 2217 to 2244 m; the climate according to Garcia (1988) is arid or dry temperate (BS); the main rock types are calcareous, shale, conglomerates, gritty with some blotches of igneous and metamorphic rocks (INEGI 1988). The second is located in the temperate region or state of Mexico, extends from 98°39' to 99° WL, and 19°22' and 19°32' NL; altitude from 1770 to 2000 m; the climate is sub-humid tempered (C(w)); geology is composed of basic vulcanoclastic extrusive and acid igneous rocks (Garcia 1988, INEGI 1988). The last study area, was in the tropical region or Veracruz, situated between the 97°10' and 97°30' WL, and 20°10' and 20°40' NL, and altitude ranges from 50 to 300 m; the climate is sub-humid warm (Garcia 1988); litology comprehends calcareous shale, alluvial silts, and small portions of tuff (INEGI 1988).

2.2 *Maps of land classes and topographical attributes*

The maps of local soil classes or ground truth (Figure 1) are from Martinez *et al.* (2003), arid region, Pajaro and Ortiz (1987), temperate region, and Cruz *et al.* (2008), tropical region, and were generated with the methodology of Ortiz *et al.* (1990). The maps were digitalized in ArcView 8.1 of ESRI™ (Shaner and Wrightsell 2000) and imported to IDRISI™ (Eastman 2006). The terrain numerical models (TNM) were extracted from the download system of the Continuo de Elevaciones Mexicano (INEGI 2007) and were resized to each study area. The pixel size was 28.5x28.5 m, as suggested by Smith *et al.* (2006), to allow a better precision for automated cartography of soil maps are those created equidistant from 6 to 20 m, and pixels from 24 to 48 m. From the TNM were used to extract slope, aspect, and profile curvature were generated in IDRISI™.

2.3 *Training sites, classifier and input data*

Polygons were digitalized onto the land class maps reported to be used as training sites. The IDRISI™ decision tree was used as a classifier. It begins from a root and the training sites.

Pixels are separated through binary rules. If the separated pixels are of a single class, they are combined to form a leaf. If the separation contains pixels from different classes, an internode is fixed and the separation process continues until the classification is finished (Quinlan 1988). The combination of the topographical attributes used in the CAC is presented in Table 1. To obtain maps 6, 7, 8, and 9 a single topographical feature was used at a time, TNM, slope, aspect, and profile curvature, respectively. The combination of the mentioned attributes was used in the three study regions.

2.4 Evaluation of precision and accuracy

For the evaluation of the precision the method proposed by Congalton (1988) was used. A representative sample for each of area whole of the maps of local soil classes (ground truth) was generated with the 1% of the total pixels. The spatial arrangement of the sample was stratified random as established by François *et al.* (2003). Those maps of sample points were cross with the 27 maps created using CAC, and with the local soil maps of the three studied areas. The results were 27 grids of the sample points (1%) with the information of the maps generated by CAC from the topographical, and three grids of sample points for the local soil classes. By the overlying of the two groups of maps (CAC and local soils) in IDRISI with the module ERRMAT confusion matrixes for each of the maps were built and the total precision for each map was obtained (Congalton 1988).

The accuracy of the maps was evaluated according to Burgess and Webster (1984) as follows: first, a grid of one-centimeter per side grid was traced at scale 1:50,000; then, the distance between two consecutive boundaries in a line was established in meters; then, the distance in the two directions North-South and East-West was measured and multiplied by 0.52 to generate the optimum distance; then the total longitude of the boundaries was divided by the mean optimum distance to produce the optimum sample size. The number of successes points on boundaries was registered.

2.5 Comparison of precision and accuracy

The binomial proportions mean test was applied to contrast the total precision and the accuracy of the maps using MINITAB™ (2003).

3. Results and discussion

3.1 *Evaluation of Precision*

The maps with the higher precision in the arid, temperate, and tropical regions were the maps 1, 2, 3 4, and 6, all of these included the TNM (Figure 2). The precisions of maps with four attributes for all areas were not better according to the statistical test to the maps 6 with TNM. This means that the maps generated with CAC can be produced using only the TNM as input data. On the other hand, the quality of the maps varies depending on the region: for the arid and temperate region the precision are less than 45%, and for the tropical region, the precision is higher 70%. These results showed that TNM is the attribute that has the high relation for the delimitation of a map with local soil classes in of tropical region. The poorest topographical attributes for delimitation of maps were that using the profile curvature, aspect, and slope as input data, individually as well as combined (maps 5, 7, 8, and 9), which indicates that elevation is more closely related with the land classes than the other three elements.

The findings of this research for the tropical area coincides with those of other authors that using also decision trees as classifier; Lagacherie and Holmes (1997) that found 75% of precision in their soil maps, and Moran and Bui (2002) who found 65% precision, that are lower that the results of this research. On the other hand, Segura *et al.* (2003) reported 70% precision in their map of local soil classes using the parallelepiped as the classifier.

3.2 *Evaluation of accuracy*

According to the technique of Burgess and Webster (1984), the highest sample size was from 448 (map 9) to 2425 (map 2) for the map of local soil classes of San Luis Potosi, since it has the greatest length of boundaries, followed by the temperate region from 16 (map 7) to 111 (map 4) in number of samples, and finally the tropical region from 26 (map 1) to 58 (map 8). The tropical region has greater length of boundaries than the temperate area, but it has a lower average distance between boundaries, this situation gives a smaller number of samples per ha (Table 2). The term "empty" in table 2, means that the decision tree classifier only was able to identify only one class for all the area.

The accuracy obtained in this research for the maps of local soil classes for the tropical region was the higher followed by the maps of the temperate, and arid regions. Similar to the results obtained in the case of precision the maps with more accuracy were those where the TNM

were used as input data for CAC as shown in Figure 3. The exception to this was map 1 generated with the four topographical features that was better than map 6 with TNM for the temperate region. Nevertheless, the highest accuracy was less than 50% in all the maps generated.

The differences between accuracy and precision of the maps found in this research coincide with that of Lleverino *et al.* (2000). They found that these parameters could have a variation of more than 20%. This means that a map can be precise but not in all the cases have the appropriate accuracy. For this study, the difference between precision and accuracy was of more than 35% higher than Lleverino *et al.* (2000) variation for the three study regions.

3.3 Analysis of elevation and better maps

As was conclude previously the TNM is the feature that produced the better maps, as observed in Figure 4. The best maps for local soil classes are explained mainly by the elevation ranges. In the SLP and the state of Mexico, the ranges of altitude are not clearly defined with intersection in most of the lines of the plot. In contrast, for Veracruz the intervals are better determined, except in local soil class 2 (Barrial) in which the minimum elevation values of 50 intersect with the values of the other lines this can produce misinterpretations. Moreover, the number of local soil class identified in the maps of the temperate and arid areas affect the precision and accuracy of the maps, by the inherent complexity of the distribution local soil classes, this situation affect the quality of the maps generated by CAC because in these areas the range of elevation is difficult to establish.

The best maps produced by CAC are presented in Figure 5. The Map 4 of the arid region is considered the best because its precision and accuracy are 39% and 10% respectively, in comparison with the other eight maps (Figure 1.5a). The low quality of the map is because the local classes Arena, Arena gravosa, Calichuda, Cuerpo and Enlamada were underestimated by more than 20% and the remaining classes were overestimated by more than 30%. The Map 6 of the temperate region had the highest precision and accuracy with 42% and 14%, correspondingly. However, it is still considered of low quality because it is not more than 50% in either attribute. The main problems found were the overestimation by more than 60% of the Arena and Lama classes, and the underestimation of the Blanca and Cacahuatuda classes underestimated by less than 16% (Figure 5b). In tropical area map 6 was the best with 72%

precision and 33% accuracy. The local soil class overestimated by more than 100% was Arenal, this could be explained because some Barrial class sites were confused for Arenal class, reducing the precision in its estimation (Figure 5c).

4. Conclusions

From the four topographical attributes considered for the generation of maps by CAC, the terrain numerical model (TNM) produced the maps with the best quality maps; is advice that for future research the TNM must be considered as only indispensable attribute for CAC for local soil classes in the three regions studied. But, the procedures for the generation of the maps to reach a better quality must be improved, because for some of the maps produced the precision is acceptable especially for the tropical area. However, the accuracy was low for the maps of the three areas. From this work could be inferred that the production of maps with CAC could generate maps with acceptable precision some regions and the accuracy of the maps remains still as a problem.

5. References

- BARRERA-BASSOLS, N. and ZINK, J.A., 2000, *Ethnopedology in a worldwide perspective: anannotated bibliography*. ITC Publication, vol 77, Enschede Netherland.
- Brown, B.R., 1988, Concerning the quality of soil survey. *Journal of Soil and Water Conservation*, 43, pp. 452-455.
- BURGESS, T.M. and WEBSTER, R., 1984, Optimal sampling strategies for mapping soil types II Risk functions. and sampling intervals. *Journal of Soil Science*, 35, pp. 655-665.
- CONGALTON, R.G., 1988, A comparison of sampling scheme use in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54, pp. 593-600.
- CRUZ, C.G., ORTIZ, S.C.A., GUTIÉRREZ, C. MA. DEL .C. and VILLEGAS, M.A., 2008, Las clases de tierras citrícolas del ejido Pueblillo, Papantla, Veracruz. *Terra Latinoamericana*, 26, pp. 11-19.
- DOBOS, E., MONTANARELLA, L., NEGRE T. and MICHELI, E., 2001, A regional scale soil mapping approach using integreted AVHRR and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, 3, pp. 30-42.

- EASTMAN, J., 2006, *IDRISI Andes: Guide to GIS and Image Processing*. Clark Labs, Clark University, Worcester, MA.
- FOODY, G.M., 1992, On compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, 58, pp. 1459-1460.
- FRANÇOIS, M.J., DÍAZ G. J.R. and PÉREZ, V.A., 2003, Evaluación de la confiabilidad temática de mapas o de imágenes clasificadas: una revisión. *Investigaciones Geográficas*, 51, pp. 53-72.
- GARCÍA, E., 1988, *Modificaciones al sistema de clasificación climática de Köppe*. Ed Ind, México.
- GIASSON, E., CLARKE, T.R., INDA JUNIOR, V.A., MERTEN, H.G. and TORNQUIST, G.C., 2006, Digital soil mapping multiple logistic regression on terrain parameters in Southern Brazil. *Scientia Agricola*, 63, pp. 262-268.
- HENGL, T. and ROSSIETER, G.D. 2003, Supervised landform to enhance and replace photo-interpretation in semi-detailed soil survey. *Soil Science Society of America Journal*, 67, pp. 1810-1822.
- INSTITUTO NACIONAL DE ESTADÍSTICA GEOGRAFÍA E INFORMÁTICA (INEGI), 1988, *Atlas nacional del medio físico*. Aguascalientes, México.
- INSTITUTO NACIONAL DE ESTADÍSTICA GEOGRAFÍA E INFORMÁTICA (INEGI), 2007, *Sistema de Descarga del Continuo de Elevaciones Mexicano*. <http://www.inegi.gob.mx/inegi/default.aspx>.
- KRASILNIKOK, V.P. and TABOR, A.J., 2003, Perspectives on utilitarian ethnopedology, *Geoderma*, 111, pp. 197-215.
- LAGACHERIE, P. and HOLMES, S., 1997, Addressing geographical data errors in a classification tree for soil unit prediction. *International Journal Information Geographical Science*, 11, pp. 183-198.
- LLEVERINO, G.E., ORTIZ, S.C.A. and GUTIÉRREZ, C. MA. DEL C., 2000. Calidad de los -mapas de suelos en el ejido de Atenco estado de México. *Terra*, 18, pp. 103-113.
- MACBRATNEY, B. A., MENDONÇA-SANTOS, L. M. and MINASNY, B., 2003, On digital soil mapping *Geoderma*, 117, pp. 3-52.

- MARTÍNEZ, M.J.F., ORTIZ, S.C.A., PÁJARO, H. D., OJEDA, T.E. and VARGAS, P.E., 2003, Clave de fotointerpretación para la cartografía de tierras campesinas a escala regional. *Terra*, 21, pp. 301-309.
- MINITAB, 2003, *Statistical Software*. State College PA, USA
- MORAN, J.C. and BUI, N.E., 2002, Spatial data mining for enhanced soil map modeling. *International Journal Information Geographical Science*, 16, pp. 533-549.
- NIEMEIJER, D. and MAZZUCATO, V., 2003. Moving beyond indigenous soil taxonomies: local theories of soils for sustainable development. *Geoderma*, 111, pp. 403-424.
- ORTIZ, S.C.A., PÁJARO, H.D. and ORDAZ, CH. V., 1990, *Manual para la cartografía de clases de tierra campesinas*. Serie de Cuadernos de Edafología 15 Centro de Edafología, Colegio de Postgraduados, México.
- ORTIZ-SOLORIO, C.A. and GUTIÉRREZ-CASTORENA, MA. DEL C., 2001, Mexican Ethnopedology: the Experience in Soil Mapping. *Science and Tradition: Roots and Wings for Development*, pp. 107-136.
- ORTIZ-SOLORIO, C.A., GUTIÉRREZ-CASTORENA, MA. DEL C., LICONA-VARGAS, A.L. and SÁNCHEZ-GUZMÁN, P., 2005, Contemporary influence of indigenous soil (land) classification in Mexico. *Eurasian Soil Science*, 38, pp. S89-S94.
- PÁJARO, H.D. and ORTIZ, S.C.A., 1987, *El levantamiento de suelos y su relación con la clasificación y cartografía de las clases de tierras campesinas*. Colegio de Postgraduados, México.
- PENG, W., WHEELER, B.D., BELL, C.J. and KRUSEMARK, M.G., 2003, Delineating patterns of soil drainage class on bare soils using remote sensing analyses. *Geoderma*, 115, pp. 261-179.
- QI, F. and ZHU, A.X., 2003, Knowledge discovery from soil maps using inductive learning. *International Journal Information Geographical Science*, 17, pp. 771-795.
- QI, F., ZHU, X. A., HARROWER, M. and BURT, J.E., 2006, Fuzzy soil mapping based on prototype category theory. *Geoderma*, 136, pp. 774-787.
- QUINLAN, R.J., 1988, Simplifying decision trees, In: B Gaines and J Boose (Eds) *Knowledge Acquisition for Knowledge-Based System*, Academic Press CA, 241-254.
- RZEDOWSKI, J. and HUERTA, L., 1985, *La vegetación de México*. Limusa, México.

- SCHMIDT, J. and HEWITT, A., 2004, Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma*, 121, pp. 243-256.
- SCULL, P., FRANKLIN, J., CHADWICK, O.A. and MCARTHUR, D., 2003, Predictive soil mapping: a review. *Progress in Physical Geography*, 27, pp. 171-197.
- SEGURA, C. M. A., ORTIZ, S. C. A. and GUTIÉRREZ, C. MA. DEL C., 2003, Localización de suelos de humedad residual a partir de imágenes de satélite: Clasificación automática supervisada de la imagen. *Terra*, 21, pp. 149-156.
- SHANER, J. and WRIGHTSELL, J., 2000, *Editing in ArcMap*. Environmental Systems Research Institute (ESRI). New York Street Redlands, CA USA.
- SMITH, P.M., ZHU, X.A., BURT, E.J. and STILES, C., 2006, The effects of DEM resolution and neighborhood size on digital soil survey. *Geoderma*, 137, pp. 58-69.
- ZHU, X.A., HUDSON, B., BURT, J., LUBICH, K. and SIMONSON, D., 2001, Soil mapping using GIS expert knowledge and fuzzy logic. *Soil Science Society of America Journal*, 65, pp. 1463-1472.
- ZIADAT, M.F., 2007, Land suitability classification using different sources of information: Soil maps and predicted soil attributes in Jordan *Geoderma*, 140, pp. 73–80.

Table 1. Topographical attributes used in the classification with the decision tree.

Map 1	Map 2	Map 3	Map 4	Map 5
TNM	TNM	TNM	TNM	Aspect
Aspect	Aspect	Profile curvature	Aspect	Profile curvature
Profile curvature	Slope	Slope	Profile curvature	Slope
Slope				

Table 2. Sample size per map type and climatic region to calculate accuracy.

Num. Map	Arid region		Temperate region		Tropical region	
	Boundary length (m)	Sample size	Boundary lengths (m)	Sample size	Boundary lengths (m)	Sample size
Map 1	406706	2350	18255	97	35783	26
Map 2	422633	2414	18255	97	35783	26
Map 3	396741	2318	14046	97	35783	26
Map 4	419463	2330	16505	111	35783	26
Map 5	468522	2108	Empty	Empty	37668	42
Map 6	435784	2425	13307	70	35783	26
Map 7	460288	1992	1386	16	37716	28
Map 8	410961	1857	Empty	Empty	107394	58
Map 9	117318	448	Empty	Empty	74828	35

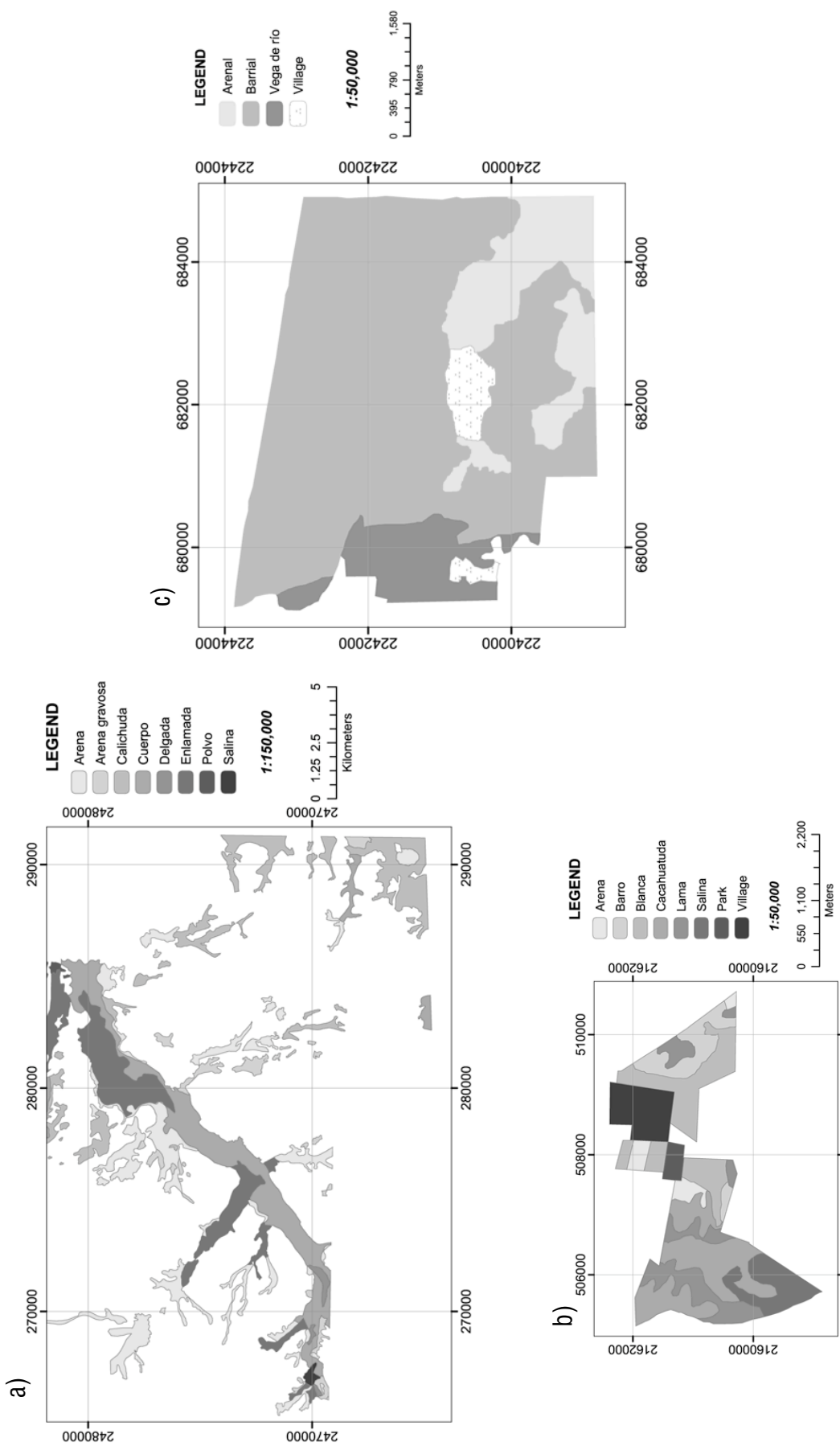


Figure 1. Maps of local soil classes; regions: arid (a), temperate (b) and tropical (c).

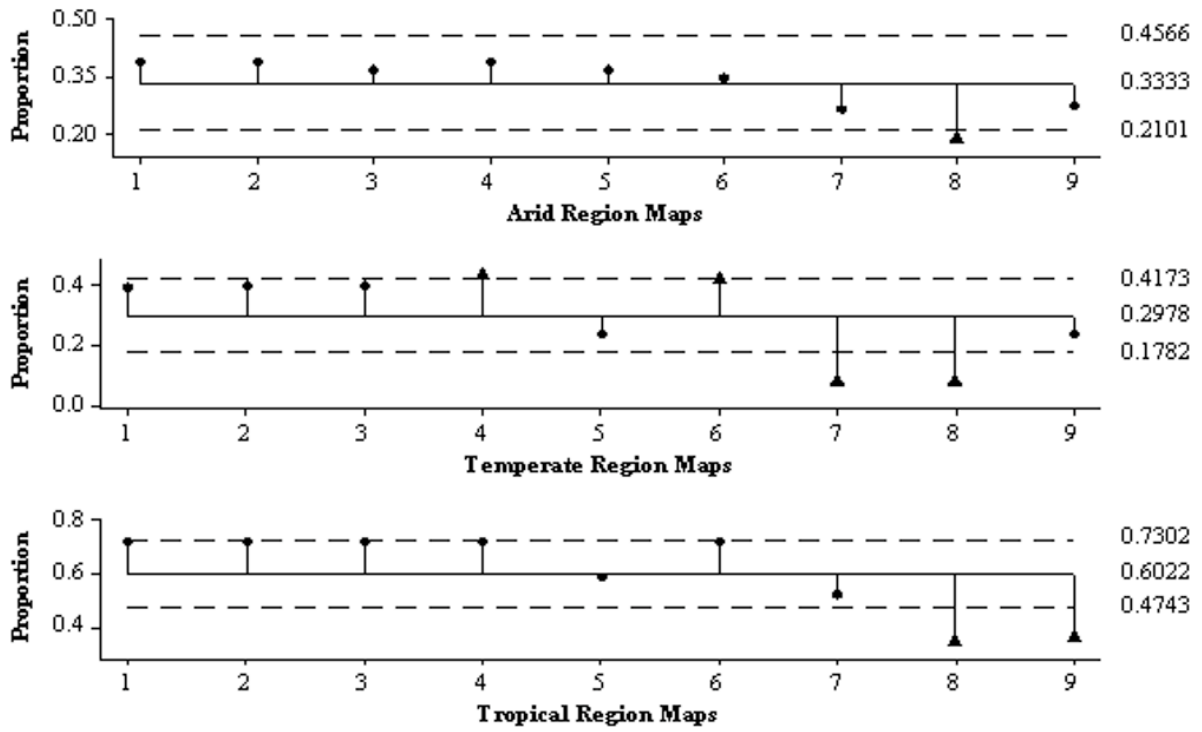


Figure 2. Binomial proportion test ($\alpha=0.05$) for precision of the three regions. †Table 1.1.

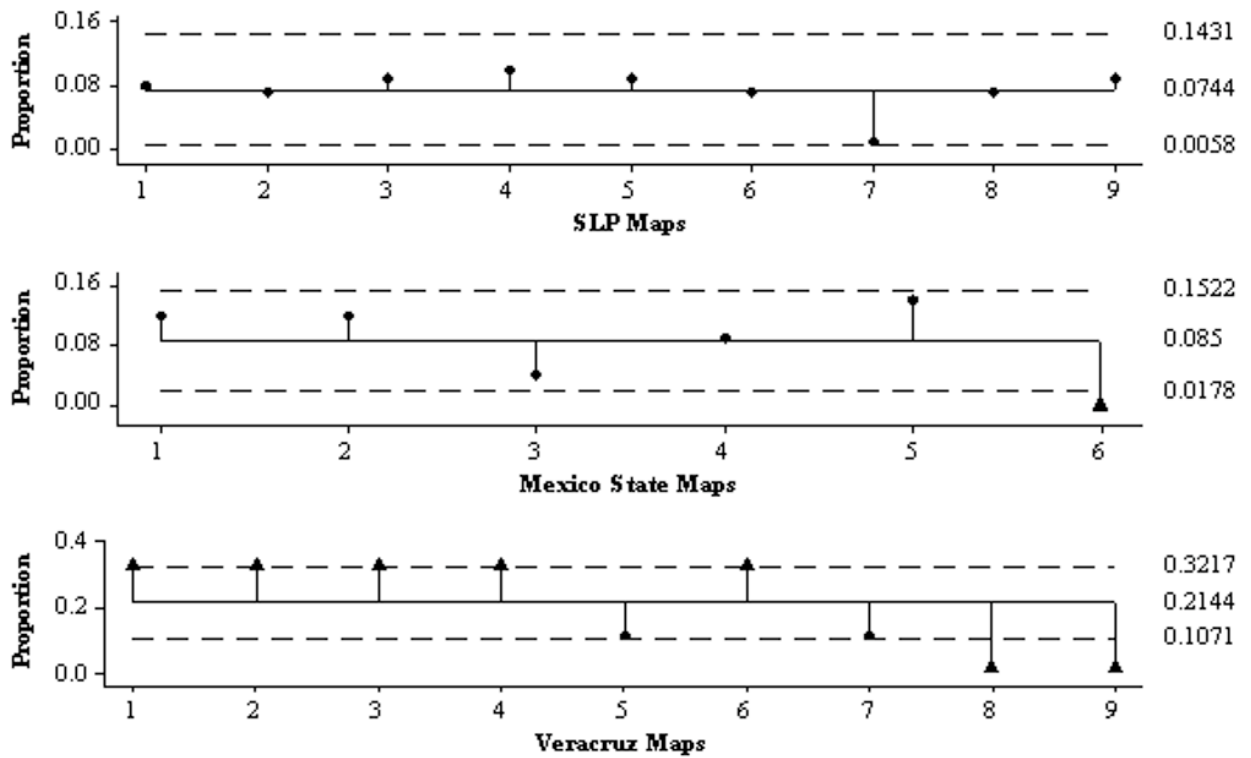


Figure 3. Binomial proportion test ($\alpha=0.05$) for accuracy of the three regions. †Table 1.1.

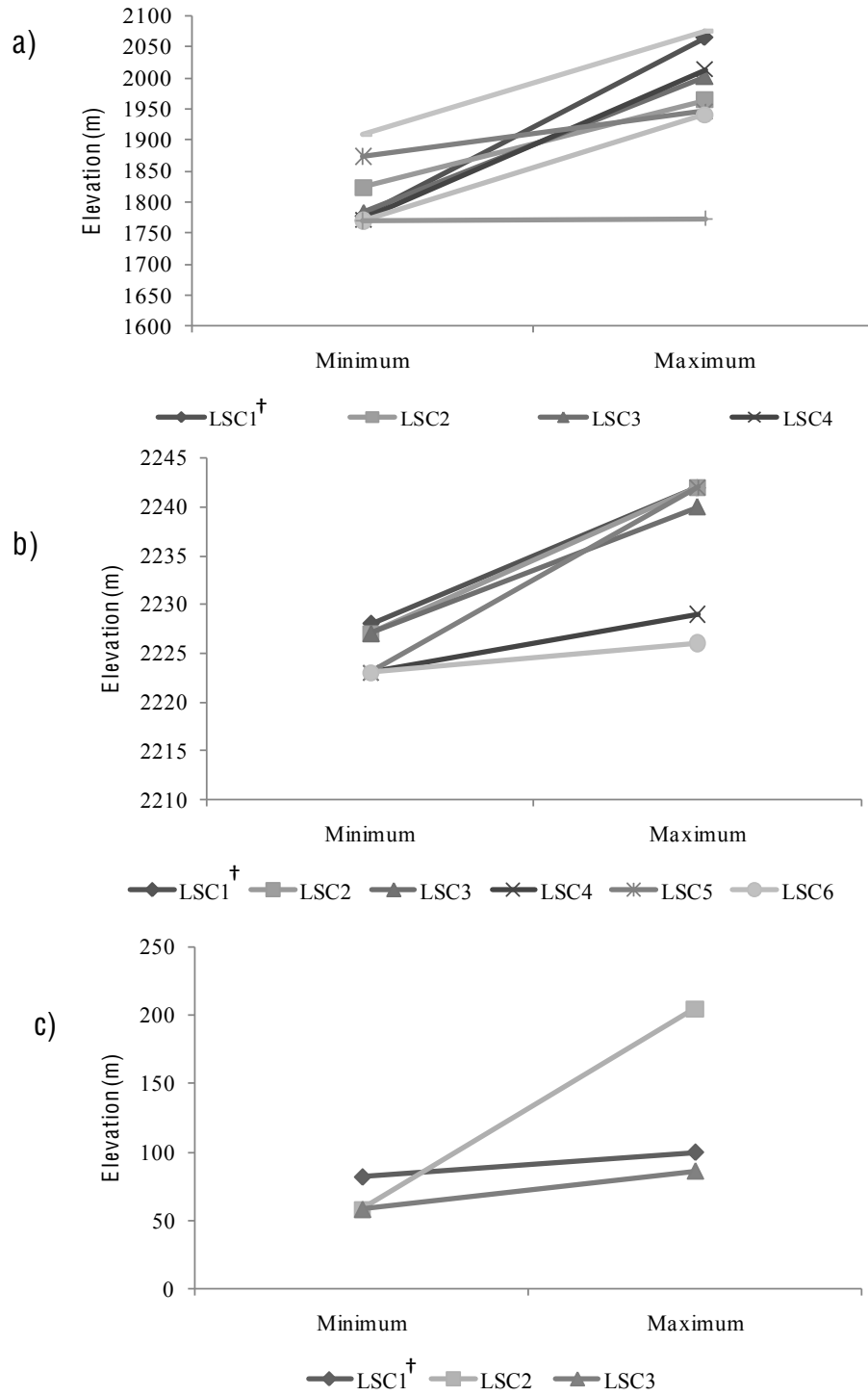


Figure 4. Minimum and maximum altitude values of local soil classes[†] of the arid (a), temperate (b) and tropical (c) regions.

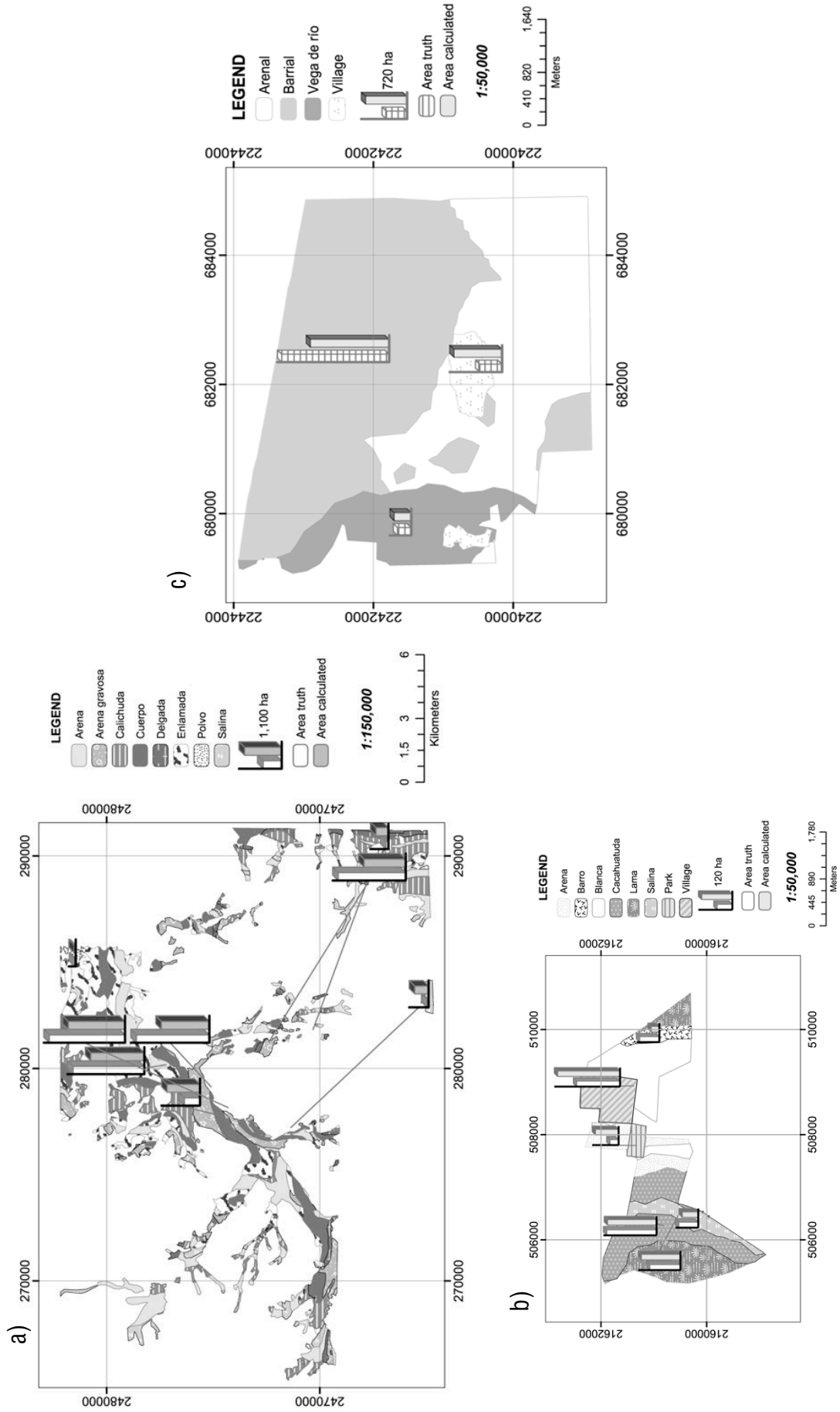


Figure 5. Map 4 (TNM, aspect, and profile curvature) of the arid region (a), Map 6 (TNM) of the temperate region (b) and Map 6 (TNM) of the tropical region (c)

CAPITULO II

EVALUATION OF FOUR DIGITAL CLASSIFIERS FOR AUTOMATED CARTOGRAPHY OF LOCAL SOIL CLASSES BASED ON REFLECTANCE AND ELEVATION IN MEXICO

EVALUATION OF FOUR DIGITAL CLASSIFIERS FOR AUTOMATED CARTOGRAPHY OF LOCAL SOIL CLASSES BASED ON REFLECTANCE AND ELEVATION IN MEXICO¹

G. Cruz-Cárdenas^a, C.A. Ortiz-Solorio^a, E. Ojeda-Trejo^a, J.F. Martínez-Montoya^b, E.D. Sotelo-Ruiz^c, and A.L. Licona-Vargas^d

^a*Colegio de Postgraduados, Campus Montecillo, Carretera México-Texcoco km 36.5, 56230, Texcoco, México, México.* ^b*Colegio de Postgraduados, Campus San Luis Potosí, Iturbide 73, 78620, Salinas de Hidalgo, Salinas, SLP, México.* ^c*Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Campo Experimental Zinacantepec, Carretera Zacatepec-Galeana km 0.5, 62780, Zinacantepec, México, México.* ^d*Universidad Autónoma Chapingo, CRUO, km 3 Carretera Huatusco-Jalapa, 94100, Huatusco, Veracruz, México.*

Resumen

El objetivo de este estudio fue evaluar la precisión y exactitud de los mapas de clases de suelos locales generados por cuatro algoritmos computacionales: mínima distancia, paralelepípedo, máxima verosimilitud y redes neuronales artificiales. Se empleó como datos de entrada el modelo digital de elevación y las firmas espectrales de las clases de suelos locales. El estudio se realizó en los estados de México, San Luis Potosí y Veracruz. La comparación de la precisión y exactitud entre mapas se efectuó con la prueba de proporción binomial. La combinación de la reflectancia y elevación mejoró la calidad de los mapas de clases de suelos locales producidos por CAC, debido a que la reflectancia de las clases de suelos locales varía acorde a la altitud, con lo cual ayuda a identificarlas de mejor manera. El mejor mapa obtuvo una precisión fue de 84% y exactitud de 30%.

Palabras clave: Modelo digital de elevación, redes neuronales artificiales, calidad de mapas.

Abstract

The objective of this research is to evaluate the precision and accuracy of local soil class maps generated with four computer algorithms: minimum distance, parallelepiped, maximum likelihood, and artificial neural networks, using digital elevation models and spectral signatures of local soil classes as input data. The study was done in the states of Mexico, San

¹Accepted for publication in *International Journal of Remote Sensing*.

Luis Potosi, and Veracruz. Statistical binomial proportion tests were done to compare the difference between maps' precision and accuracy. The conclusion was that the combination of reflectance and elevation improved the quality of soil class maps produced by CAC, due to the reflectance variation of local soil classes according to altitude, which helped to better identify them. The best precision was of 84% and the best accuracy of 30%.

Key words: Digital elevation models, artificial neural networks, map quality.

1. Introduction

In Mexico, local soils, or etnopedological, classification works began in the 1980s. The most outstanding results up date are: a) farmers do not recognize the term soil in its strict technical sense, but have an integrated holistic concept similar to the one of land; b) the earth types identified by farmers are related to several soil properties (technical criteria), such as color, texture, consistency, water retention, workability, etc.; moreover, they consider other characteristics, or soil formation processes, as sub-surface conditions or the relationship with native plants, among others (Ortiz-Solorio *et al.* 2005); c) earth local knowledge through an observation scale has a close soil-landscape relationship compared to conventional technical systems, such as Soil Taxonomy or Food Agriculture Organization (FAO) system, usually used in Mexico in general observation scales (1:1 000 000 to 50 000) (Ortiz and Gutierrez 2000); d) technical soil surveys in Mexico and other countries, 1:50 000 scale, do not provide enough information to give recommendations at farm level, thus the information they produce is often useless to support decisions made by farmers (Ortiz-Solorio and Gutierrez-Castorena 2001, Krasilnikov and Tabor 2003). On the other hand, Barrera-Bassols *et al.* (2006) evaluated spatial correlation between technical and local soil mapping and found that there are some discrepancies, but also similarities which make them both necessary for a soil system better understanding.

Technical soil surveys are conducted by experts with technical standardized procedures and tools such as aerial photographs, satellite images and laboratory analysis. Local soil surveys are produced at local level and carried out using farmers' knowledge of their lands with a participatory approach; by these means individual mental maps of boundaries and different types of local soil classes are spatially integrated in an area of local lands. Farmers use their own experience about the relation soil-land use by long periods of time as tools for local soil

differentiation. Bassols *et al.* (2006) indicate that the farmers can distinguish with mental accuracy the different classes of local soils using factors such as relief, vegetation, land use, and mesoclimate. Ortiz-Solorio *et al.* (2005) in a summary of works about ethnopedological studies in Mexico compared with technical surveys concludes that both systems have a similar classification structure. The local soil classes are considered as tridimensional bodies, both of them have hierarchical levels and formal taxonomy; in addition, there are coincidences among the maps of local soil classes at the level of Word Reference Base or Soil Taxonomy subunits or subgroups, respectively (Ortiz-Solorio *et al.* 2005).

Based on this, we can say that the conjunction of soils local and technical knowledge is fundamental to produce quality maps from the perspective of cartography and to help farmers and technicians in making decisions. Geographical Information Systems (GIS), remote sensing, and image classifiers are useful tools in the process of map production (Payton *et al.* 2003).

In soil cartography, the use of satellite images has proven its usefulness (McBratney *et al.* 2003). However, in cartography of local soil classes, contrasting results have come out when using conventional and automated methods (Cajuste 1991, Martinez 1993). One of the main obstacles of digital cartography has been that only reflectance (tone) has been used as main input data to differentiate local soil classes, with generally poor results. On the other hand, cartographic approaches where processes similar to traditional photointerpretation have been used, taking into account other characteristics such as location, form, drainage, etc., the results have been better than when using only one variable (Licona *et al.* 1993, Martinez *et al.* 2003). Nowadays, the availability of new digital classifiers and data bases, like digital elevation models (DEM), as well as advances in GIS capabilities, allow to generate automatically new parameters, such as altitude layers, slope, aspect, among others, which are criteria that under certain conditions discriminate local soils or soil classes from aerial photographs (Martinez *et al.* 2003, McBratney *et al.* 2003). Peng *et al.* (2003) mention that soil reflectance varies with its slope, which means that the integration of altitude into the process can improve the classification of local soil classes. There is evidence that DEM has been used in forest cover computer assisted cartography (CAC), improving map accuracy by over 20% (Shuoep and Marsh 2004). Diverse authors have done work on soil cartography using GIS and DEM (Heng and Rossiter 2003, Qi and Zhu 2003). The variables used in cartography have been altitude,

slope, curvature profile, shadow, and aspect, extracted from DEM. The precision of some soil maps generated through this technique was over 50%. This indicates that DEM can be useful in local soil class cartography, generating maps of acceptable quality using these methods.

The indicators to evaluate a map's quality are its precision and its boundary accuracy, there being little information on the second parameter and its usefulness in CAC. The cartographic precision of a map is evaluated by the purity of the cartographic units classified in the legend (Nordt *et al.* 1991). The purity of a soil class uses as a central concept the typical soil profile. The precision of the soil unit is evaluated by the measurement of the level of dispersion from the properties of the soil or from the typical profile for each cartographic unit (vgr. Fluvisol) (Brown 1988). The evaluation of the purity of a cartographic unit is obtained by the computation of the percentage of sample points that belong to the cartographic unit classified in the legend (Bie and Beckett 1973). The accuracy of the map is related with the localization of the boundaries of each class of legend. This depends on the location of the boundaries (Brown 1988). A boundary is more accurate if the soils in both sides of the line are classified as different soil types. The accuracy of the boundary is the percentage of correct boundaries (Young and Stoeckeler 1956). Lleverino *et al.* (2000) found, in local soil class maps in Atenco, Mexico, that accuracy can vary by more than 20% when compared with its precision. This means that a map with good precision not always has an adequate accuracy. The main objective of this work is to evaluate the precision and accuracy of local soil maps in contrasting conditions of climate and landscape using four digital classifiers, reflectance and DEM as input data.

2. Materials and methods

2.1 *Zone of study*

The study was done in three regions of Mexico (Figure 1). The first region corresponds to the arid zone (SLP), and is located between coordinates 101° and 101 ° 20' W and 22 ° 16' and 22 ° 34' N altitude varies from 2,217 to 2,244 m. and consists of 9,526 ha. The climate is arid (Bs) according to Garcia (1988); rock types include limestone, shale, conglomerates, and sandstone with some patches of igneous and metamorphic rocks (INEGI 1988). The second region, temperate region in the State of Mexico, lies between 98°39' and 99° W, and 19°22' and 19°32' N; altitude from 1,770 to 2,000 m, 685 ha. The climate is sub-humid temperate

(C(w)); existing geology is made up of basic volcanoclastic and acid igneous extrusive rocks (Garcia 1988, INEGI 1988). The last region, tropical area or Veracruz, is located between 97°10' and 97°30' W, and 20°10' and 20°40' N; altitude varies from 50 to 300 m, with 2,300 ha. The climate is sub-humid warm (Garcia 1988). Rock formations include calcareous shale, alluvial sediments, and small proportions of tephra (INEGI 1988).

2.2 Local soil class maps, remote sensing imagery and DEM

The local soil class maps were taken from Martinez et al. (2003), Pajaro and Ortiz (1987), and Cruz (2006) in these maps field information was generated according to the methodology by Ortiz *et al.* (1990). The maps were elaborated at different scales 1:20 000, 1:30 000 and 1:37 500 respectively, but generalized to scale 1:50.000. They were digitized using ArcView 8.1, by ESRI™ (Shaner and Wrightsell 2000) and then imported to IDRISI™. Landsat ETM+ images were used, dated 28/11/1999 in SLP, 21/03/2000 in the state of Mexico, and 06/09/2000 in Veracruz. Later, each image was resized to fit the area; digital numbers were transformed to reflectance according to the User's Manual of Landsat-7 (NASA-GSFC 2007). The Jeffries-Matusita model was used to identify the combination of bands with the best separation of spectral signatures of local soil classes. DEM was extracted from the download system of the Continuo de Elevaciones Mexicano (INEGI 2007: <<http://www.inegi.gob.mx/inegi/default.aspx>>) and it was generated at scale 1:50, 000 . The grid is of 1 x 1 second. The georeference is in projection UTM, datum ITRF92, 1988.0, and the ellipsoid is GRS80. It was configured according to the characteristics of the window of the remote sensing imagery.

2.3 Training sites and classifier

For the classification of the local soil classes, randomized training sites were selected covering most of the variation inside each soil classes. Four classifiers were used maximum likelihood (MAXLIKE), minimum distance (MINDIST) parallelepiped (PIPED), and inverse propagation artificial neural networks (ANNs). The sample size of the training was 30 pixels for the statistical classifiers; in the ANNs, the sample size was 20 pixels: 10 pixels were used for training of the network and 10 pixels to testing.

The MAXLIKE computes the probability that a pixel belongs to a class. Data from the training

sets uses the mean vector and the covariance matrix of the cluster of each category of brightness values to be computed. The pixel is assigned with the highest probability value; the MINDIST computes the Euclidian distance between a pixel and the cluster of categories derived from training areas in a feature space defined by the number of spectral bands. A pixel is assigned to the closest cluster; the PIPED uses the upper and lower limits of brightness values in each band for each category of class (as determined from training) and is fixed as a box. The brightness value of each pixel is then compared to each of these boxes to determine to which box the pixel's brightness value should belong (Lo and Yeung 2007).

The structure of the ANNs was built of an input layer, two hidden layers, and an output layer (Figure 2). The Delta learning rate was of 0.01 (Serpico *et al.* 1996); this process minimizes iteratively an error function over the network outputs taken from the training data set. The learning process starts by entering the training data in the network and setting the weights of interconnecting channels randomly. These data flow forward through the network to the output units. The difference between the desired and actual output is calculated and represents the network error. This error is sent backwards through the network toward the input layer, changing the weights of the interconnecting datum input in relation to the error. The momentum was of 0.05 (Blackard and Dean 1999), that helps the network avoid local minima in the error surface and weight oscillations around the minimum. The momentum term is added from the weight change determined in the preceding iteration to the calculation. The number of nodes of the hidden layers was empirically determined (Chen and Burrell 2002). The whole process is repeated until acquiring 100% accuracy rate or 10, 000 iterations (maximum iterations). The structure of the hidden layers (HL) starts with 6 and 8 nodes from the first and second layer respectively (Jensen *et al.* 1999). To get the highest accuracy rate, the HL structure was changed and the program was run several times.

The following input data combinations were used with each classifier: reflectance, DEM, and both reflectance and DEM, except maximum likelihood where only the first and third used, since problems arose when using only DEM to classify. This configuration was used in the three climatic regions.

2.4 Evaluation of precision and accuracy

Sample sizes were generated with 1% of the total pixels of the maps from each region (Congalton 1988). The spatial arrangement of the sample was random systematic (François *et al.* 2003). The sample points were crossed with the CAC maps and the ones representing field information.. Confusion matrixes were generated; the sum of the diagonal in each matrix represents the global accuracy of the maps (Figure 3).

The evaluation of accuracy was achieved in the maps of the three areas at scale 1:50,000 according to the procedure of Burgess and Webster (1984) The procedure signals that a grid of 1 cm² aligned from north to south and east to west, for each maps have to be produced. After this the distance from each consecutive pair of boundaries is measured and transformed to meters according to the scale The distances of each direction is added and multiplied by 0.52. The product is divided by the total surface of study to obtain the optimal sampling spacing. This result is divided between the total length of the boundaries to determine the sample size. The sample points are placed to 50 m (1 mm in 1:50 000) in both sides of the boundary. Then the correct proportion of sample points is computed and the accuracy of boundaries is produced. This procedure applied to the maps is showed in (Figure 4).

The quality of the maps was evaluated with the MINITAB (2003) statistical software using the mean contrast for binomial proportions of precision and accuracy test.

3. Results and discussion

The greatest separation of the local soil classes spectral signatures for the three areas using the Jeffries-Matusita distance method was obtained by using bands 1, 2, 3, 4, 5, and 7. These bands were used to generate the maps through CAC (Figure 5). Dwivedi *et al.* (2004) also found that this combination of bands gave the best separations to identify changes in land use. The map precision was obtained by applying the Congalton (1988) method, generating the confusion matrix of the local soil class maps. In general, the use of DEM in conjunction with reflectance increases the precision of the generated maps by over 15% with most classifiers and contrasting climatic areas (Table 1).

Precision in some maps, like the tropical region one, is increased by more than 30% because one of the factors that differentiate local soil classes is topography, especially on locations with significant altitude differences (Licona *et al.* 1993, Martinez *et al.* 2003). This is similar

to stereoscopic photointerpretation, where areas are separated based on altitude; however, Hengl and Rossiter (2003) point out that automated interpretation is faster, and the boundaries are better defined and georeferenced. The results found are similar to those by Moran and Bui (2002), which establish that the use of elevation and Landsat MSS images increased the precision of their soil maps from 60 to 70% more than when they used only spectral signatures. Other works where it has been demonstrated that the use of DEM increases map precision are those by Dobos *et al.* (2001), and Hengl and Rossiter (2003), whose maps have a precision of over 50%. On the other hand, maps generated using reflectance alone had the lowest precision (Table 1). Instead, the ones generated with MAXLIKE, reflectance, and DEM had better precision for the arid region, 42% (Figure 6b), and the temperate region, 62% (Figure 7b) than those generated with ANN. Contrary to the results obtained in forest cover classification (Blackard and Dean 1999), soil moisture prediction (Persson *et al.* 2002), and identification of tepetate types (Trueba *et al.* 2004), where the ANNs increased the precision of the maps from 60 to 95%, in this research, the results using ANNs were only satisfactory in the tropical region (Figure 8b). Table 2 shows that the numbers of pixels of the training sample do not have a direct relation with the accuracy rate of the ANNs. For example, in the tropical area the network training was of 30 pixels and an accuracy rate of 97% in contrast with the temperate region which was of 60 pixels and an accuracy rate of 20%. Despite the fact that the tropical areas are more than three times its size.

When comparing this classifier to the MINIDIST and PIPED, the difference in precision is more than 40%, and precision remains at 80%. This is similar to what was found by Zhu *et al.* (2001) and Qi and Zhu (2003), using the Fuzzy logic method and the decision tree as classifiers, respectively, with the difference that they used altitude, aspect, geological layer, and drainage as input data to discriminate soil units in abrupt zones. Both studies had a precision of around 84%.

The number of sampling points to assess the accuracy of the maps is represented in Figure 9: the greater the total length of the boundary (Figure 9a), the greater the sample size (Figure 9b). Firstly, this is due to the size of the study area, and, secondly, the number of cartographic units contained in each map. Thus, the maps with a greater boundary length are those of the arid zone, which is the one generated by MAXLIKE and reflectance.

In general, the accuracy of the maps is less than 30%. Like in the case of precision, DEM is an element that helps to increase the accuracy of the maps with most classifiers and areas of study. For the arid region, PIPED is the best classifier with 14% accuracy; so are MAXLIKE for the temperate region with 17%, and ANNs for the tropical region with 30% (also MAXLIKE) (Table 3). This latter map has the digital elevation model in common with higher precision of the temperate and tropical regions maps, indicating that elevation is an important feature, which is closely related with local soil classes.

On the other hand, Lleverino *et al.* (2000) found that the accuracy of maps using field evaluations was of 94% in local soil classes, 8% in traditional soil surveys, and 0% in INEGI. When comparing these to the present study, we can notice that the local soil class maps done in the traditional form are the only ones that gave better results than all those generated by CAC (Figure 9) and those done by soil surveys.

4. Conclusions

The combination of bands 1, 2, 3, 4, 5, and 7 can be used to separate the signatures of local soil classes. The combination of reflectance with the digital elevation model in computer assisted cartography improves precision and accuracy of the local soil class maps in the three regions by more than 30%. The best classifier was MAXLIKE for all three climatic regions. The quality of some maps of the temperate and tropical regions had above 60% of precision; however, accuracy in all maps was less than 30%. In previous studies, only precision had been used as an indicator of quality, and as it was proved in this research, it is not enough to have good precision but it is also necessary to evaluate accuracy to really assess a map's quality.

5. References

- BARRERA-BASSOLS, N., J.A. ZINCK, J. A. and VAN RANST, E., 2006, Local soil classification and comparison of indigenous and technical soil maps in a Mesoamerican community using spatial analysis. *Geoderma*, 135, pp. 140-162.
- BIE, S.W. and BECKETT, H. P., 1973, Comparison of four independent soil surveys by airphoto interpretation. Paphos area (Cyprus). *Photogrametría*, 29, pp. 189-202.

- BLACKARD, A. J. and DEAN, D.J., 1999, Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and Electronics in Agriculture*, 24, pp. 131-151.
- BROWN, B.R. 1988, Concerning the quality of soil survey. *Journal of Soil and Water Conservation*, 43, pp. 452-455.
- BURGESS, T.M. and WEBSTER, R., 1984, Optimal sampling strategies for mapping soil types. II. Risk functions and sampling intervals. *Journal of Soil Science*, 35, pp. 655-665.
- CAJUSTE, B.L., 1991, *Procedimiento para la Cartografía de Clases de Tierras Campesinas mediante Sensores Remotos*. MSc thesis, Colegio de Postgraduados, México.
- CHEN, D. and BURREL, P., 2002, On the optimal structure design of multilayer feedforward neural networks for pattern recognition. *Internacional Journal of Pattern Recognition and Artificial Intelligence*, 16, pp. 375-398.
- CONGALTON, R. G. 1988. A comparison of sampling scheme use in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54, pp. 593-600.
- CRUZ, C.G., 2006, *Las Clases de Tierras con producción de Naranja (Citrus sinensis L. Osbeck) en Papantla*, Veracruz. MSc thesis, Colegio de Postgraduados, México.
- DOBOS, E., MONTANARELLA, L., NEGRE, T. and MICHELI, E., 2001, A regional scale soil mapping approach using integrated AVHRR and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, 3, pp. 30-42.
- DWIVEDI, S. R., KANDRIKA, S., and RAMANA, V. K., 2004, Comparison of classifiers of remote sensing data for land use/land cover mapping. *Current Science*, 86, pp. 328-335.
- ESWARAN, H., FORBES, R.T. and LAKER, C.M., 1981, Soil map parameters and classification. *In Proceeding of Workshops Soil Resource Inventories and Development Planning, at Cornell University 1977-1978*, US Department of Agriculture, pp. 37-57.
- FRANÇOIS, M.J., DÍAZ G., J.R. and PÉREZ, V.A., 2003, Evaluación de la confiabilidad temática de mapas o de imágenes clasificadas: una revisión. *Investigaciones Geográficas*, 51, pp. 53-72.
- GARCÍA, E., 1988, *Modificaciones al sistema de clasificación climática de Köppen*. (Mexico: Ind).

- GOWARD, S., MASEK, G.J., WILLIAMS, L.D., IRONS, R.J. and THOMPSON, J.R. 2001, The Landsat 7 mission 7 terrestrial research and applications for 21st century. *Remote Sensing of Environment*, 78, pp. 3-12.
- HENGL, T. and ROSSIETER, G.D., 2003, Supervised landform to enhance and replace photo-interpretation in semi-detailed soil survey. *Soil Science Society of America Journal*, 67, pp. 1810-1822.
- INSTITUTO NACIONAL DE ESTADÍSTICA, GEOGRAFÍA E INFORMÁTICA (INEGI)., 1988, *Atlas nacional del medio físico*. Aguascalientes, México.
- INSTITUTO NACIONAL DE ESTADÍSTICA, GEOGRAFÍA E INFORMÁTICA (INEGI)., 2007, *Sistema de Descarga del Continuo de Elevaciones Mexicano*. [online] Available from: <http://www.inegi.gob.mx/inegi/default.aspx> (Accessed 4 April 2007).
- JENSEN, R.J. QIU, F. and JI, M., 1999, Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. *Internacional Journal of Remote Sensing*, 20, pp. 2805-2822.
- KRASILNIKOK, V.P. and TABOR, A.J., 2003, Perspectives on utilitarian ethnopedology. *Geoderma*, 111, 197-215.
- LICONA, V.A., ORTIZ, S.C.A. and PÁJARO H., D., 1993. El uso de la fotointerpretación en la cartografía de clases de tierra campesina. *Revista Geografía Agrícola*, 18, pp. 85-93.
- LLEVERINO, G. E., ORTIZ, S.C.A. and GUTIÉRREZ C., Ma. del C., 2000, Calidad de los -mapas de suelos en el ejido de Atenco, estado de México. *Terra*, 18, pp. 103-113.
- LO, P. C., and YEUNG, W. A. K., 2007, *Concepts and techniques in Geographic Information Systems*. Upper Saddle River, NJ: Pearson Prentice Hall.
- MACBRATNEY, B.A., MENDONÇA-SANTOS, L.M. and MINASNY, B., 2003, On digital soil mapping. *Geoderma*, 117, pp. 3-52.
- MARTÍNEZ, M.J.F., 1993, *El uso de Fotografías Aéreas e Imágenes de Satélite en la Cartografía de Tierras*. MSc thesis, Colegio de Postgraduados, México.
- MARTÍNEZ, M.J.F., ORTIZ, S.C.A., PÁJARO, H.D., OJEDA, T. E. and VARGAS, P.E., 2003, Clave de fotointerpretación para la cartografía de tierras campesinas a escala regional. *Terra*, 21, pp. 301-309.
- MINITAB. 2003. *Statistical Software*. State College PA, USA.

- MORAN, J.C. and BUI, N.E., 2002, Spatial data mining for enhanced soil map modeling. *International Journal Information Geographical Science*, 16, pp. 533-549.
- NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA) Goddard Space Flight Center (GSFC), 2007, *The Landsat-7 Science Data User's Handbook*. Greenbelt, MD. Available online at: http://landsathandbook.gsfc.nasa.gov/handbook/handbook_toc.html (Accessed 28 August 2007).
- NORDT, L.C., JACOB, S.J. and WILDING, P.L., 1991, Quantifying map unit composition for quality control in soil survey. In *Spatial variabilities of soils and landforms*, M. J. Mausbach and L. P. Wilding (eds.), pp. 183-197. Madison, WI: Soil Science Society of America.
- ORTIZ S., C. A., and GUTIÉRREZ C., MA. DEL C., 2000. Evaluación taxonómica de sistemas locales de clasificación de tierras. *Terra*, 17, pp. 277-286.
- ORTIZ, S.C. A., PÁJARO, H.D. and ORDAZ, CH.V.O., 1990, *Manual para la cartografía de clases de tierra campesinas*. (Montecillo, México: Serie de Cuadernos de Edafología).
- ORTIZ-SOLORIO, C. A. and GUTIÉRREZ-CASTORENA, MA. DEL C., 2001, Mexican Ethnopedology: the Experience in Soil Mapping. Science and Tradition: *Roots and Wings for Development*, pp. 107-136.
- ORTIZ-SOLORIO, C.A., GUTIÉRREZ-CASTORENA, MA. DEL C., LICONA-VARGAS, A.L. and SÁNCHEZ-GUZMÁN, P., 2005, Contemporary influence of indigenous soil (land) classification in Mexico. *Eurasian Soil Science*, 38, pp. S89-S94.
- PÁJARO, H.D., and ORTIZ, S.C.A., 1987, *El levantamiento de suelos y su relación con la clasificación y cartografía de las clases de tierras campesinas*. (Chapingo, México: Colegio de Postgraduados).
- PAYTON, W.R., BARR, F.J.J., MARTIN, A., SILLITOE, P., DECKERS, F.J., GOWING, W.J., HATIBU, N., NASEEM, B.S., TENYWA, M. and ZUBERI, I.M., 2003, Contrasting approaches to integrating indigenous knowledge about soil and scientific soil survey in East Africa and Bangladesh. *Geoderma*, 111, pp. 355-386.
- PENG, W., WHEELER, B.D., BELL, C.J. and KRUSEMARK, M.G., 2003, Delineating patterns of soil drainage class on bare soils using remote sensing analyses. *Geoderma*, 115, 261-179.

- PERSSON, M., SIVAKUMAR, B., BRENDTSSON, R., JACOBSEN, H.O. and SCHJØNNING, P., 2002, Predicting the dielectric constant–water content relationship using artificial neural Networks. *Soil Science Society of America Journal*, 66, pp. 1424–1429.
- QI, F., and ZHU, A.X., 2003, Knowledge discovery from soil maps using inductive learning. *International Journal Information Geographical Science*, 17, pp. 771-795.
- RZEDOWSKI, J. and HUERTA, L., 1985, *La vegetación de México*. (Mexico, DF: Limusa).
- SÁNCHEZ G.P., ORTIZ, S.C.A., GUTIÉRREZ C. MA. DEL C. and GÓMEZ., D.J.D., 2002, Clasificación campesina de tierras y su relación con la producción de caña de azúcar en el sur de Veracruz. *Terra*, 20, pp. 359- 369.
- SERPICO, S., BRUZZONE, L. and ROLI, F., 1996, An experimental comparasion of neural statistical non-parametric algorithms for supervised classification of remote sensing images. *Pattern Recognition Letters*, 17, pp. 1331-1341.
- SHANER, J. and WRIGHTSELL, J., 2000, *Editing in ArcMap. Environmental Systems Research Institute (ESRI)*. New York Street, Redlands, CA, USA.
- SHUPE, M.S. and MARSH, E.S., 2004. Cover and density-based vegetation classifications of the Sonoran Desert using Landsat TM and ERS-1 SAR imagery. *Remote Sensing of Environment*, 93, pp. 131-149.
- TRUEBA E., A., OROPEZA M., J. L., ORTIZ S., C. A., MARTÍNEZ A., A., and RUELAS A., G., 2004, Identificación de zonas erosionadas mediante el tratamiento de imágenes digitales con una Red Neuronal. *Agrociencia*, 38, pp. 573-581.
- YOUNG, H.E. and STOECKELER, G.E., 1956, Quantitative evaluation of photo interpretation mapping. *Photogrammetric Engineering*, 22, pp. 137-143.
- ZHU, X.A., HUDSON, B., BURT, J., LUBICH, K. and SIMONSON, D., 2001, Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal*, 65, pp. 1463-1472.

Table 1. Precision of the maps of the three regions of study.

Map	Region		
	Arid [†] (%)	Temperate [‡] (%)	Tropical [£] (%)
MAXLIKE and Refl+DEM	42*	62*	84*
MAXLIKE and Reflectances	29	47	49
MINDIST and Ref+DEM	14	47	44*
MINDIST and Reflectances	9	47	38*
MINDIST and DEM	14	31	43*
PIPED and Ref+DEM	22	37	63
PIPED and Reflectances	3	22*	18
PIPED and DEM	21*	19*	71
ANN and Ref+DEM	23	29	80*
ANN and Reflectances	10	18*	68
ANN and DEM	18	35	80*

*Significant difference of binomial proportion test ($\alpha=0.05$). Confidence interval: †18.64 ± 10.54; ‡35.82 ± 12.97; £58.00 ± 13.40.

Table 2. Accuracy rate of the training of three ANNs.

Region	Map	Total pixels	Tr/Ts*	Accuracy rate (%)
Arid	ANN and DEM	98295	80/80	51
Temperate	ANN and Reflectances	8306	60/60	20
Tropical	ANN and DEM	28262	30/30	97

*Tr= training sample; Ts= testing sample

Table 3. Map boundaries accuracy of the three study regions.

Map	Region		
	Arid [†] (%)	Temperate [‡] (%)	Tropical [£] (%)
MAXLIKE and Refl+DEM	6	16*	17*
MAXLIKE and Reflectances	5	8	1
MINDIST and Ref+DEM	5	9	1
MINDIST and Reflectances	2	9	10
MINDIST and DEM	7	6	1
PIPED and Ref+DEM	6	4	1
PIPED and Reflectances	2	8	1
PIPED and DEM	12*	0*	1
ANN and Ref+DEM	1	0*	22*
ANN and Reflectances	2	0*	1
ANN and DEM	6	0*	30*

*Significant difference of binomial proportion test ($\alpha=0.05$). Confidence interval: [†]4.91 ± 5.84; [‡]5.45 ± 6.15; [£]7.82 ± 7.27.

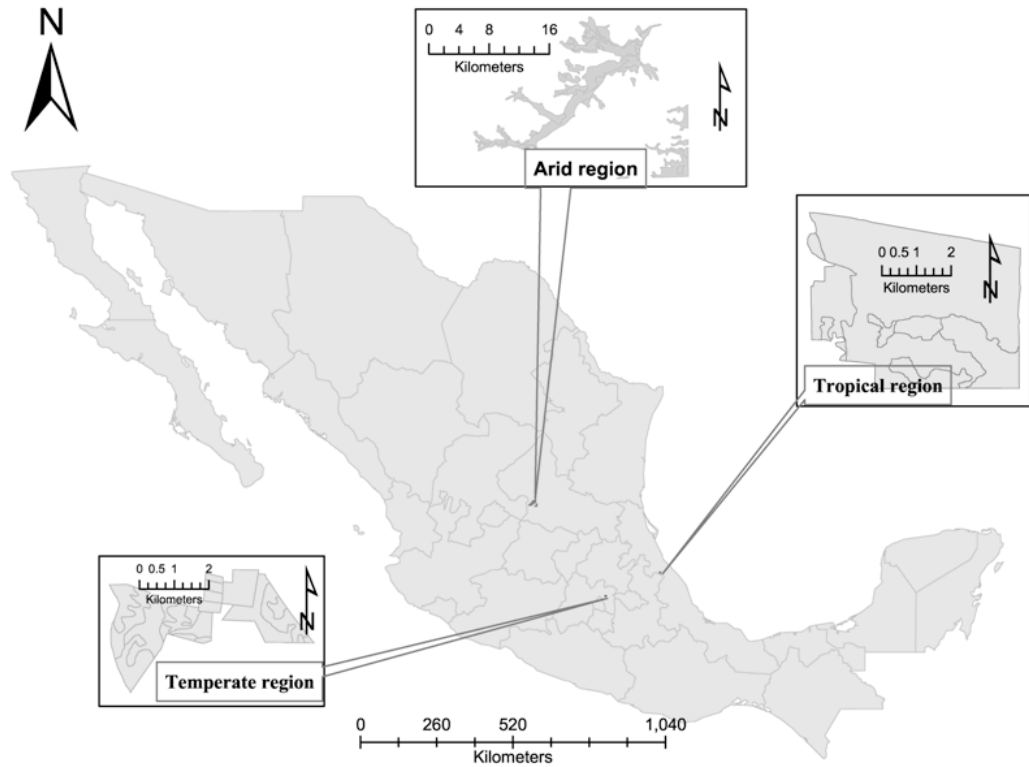


Figure 1. Regions of Study.

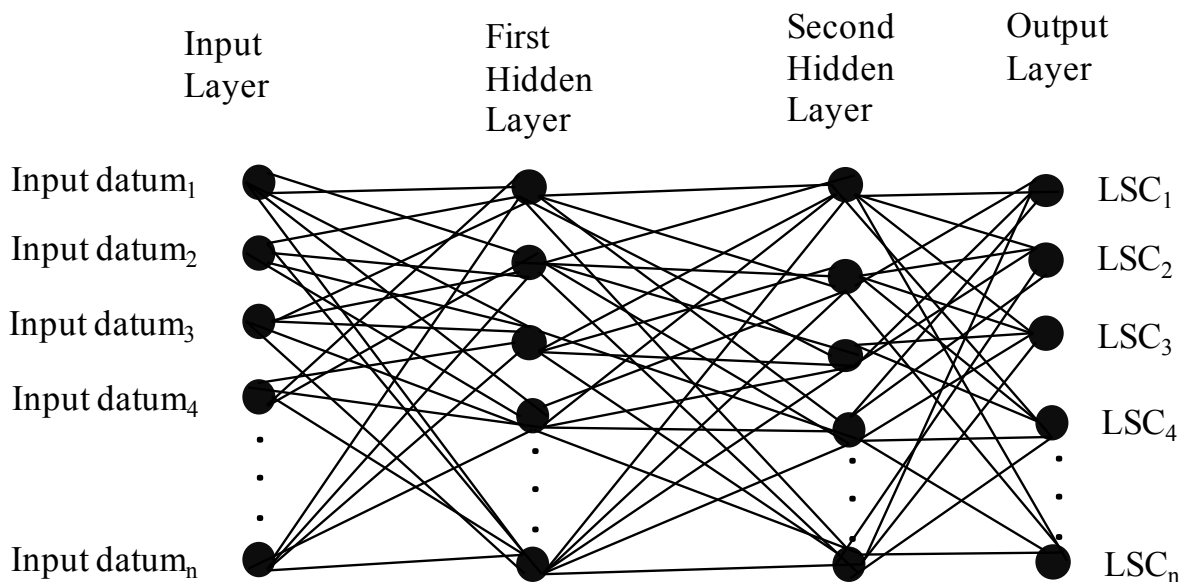


Figure 2. ANNs Architecture. LCS: Local Soil Class.

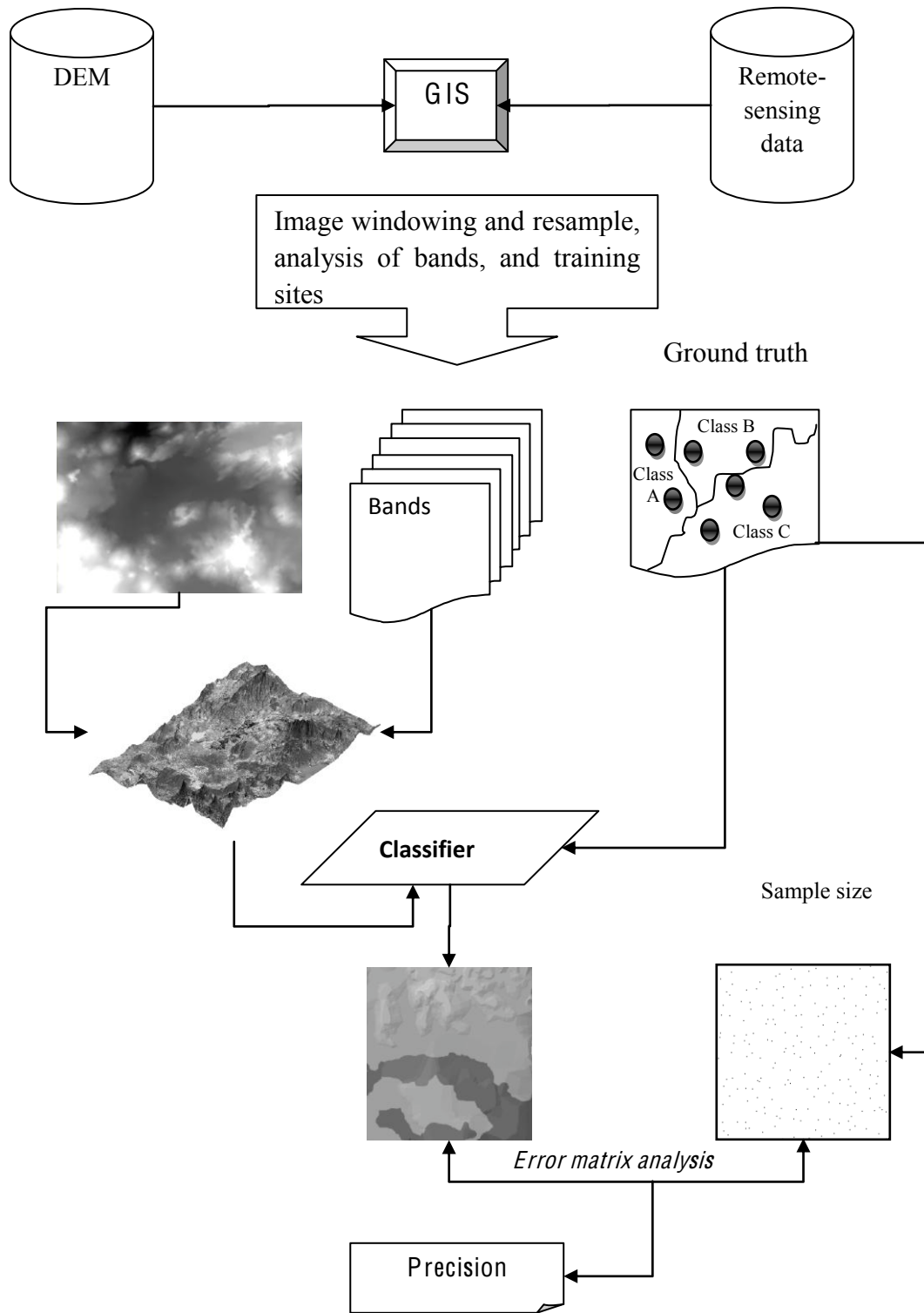


Figure 3. Evaluation of precision.

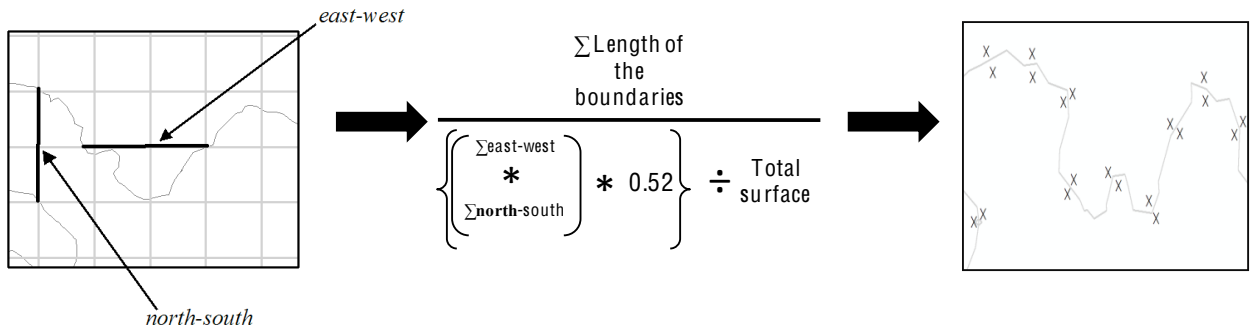


Figure 4. Evaluation of boundaries accuracy.

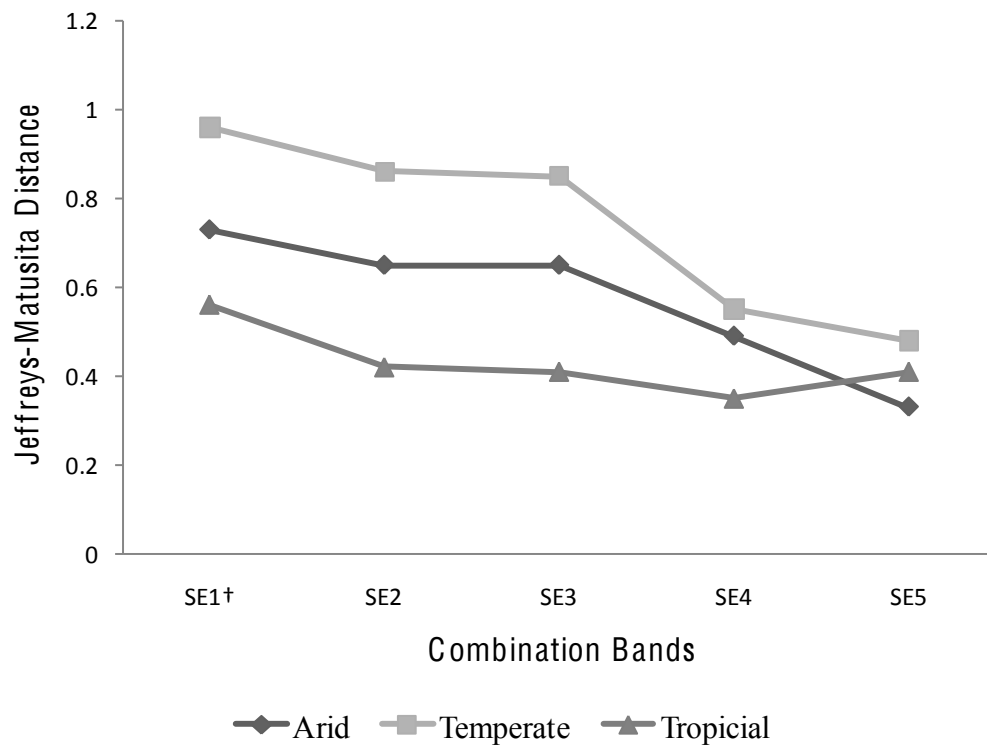


Figure 5. Jeffreys-Matusita distance of local soil classes spectral signatures. SE1= 123457 (bands); SE2= 12347; SE3= 12345; SE4= 247; SE5= 457.

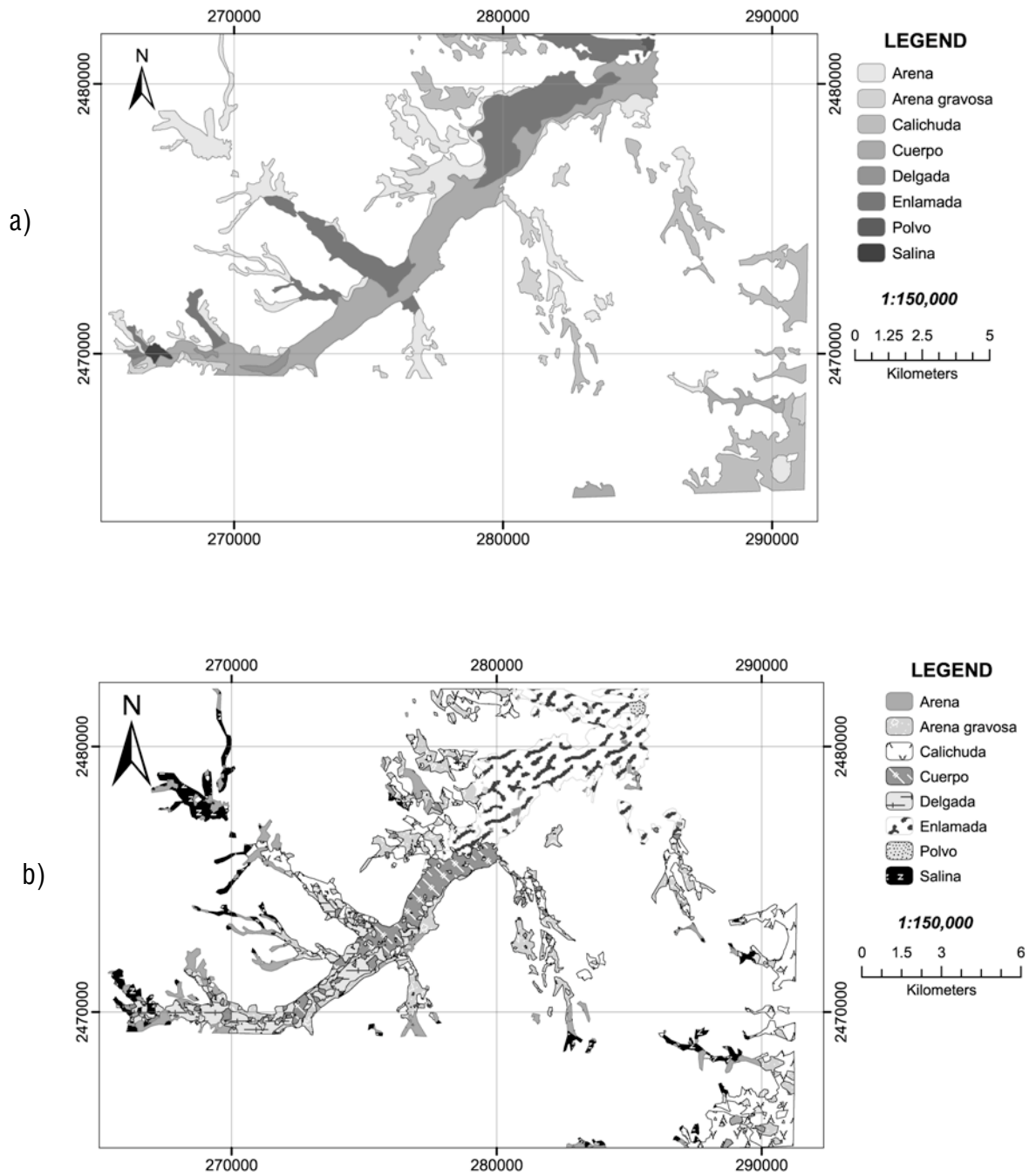


Figure 6. Arid region; a) map by Martinez et al. (2003), b) Map 1 (Maximum likelihood with reflectances and DEM) of SLP.

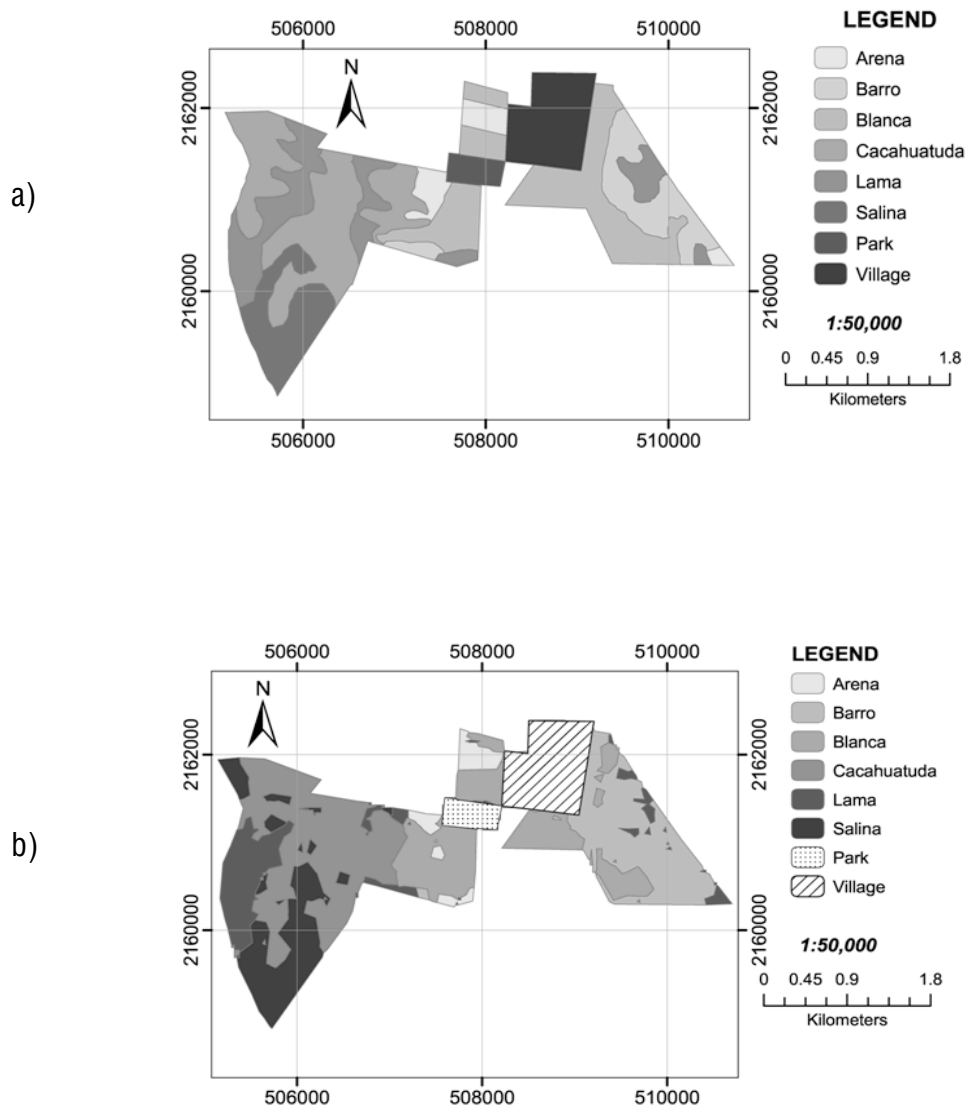


Figure 7. Temperate region; a) map by Pajaro and Ortiz (1987), b) Map 1 (Maximum likelihood with reflectances and DEM) of the state of Mexico.

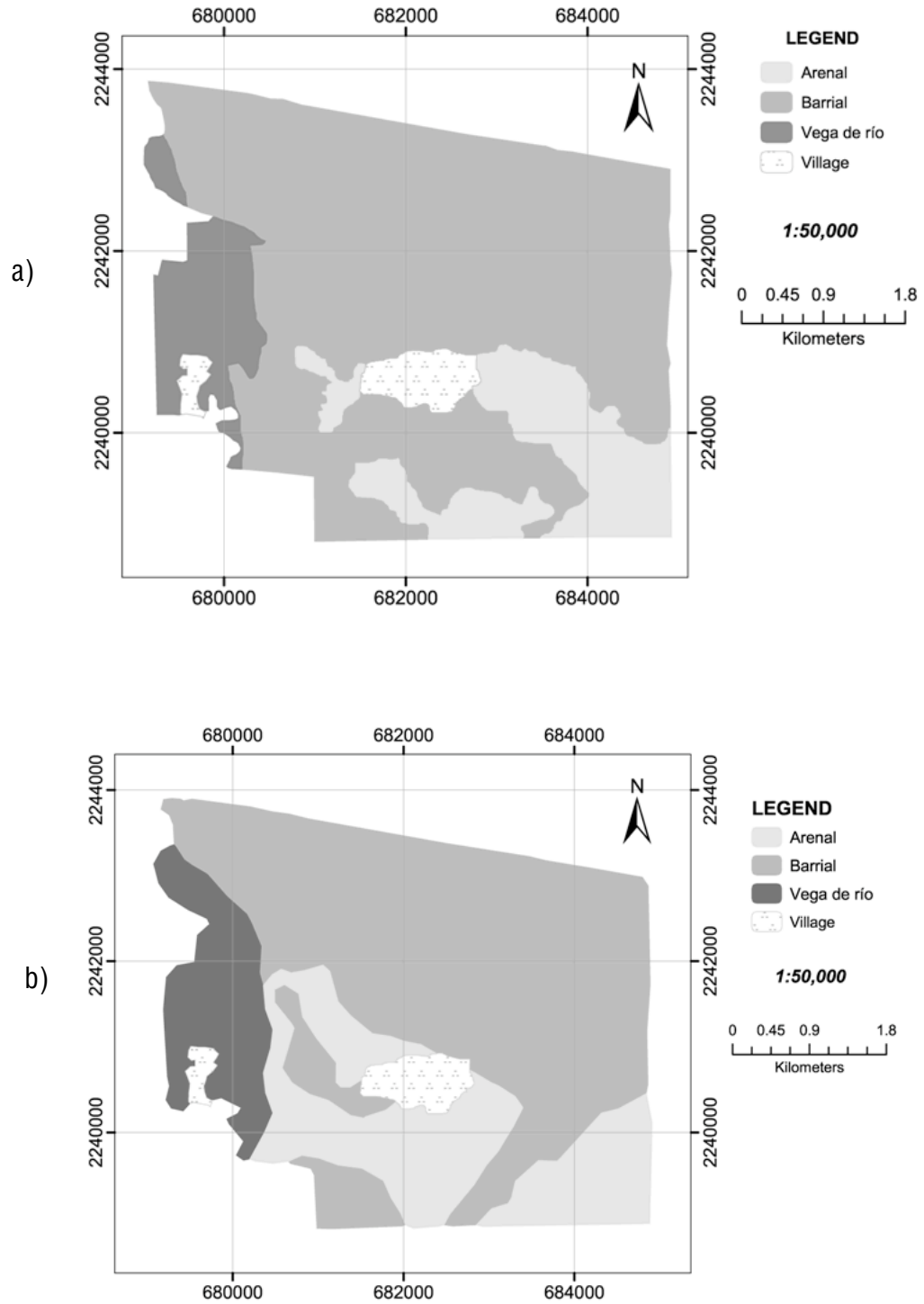


Figure 8. Tropical region; a) map by Cruz (2006), b) Map11 (ANN and DEM) of Veracruz.

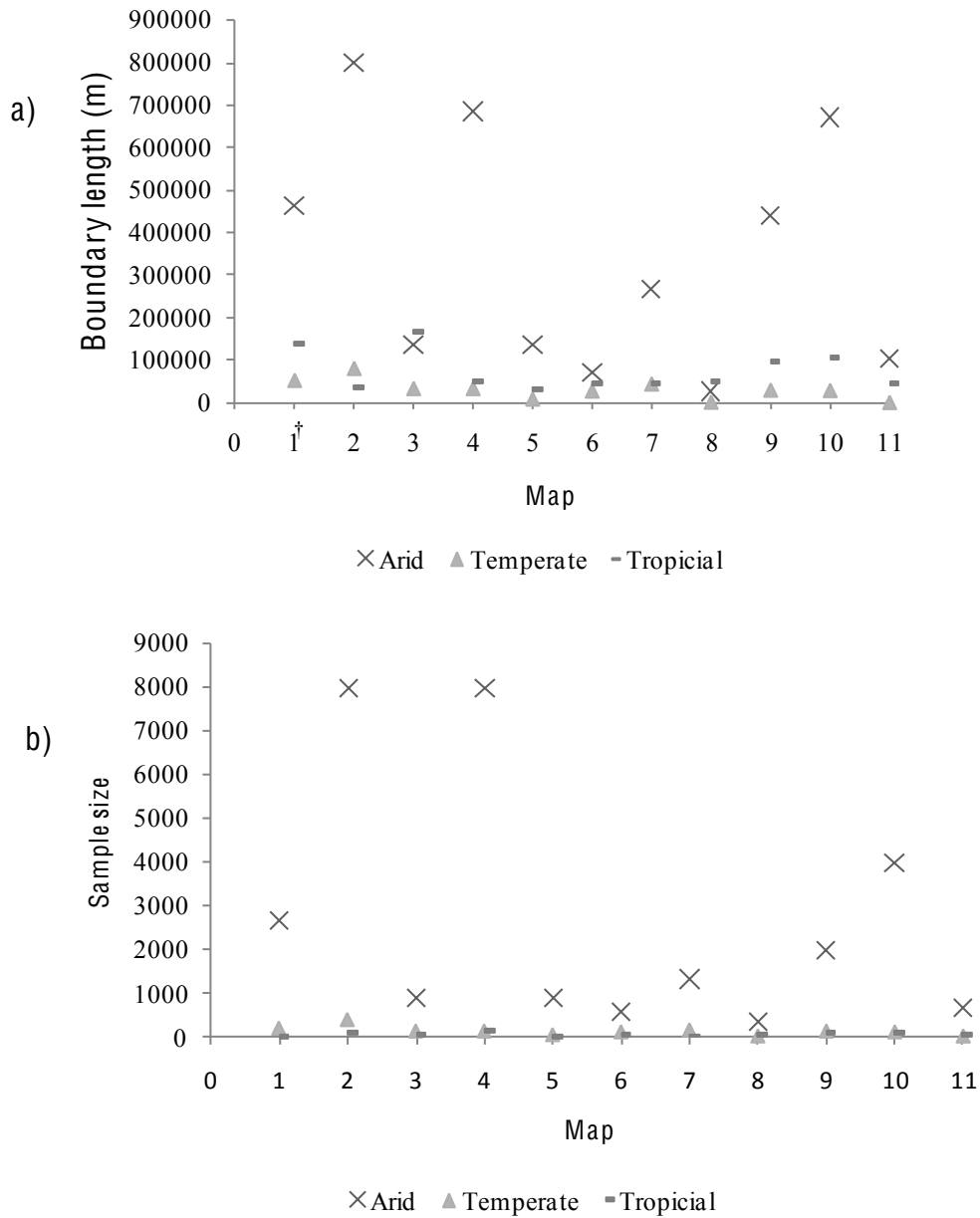


Figure 9. Total boundary length (a) and sampling points (b) to evaluate the accuracy of the maps of the three regions. Classifier and input data of the maps generated with CAC: (1) MAXLIKE and Reflectances+DEM, (2) MAXLIKE and Reflectances, (3) MINDIST and Reflectances + DEM, (4) MINDIST and Reflectances, (5) MINDIST and DEM, (6) PIPED and Reflectances + DEM, (7) PIPED and Reflectances, (8) PIPED and DEM, (9) ANN and Reflectances + DEM, (10) ANN and Reflectances, (11) ANN and DEM.

CAPITULO III
DIGITAL MAPPING OF FARMLAND CLASSES

DIGITAL MAPPING OF FARMLAND CLASSES¹

G. Cruz-Cárdenas^a, C.A. Ortiz-Solorio^a, E. Ojeda-Trejo^a, J.F. Martínez-Montoya^b, E.D. Sotelo-Ruiz^c, and A.L. Licona-Vargas^d

^a*Colegio de Postgraduados, Campus Montecillo, Carretera México-Texcoco km 36.5, 56230, Texcoco, México, México.* ^b*Colegio de Postgraduados, Campus San Luis Potosí, Iturbide 73, 78620, Salinas de Hidalgo, Salinas, SLP, México.* ^c*Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Campo Experimental Zinacantepec, Carretera Zacatepec-Galeana km 0.5, 62780, Zinacantepec, México, México.* ^d*Universidad Autónoma Chapingo, CRUO, km 3 Carretera Huatusco-Jalapa, 94100, Huatusco, Veracruz, México.*

Resumen

La cartografía de clases de tierra campesinas permite generar mapas de tierras utilizando una metodología basada en el conocimiento local de manera rápida y a un costo bajo y con un mayor número de unidades cartográficas que los levantamientos de suelos convencionales. Sin embargo, los resultados encontrados al producir estos mapas con técnicas de cartografía automatizada son contrastantes. Se evaluó la precisión y la exactitud de linderos de 324 mapas de clases de tierra campesina (CTC) generados por computadora, mediante la aplicación del interpolador de Inverso de la Distancia (IDW). Estos mapas se obtuvieron al variar el tamaño de la muestra para el entrenamiento, el diseño espacial de la misma y el valor del Poder del interpolador. Además, de cuantificar el esfuerzo que se necesita para obtener mapas de confiabilidad aceptable. El procedimiento se aplicó a mapas de CTC obtenidos mediante recorridos con productores de tres zonas ambientalmente contrastantes de la república mexicana. Los resultados muestran que el mejor esquema de muestreo en las tres aéreas es el sistemático y el Poder de 8, ello con la obtención de los mapas con mayor confiabilidad. A través del criterio de la confiabilidad del mapa y al esfuerzo que representa dicho muestreo, el tamaño de la muestra recomendable es de 10% a 25% del total de parcelas.

Palabras Clave: Precisión y exactitud de linderos cartográficos, tamaño de parcela, interpolador IDW, tiempo de muestro.

¹Submitted in *Communications in Soil Science and Plant Analysis*, with Editor.

Abstract

The cartography of farmland classes allows to generate soil maps, using a methodology based on local knowledge, rapidly and at low cost, and with a greater number of cartographic units than maps based on conventional soil surveys. However, the results found when producing these maps with automated cartography techniques are contrasting. Precision and accuracy were evaluated in 324 computer generated farmland class (FLC) maps by applying the Inverse Distance Weighted (IDW) interpolation model. These maps were obtained by varying the sample size for the training, its spatial design, and the Power value of the interpolator. Moreover, the effort needed to obtain maps with acceptable reliability was quantified. The procedure was applied to FLC maps obtained from surveys with producers from three contrasting environmental zones in Mexico. The results show that the best sampling scheme in the three areas is the systematic sampling, and Power 8, giving the maps with the highest reliability. Through the criterion of map reliability and effort needed for sampling, the recommended sample size is 10% to 25% of the total plots.

Key words: Precision and accuracy, plot size, IDW interpolator, sampling time.

1. Introduction

A farmland class is defined as a specific land area that includes all the directly or indirectly observable attributes of the biosphere, in time or space, and which are affected by their use or handling (Ortiz-Solorio *et al.* 2005). Diverse studies on FLC have shown that it is a good alternative to relate them to physical and chemical soil properties (technical concept) and their formation factors, as well as color, texture, drainage, agricultural practices, type of vegetation, and crop (Ericksen and Ardon 2003, Barrera-Bassols *et al.* 2006, Licona-Vargas *et al.* 2006, Cruz *et al.* 2008). Also, it is a rapid, inexpensive methodology which does not require high specialization of the personnel in cartography, as opposed to technical soil surveys (Ortiz 1999). The maps generated under this approach have a high degree of precision and accuracy, as mentioned by Lleverino *et al.* (2000). Also, the cartographic units delimited are more detailed than the Subunit or Subgroup levels of the World Reference Base or Soil Taxonomy, respectively (Ortiz-Solorio *et al.* 2005). With regard to digital mapping of FLC, some studies have been done to automate cartography, with contrasting results. For example, Martinez (1993) and Ortiz (1999) mention that digital mapping of FLC cannot be done since the classes

cannot be satisfactorily identified. On the contrary, Segura *et al.* (2004) got 80% reliability in their FLC map, but only when grouping land classes in the groups with and without residual moisture. Therefore, there's still to be found an automated technique to generate FLC maps with acceptable reliability.

Some factors taken into account to generate high quality computer assisted soil maps (CALM) are: a) sample size to do the classification, b) spatial design of the sampling scheme, and c) the configuration of the interpolator or classifying algorithm, specifically regarding Power with the IDW model.

Sample size is an important factor to carry out the classification since the precision of each class and global map precision depend on it (Foody and Mathur 2006). In some cases, a value determined as 30ρ is taken, meaning 30 pixels times the number of bands or layers (ρ) that intervene in the classification. In other cases, it is established based on statistical models (Foody *et al.* 2006, Carré *et al.* 2007). An exploration can also be done determining percentages, for example, Grinand *et al.* (2008) analyzed sample sizes to generate CALM, from 10 to 90 % of the total area, and found that there are no changes in the map precision after 30%. Regarding spatial design of the sampling, Hengl *et al.* (2003) pointed out that by graphing two principal components (topographical variables) on an X – Y plane (UTM coordinates) the best spatial arrangement or sampling design is obtained. On the other hand, Moran and Bui (2002) recommend the Area-Weight method, similar to a random design, but unlike the random design, it takes into account all classes, this is, the sample number per class is proportional to the area occupied by each one. Finally, the configuration of the interpolator or classifying algorithm affects the outline of the resulting maps. In the IDW model, Power plays an important role in the reliability of the created map. Robinson and Metternicht (2006) state that the best maps are obtained using Power 1, but Kravchenko and Bullock (1999) say that it is so with Power 4.

The main goal of this work is to create a methodology to generate high quality computer assisted FLC maps. The following specific objectives were established: 1) to evaluate the factors that intervene in the generation of computer assisted soil maps in digital mapping of farmland classes; 2) to quantify the sampling time needed to obtain maps with acceptable reliability.

2. Materials and Methods

2.1 *Study zones*

Three study zones were selected, with different climatic, lithologic, and topographic conditions. The first zone is located in the municipality of Villa Hidalgo, Zacatecas: extreme coordinates 101°45' WL and 22°24' NL, 101°42' WL and 22°18' NL; climate type BS₁kw, semi-arid with mean yearly temperature between 12 °C and 18 °C; temperature during the coldest month varies from -3 °C to 18 °C; temperature during the warmest month is below 22 °C; with summer rains, and winter rain percentage from 5% to 10.2% of the total annual rainfall (Garcia 1988); parent material is mainly sedimentary rock such as sandstone, conglomerate, and lutite (INEGI 1988); average slope is less than 5%.

The second zone is located in the municipality of Texcoco, Mexico: extreme coordinates 98°57' and 98°53' WL, 19°33' and 19°30' NL; climate type C (wo), temperate, sub-humid mean annual temperature from 12 °C to 18 °C; temperature during the coldest month from -3 °C to 18 °C; temperature during the warmest month is below 22 °C; rainfall during the driest month less than 40 mm; summer rains with a P/T index below 43.2; winter rain percentage from 5% to 10.2% of the total annual rainfall (Garcia 1988); it is a lacustrine alluvial plain; its geology corresponds to the Quaternary (INEGI 1988); slope is less than 2%.

The third zone is located in the municipality of Papantla, Veracruz: extreme coordinates 97°17' WL and 20°17' NL, 97°10' WL and 20°14' NL; climate type Am (f), warm humid, mean annual temperature over 22 °C, temperature of the coldest month over 18 °C; rainfall during the driest month less than 60 mm; summer rains, winter rain percentage over 10.2% of the total annual rainfall (Garcia 1988); lithology is made up of lutites, sandstone, alluvial sediments and extrusive acid igneous rocks (INEGI 1988); in the plains, the slope is 2%, and in the hillsides the slope varies from 14% to 72%.

2.2 *Cartography of the farmer land classes*

The FLC maps for each zone were generated through the methodology by Ortiz *et al.* (1990), which consists of 1) selecting a base map on which boundaries are drawn, 2) surveys are made with the farmers of the study zone, and 3) the following questions are made: Where does land class change? And how is it different from neighboring ones? As a rule, it is always located with the farmers in the sites where land changes occur, which is equal to verifying the

boundaries. According to the landscape survey theory, this last step makes it a detailed land survey (Ortiz 1999).

2.3 Sample size and sampling scheme

As mentioned before, there are several ways to determine the sample size for training and sampling scheme. In this study, the number of points for the interpolation was determined based on proportions of the total land surface with the following percentages: 1, 5, 10, 15, 20, 25, 30, 40, and 50. Moreover, in order to find the best sampling scheme for each study zone, the three most common schemes were used: random, systematic, and random-stratified.

2.4 Configuration of the classifier

The IDW model calculates the weight of the values according to the inverse relationship of the distance with the following equation (Lloyd 2007):

$$\hat{Z}(X_0) = \frac{\sum_{i=1}^n Z(X_i) * d_{i0}^{-r}}{\sum_{i=1}^n d_{i0}^{-r}}$$

(1)

Where the prediction of X_0 is a function of n neighboring observations, $Z(X_i)$, $i = 1, 2, \dots, n$, r is an exponent that determines the weight assigned to each observation, and d is the separation distance between X_0 and X_i . As the distance between these two points increases the weight decreases. As the distance decreases, the weight increases. An important parameter of this model is the value of the exponent, or Power, where 2 is the most common value. Although, according to Gotway *et al.* (1996), the accuracy of the predictions with IDW increases as the Power increases. Because of this, in this research different Power values were used: 1, 2, 4, and 8.

2.5 Computer assisted farmer land class maps (CAFLCM)

For each study zone 108 CAFLCM were generated, resulting from the application of the variations of each of the three factors considered (sections 2.3. and 2.4.): Nine percentages for sample size, three sampling schemes, and four Power values (9x3x4). Thus, 324 CAFLCM were generated.

2.6 Evaluation of map reliability

One hundred pixels were considered for each farmer land class to evaluate global precision of the CAFLCM with the confusion matrix (Congalton 1991). However, it is also important to evaluate the location of the classes, which is directly related to the accuracy of each farmland class. For this, the Klocation index, widely described by Pontius (2000), was used. The sample size to evaluate accuracy was the same as was used for precision. The sampling scheme used was random-stratified since it gives satisfactory results when evaluating map reliability (Congalton 1988). Figure 1 shows the general methodology used.

2.7 Sample size determination by plot size

The definition of the sample size for the training in the generation of CAFLCM was done according to the number of pixels that have to be taken. However, this amount depends on the size to which the pixels are configured, which may not be practical in the field. Therefore, a second option is to consider the size of the plot to determine the size of the sample. The average plot size in Villa Hidalgo and Texcoco was 2 ha, and in Papantla it was 12 ha. The plot size was divided in the surface area of each zone to obtain the sample size (100% of the sampled plots). Moreover, this value was divided by 2, 4, and 10 to determine the sampling points shown in Table 1. The sampling scheme and Power defined in sections 2.3. and 2.4. were used. The evaluation of reliability of the maps was done as mentioned in section 2.6.

2.8 Determination of sampling time

To calculate the time that it would take to carry out the sampling according to the proportion of visited plots, the following formula was generated:

$$TP = S \times \sqrt{PS}$$

(2)

Where TP is the time needed to traverse two consecutive points (hours), S is the average walking speed of a person (km/hr), in this case 5, PS is the average size of the plot (km²).

$$ST = (TP + T)(SS - 1)$$

(3)

Where ST is the sampling time (hours), T is the time spent stationary at a single point (hours), under the assumption that the farmer is only asked what land class that specific point belongs

to, and some other characteristics, it takes 10 minutes or 0.167 hours as a constant, $SS-1$ is the sample size, not considering the time needed to reach the first point.

3. Results and Discussion

The number of farmland classes varies in each area. In Villa Hidalgo, Zacatecas, there are five classes: Canelona, Chautosa, Colorada, Parda, and Pardusca on a surface area of 1831 ha (Figure 2a). In Texcoco, Mexico, there are eight classes: Arena, Barro, Blanca, Cacahuatuda, Jaboncillo, Lama, Pantano, and Salina in 4174 ha (Figure 2b). Lastly, in Papantla, Veracruz, there are three classes: Arenal, Barrial, and Vega de Río in 3462 ha (Figure 2c).

The behavior of precision and accuracy of the CAFLCM is similar for both parameters measured in all three study zones (Figure 3). The best sampling scheme was the systematic scheme for both precision and accuracy: 89% and 86% in Villa Hidalgo; 93% and 92% in Texcoco; and 98% and 95% in Papantla. In none of the three zones were precision and accuracy less than 80% using this sampling design.

The sample size for training of the interpolator does have an influence on precision and accuracy of the CAFLCM (Figure 4). In Villa Hidalgo, we recommend 10% of the total pixels (2265) to do the interpolation, since the results obtained are similar to those of a 50% sample size (11325). In Texcoco, 15% of the total pixels (7716) is enough to carry out the computer assisted classification, obtaining over 95% reliability of the maps. Nevertheless, 5% of the total pixels (2572) is also acceptable, since the CAFLCM would have over 85% precision and accuracy. Like in the case of Villa Hidalgo, the CAFLCM of Papantla can be done with 10% of the sample (4262) to feed the interpolator and obtain maps with acceptable reliability. Also, like in the case of Texcoco, a smaller sample can be used; in this case 1% of the total pixels (426) and the maps would have over 85% reliability. On the other hand, Moran and Bui (2002) obtained these same results of precision and kappa index with 50% sampling density and using a decision tree as classifying algorithm. Using this same sample size, Grinand *et al.* (2008) obtained 65% and 63% precision and kappa index in their land maps.

The influence of the Power on the reliability of the CAFLCM varies with each study region (Figure 5). For Villa Hidalgo and Texcoco the difference between Power 1 and Power 8 is more than 6% and 8%, respectively. Contrarily, the CAFLCM of Papantla had a difference below 1%. Generally speaking, the best maps were obtained using Power 8, unlike what was

found by Robison and Metternicht (2006), whose land maps with the greatest reliability were those using Power 1, as compared to Power 2, 3, and 4. Then again, Kravchenko and Bullock (1999) obtained their best maps using Power 4, followed by Power 1, 3, and 2.

To determine the sample size according to the average plot size that will generate CAFLCM with acceptable reliability, the systematic sampling design and Power 8 were chosen for the interpolation. The results of this analysis are shown in Table 2.

In Texcoco and Papantla, with 10% of the sampled plots, 209 and 29 respectively, map precision was over 80% and accuracy was over 65%. Contrastingly, in Villa Hidalgo only global precision (62%) is acceptable with 92 points, and accuracy barely reaches 53%. In general, no more than 300 points are necessary to obtain CAFLCM with over 75% precision and 70% accuracy. On the other hand, Foody and Mathur (2006) recommend 90 points for each class. Therefore, considering this recommendation the sampling size in Villa Hidalgo would be 450 points, in Texcoco 720 points, and in Papantla 270 points. For this study, it is equal to 49%, 35%, and 94% of the total sampled plots for each zone, respectively.

The effort required increasing precision and accuracy of the CAFLCM of Villa Hidalgo from 62% and 53% to 75% and 72% is 11 and 7 hours, respectively, for each percentage point. However, in order to increase from 75% and 72% to 84% and 79%, the effort increases to 19 and 25 hours. Likewise, to reach maximum precision and accuracy when sampling all the plots, the sampling time is 44 and 33 hours for each percentage point in each parameter. The same behavior is true for the other two zones, this is to say, the effort needed to increase precision and accuracy increases as it nears 100% of the sampled plots for both criteria of map reliability (Table 2).

4. Conclusions

Automated cartography of farmland classes was applied in areas with different environmental conditions and local farmer land classes, which allows to evaluate the methodology used under contrasting conditions. An important factor in the sampling design was to consider plot size, which is related with their handling, which in turn has influence on the identified farmland class. This can be observed in the distribution of their boundaries.

In general, the recommended sample size varies from 15 to 20% of the sampled plots to obtain maps with 68% to 80% reliability. This would take from 22 to 45 days (8-hour work days) in

areas from 2000 to 4000 ha, all in function of the plot size. The best sampling scheme was the systematic scheme, and Power 8 for all three zones.

5. References

- BARRERA-BASSOLS, N., ZINCK, A.J. and VAN RANST, E., 2006, Local soil classification and comparison of indigenous and technical soil maps in a Mesoamerican community using spatial analysis. *Geoderma*, 135, pp. 140-162.
- CARRÉ, F., MACBRATNEY, B.A. and MINASNY, B., 2007, Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. *Geoderma*, 141, pp. 1-14.
- CONGALTON, G.R., 1988, A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54, pp. 593-600.
- CONGALTON, G.R., 1991, A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, 37, pp. 35-46.
- CRUZ, C.G., ORTIZ, S.C.A., GUTIÉRREZ, C. MA. DEL .C. and VILLEGAS, M.A., 2008, Las clases de tierras citrícolas del ejido Pueblillo, Papantla, Veracruz. *Terra Latinoamericana*, 26, pp. 11-19.
- ERICKSEN, P.J. and ARDÓN, M., 2003, Similarities and differences between farmer and scientist views on soil quality issues in central Honduras. *Geoderma*, 111, pp. 233-245.
- FOODY, M.G. and MATHUR, A., 2006, The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, 103, pp. 179-189.
- FOODY, M.G., MATHUR, A., SANCHEZ-HERNANDEZ, C. and BOYD, S.D., 2006, Training set size requirements for the classification of a specific class. *Remote Sensing of Environment*, 104, pp. 1-14.
- GARCÍA, E., 1988, *Modificaciones al sistema de clasificación climática de Köppen*. Ind. México.
- GOTWAY, C.A., FERGUSON, R.B., HERGERT, G.W. and PETERSON, T.A., 1996, Comparison of Kriging and Inverse Distance Methods for mapping soils parameters. *Soil Science Society of American Journal*, 60, pp. 1237-1247.

- GRINAND, C., ARROUAYS, D., LAROCHE, B. AND MARTIN, P.M., 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143, pp. 180-190.
- HENGL, T., ROSSITER, D. and STEIN, A., 2003, Soil Sampling strategies for spatial prediction by correlation with auxiliary maps. *Australian Journal of Soil Research*, 41, pp. 1403-1422.
- INSTITUTO NACIONAL DE ESTADÍSTICA, GEOGRAFÍA E INFORMÁTICA (INEGI), 1988, *Atlas nacional del medio físico*. Aguascalientes, México.
- KRAVCHENKO, A. and BULLOCK, G.D., 1999, Comparative Study of Interpolation Methods for Mapping Soil Properties. *Soil Science Society of American Journal*, 91, pp. 393-400.
- LICONA-VARGAS, A.L., ORTIZ-SOLORIO, C.A., GUTIÉRREZ-CASTORENA MA. DEL C. and MANZO-RAMOS, F., 2006, Clasificación local de tierras y tecnología del policultivo café-plátano para velillo-sombra en comunidades cafetaleras. *Terra Latinoamericana*, 24, pp. 1-7.
- LLEVERINO, G.E., ORTIZ, S.C.A. and GUTIÉRREZ, C. MA. DEL C., 2000, Calidad de los mapas de suelos en el ejido de Atenco, estado de México. *Terra*, 18, pp. 103-113.
- LLOYD, D.C., 2007, *Local Models for Spatial Analysis*. CRC Press, Taylor & Francis. Boca Raton, FL.
- MARTÍNEZ, M.J.F., 1993, *El uso de Fotografías Aéreas e Imágenes de Satélite en la Cartografía de Tierras*. MSc thesis, Colegio de Postgraduados, México.
- MORAN, J.C. and BUI, N.E., 2002, Spatial data mining for enhanced soil map modeling. *International Journal Information Geographical Science*, 16, pp. 533-549.
- ORTIZ, C., A., PÁJARO, H.D. and ORDAZ, CH., V.M. 1990. *Manual para la cartografía de clases de tierra campesinas*. Serie de Cuadernos de Edafología. 15. Colegio de Postgraduados, Montecillo, México.
- ORTIZ, S.C.A. 1999, *Los levantamientos etnoedafológicos*. Ph. D thesis. Colegio de Postgraduados, Montecillo, México.
- ORTIZ-SOLORIO, C.A., GUTIÉRREZ-CASTORENA, MA. DEL C., LICONA-VARGAS, A.L. and SÁNCHEZ-GUZMÁN, P., 2005, Contemporary influence of indigenous soil (land) classification in Mexico. *Eurasian Soil Science*, 38, pp. S89-S94.

- PONTIUS, G.R., 2000, Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, 66, pp. 1011-1016.
- ROBINSON, P.T. and METTERNICHT, G., 2006, Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture*, 50, pp. 97-108.
- SEGURA, C. M.A., ORTIZ, S. C.A. and GUTIÉRREZ, C. MA. DEL C., 2004. Localización de suelos de humedad residual a partir de imágenes de satélite: Parte 2. Factores que influyen en su reflectancia y clasificación supervisada con los procedimientos: Mínima distancia y máxima verosimilitud. *Terra*, 22, pp. 135-142.

Table 1. Sampling density per plot size.

Zone	Sampling plot (%)	Points	Point/ha
Villa Hidalgo	100	916	1/2
	50	458	1/4
	25	229	1/8
	10	92	1/20
Texcoco	100	2087	1/2
	50	1044	1/4
	25	522	1/8
	10	209	1/20
Papantla	100	288	1/12
	50	144	1/24
	25	72	1/48
	10	29	1/120

Table 2. Precision and accuracy, and their effort in function of sampling density, considering plot size

Zone	Sampling plot (%)	Precision (%)	Accuracy of boundaries (%)	Sampling time (hours)
Villa Hidalgo	100	90	87	800
	50	84	79	533
	25	75	72	361
	10	62	53	219
Texcoco	100	94	94	1823
	50	91	90	1217
	25	86	84	824
	10	80	78	500
Papantla	100	94	87	545
	50	91	79	374
	25	89	74	258
	10	85	68	174

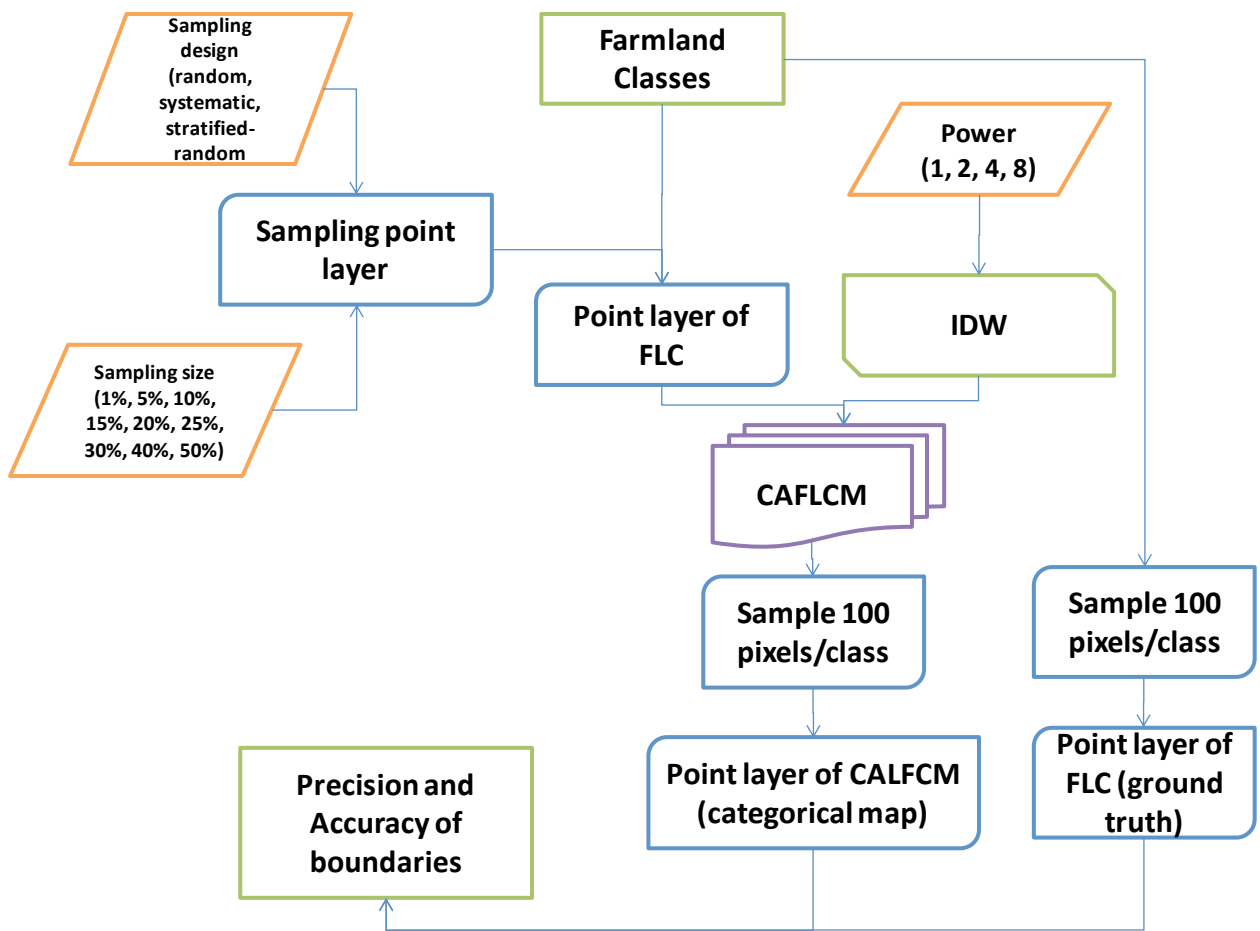


Figure 1. General scheme of methodology. Farmland Classes (FLC); Computer Assisted Farmland Classes Map (CAFLCM).

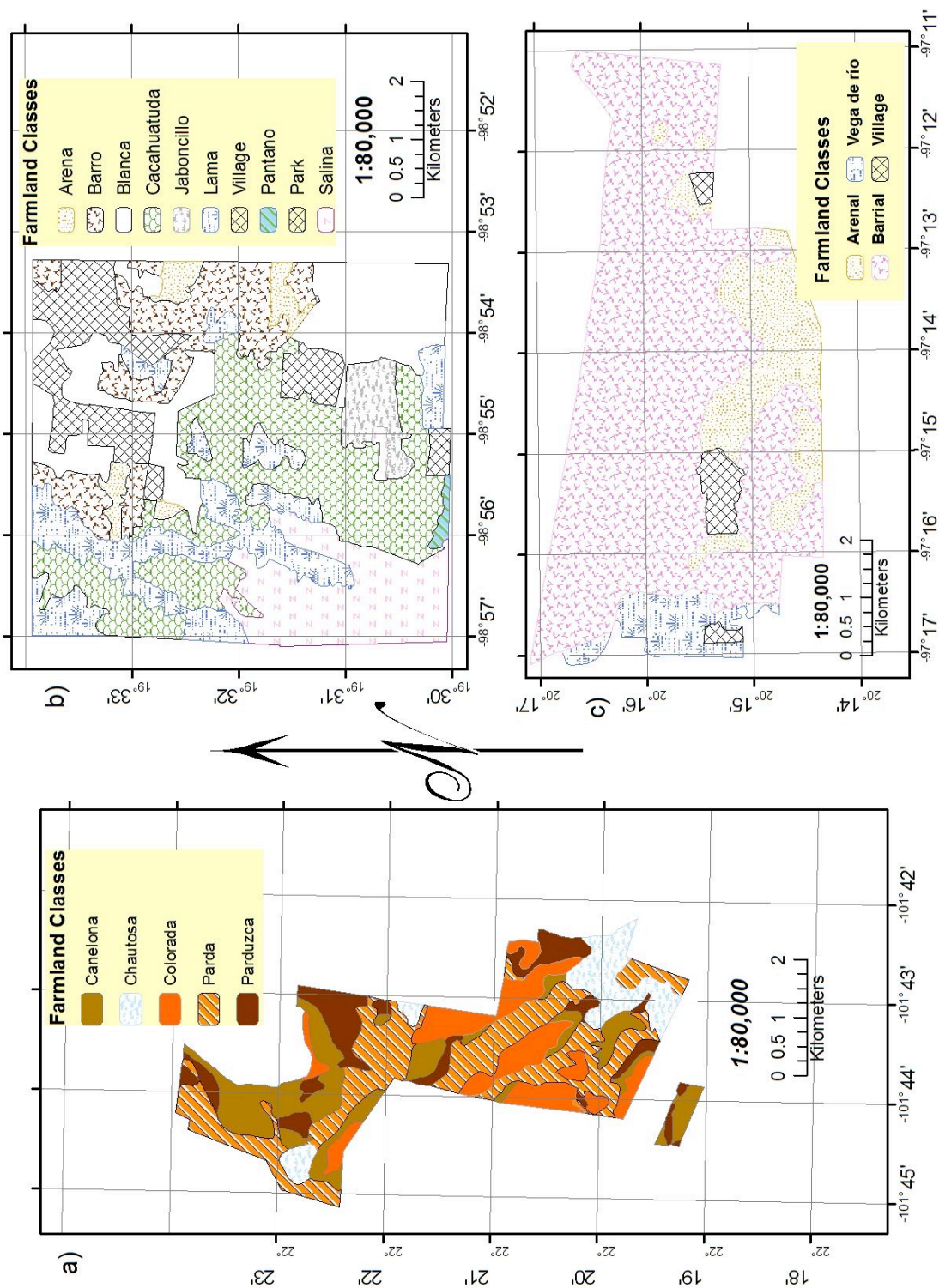


Figure 2. Farmland Classes maps a) Villa Hidalgo, Zacatecas; b) Texcoco, Mexico; c) Papantla Veracruz.

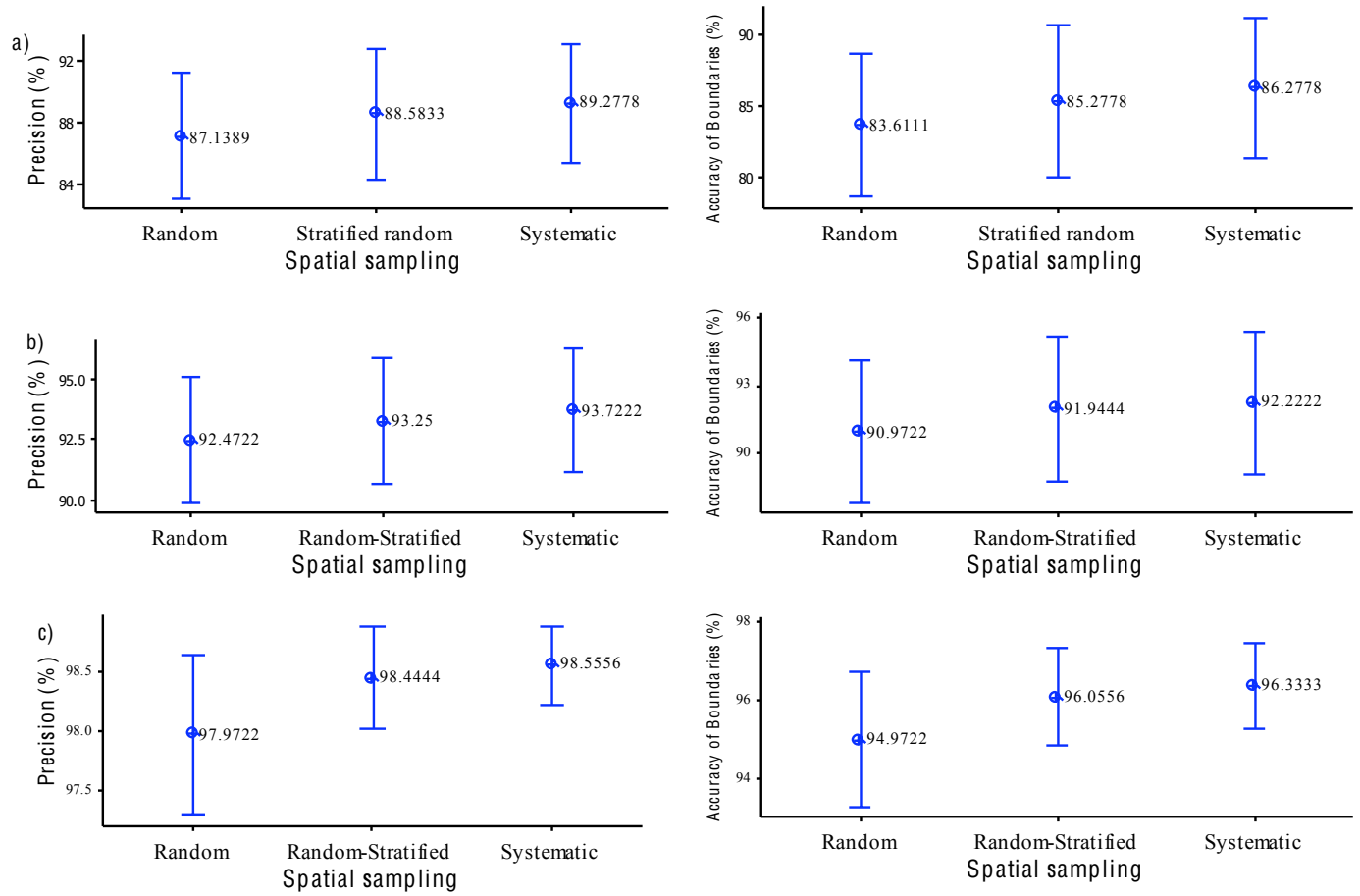


Figure 3. Precision and accuracy based on the sampling design. a) Villa Hidalgo, Zacatecas; b) Texcoco, Mexico; c) Papantla, Veracruz.

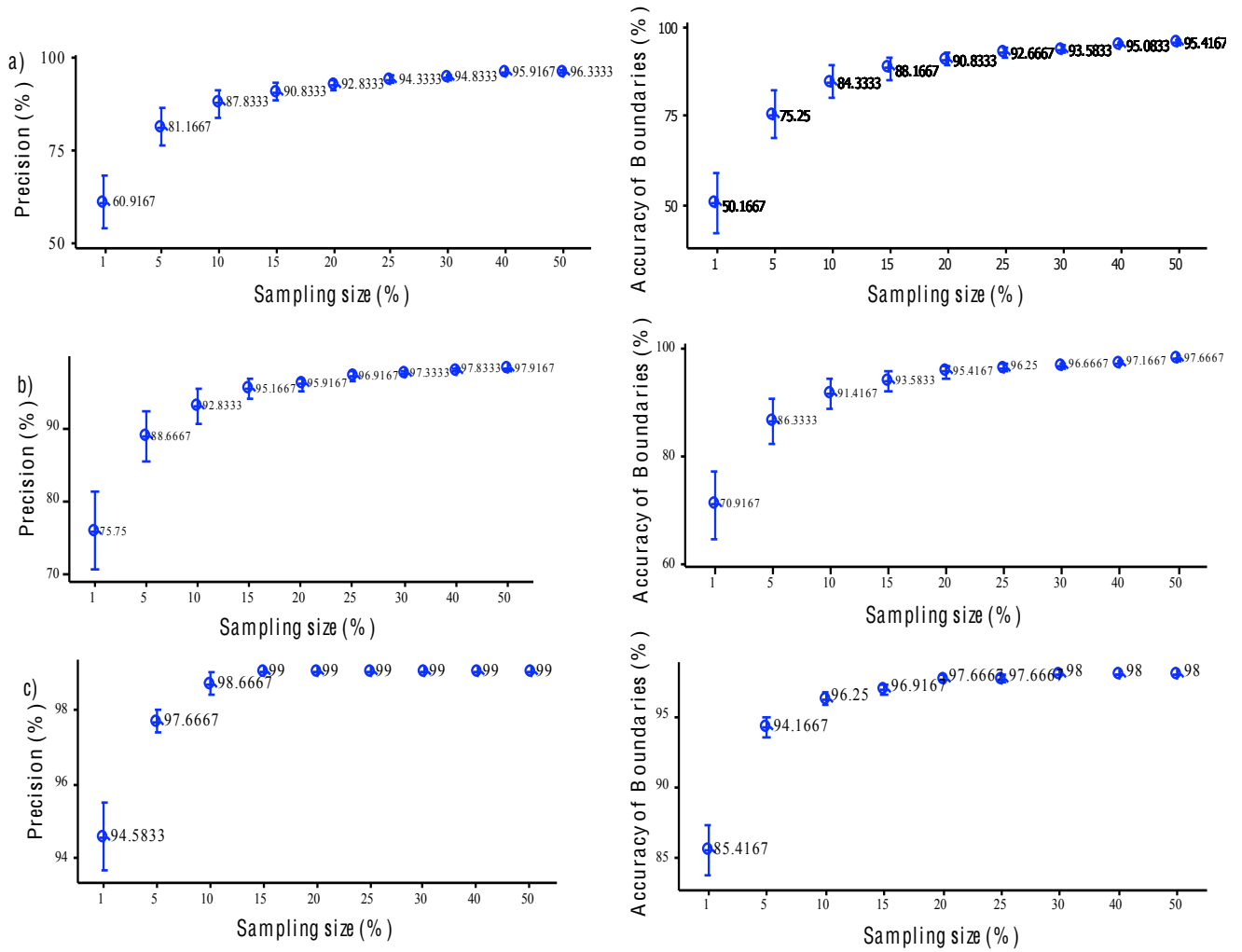


Figure 4. Precision and accuracy based on sample size. a) Villa Hidalgo, Zacatecas; b) Texcoco, Mexico; c) Papantla, Veracruz.

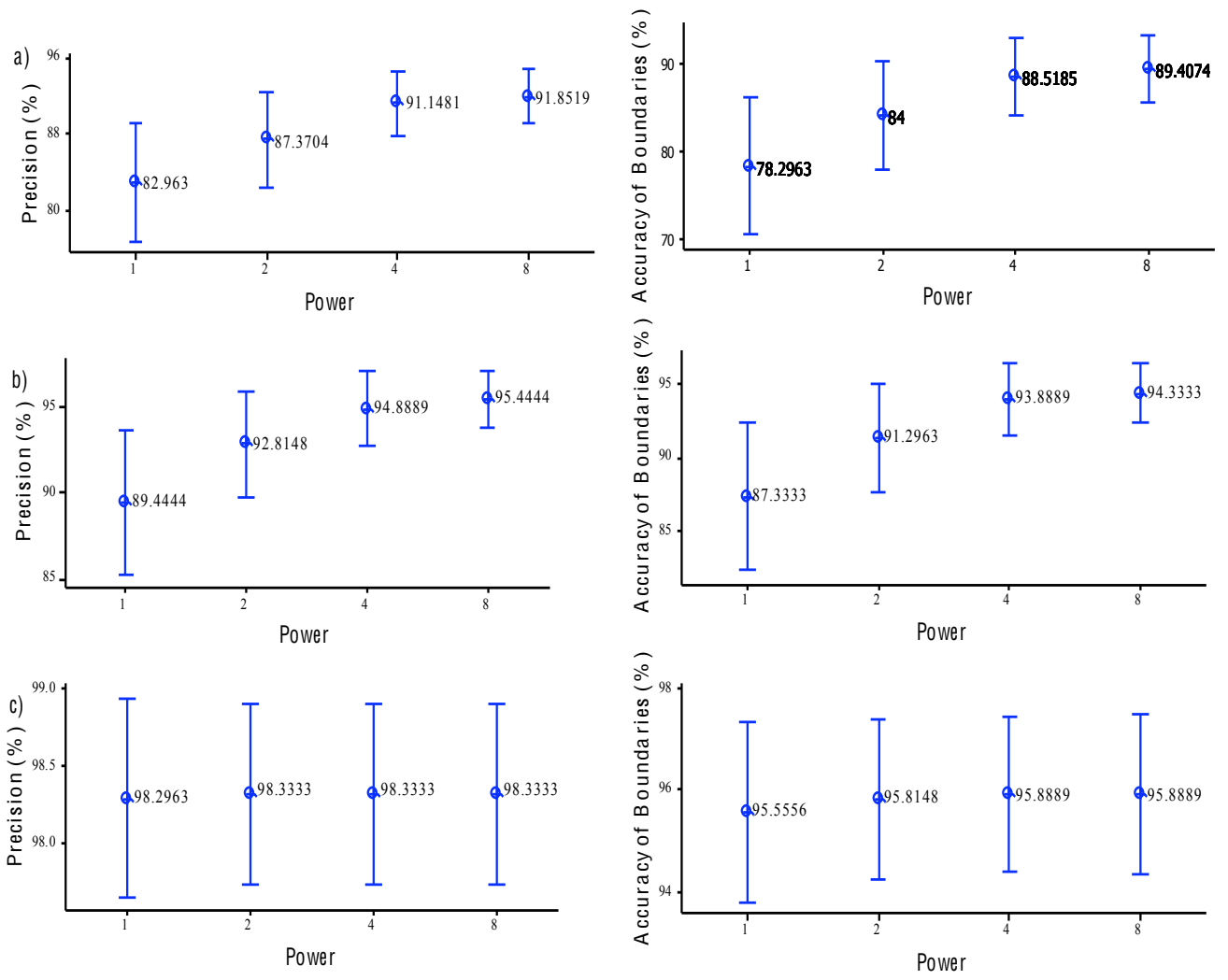


Figure 5. Precision and accuracy based on the Power. a) Villa Hidalgo, Zacatecas; b) Texcoco, Mexico; c) Papantla, Veracruz.

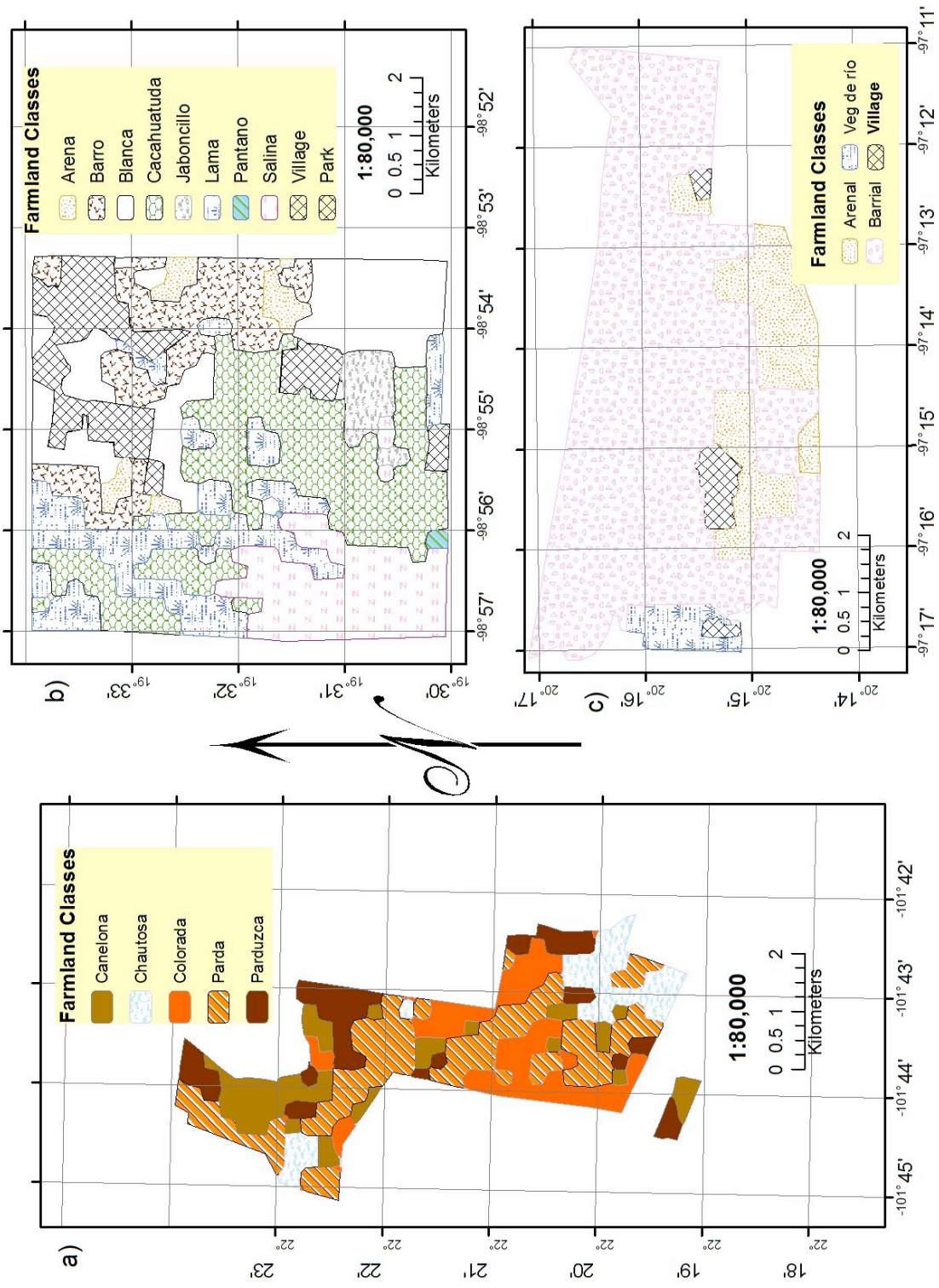


Figure 6. Farmland class maps generated with IDW model (with 25% number of points of sampled plots). a) Villa Hidalgo, Zacatecas; b) Texcoco, Mexico; c) Papantla, Veracruz.

DISCUSIÓN Y CONCLUSIÓN GENERAL

1. Discusión

Zona árida

El mapa de mayor confiabilidad para la zona árida fue el que se generó con el MDE, curvatura del perfil y pendiente (Figura 1c) con 39% de precisión y 10 de exactitud. En contraste, en donde sólo se utilizó como predictor a la curvatura se produjo el mapa de más baja confiabilidad (Figura 1g). Empleando sólo la curvatura del perfil en el mapeo digital de clases de tierra campesinas, seis de ocho clases se traslapan con lo cual se genera confusión para cartografiar estas clases (de 1 a 6). Cuando se utilizan los valores de MDE, curvatura del perfil y pendiente solo cuatro de las ocho clases tienen traslape (1, 2, 3 y 5). Estas representan más de 52% de la superficie del área de estudio, lo cual influye en la precisión y exactitud del mapa porque esta superficie no es cartografiada en forma adecuada.

La combinación de la reflectancia y el MDE como predictores para el mapeo digital de clases de tierra campesinas generó el mapa de mayor precisión y exactitud con 42% y 6% respectivamente, sólo cuando se empleo máxima verosimilitud como clasificador. Los mapas con menor confiabilidad en cuanto a precisión y exactitud, menos de 29% y 2% respectivamente, resultaron cuando se utilizó como predictores las reflectancias de las siete bandas; entre ellos, el de más baja calidad fue el mapa que se creó con el clasificador del paralelepípedo. Las clases 1, 2, 4, 5, 6 y 8, de un total de ocho, no pueden ser discriminadas porque existe traslape en sus valores de reflectancia (Figura 5h), por lo que presentan problemas para ser cartografiadas correctamente. El empleo del MDE disminuye la confusión entre clases (Figura 1i) y esto se refleja en el aumento de la confiabilidad de los mapas en más del 10%. Sólo cuatro de las ocho clases de tierra campesinas se traslapan, 1, 2, 3 y 5 (Arena, Arena gravosa, Calichuda y Delgada), con el análisis de atributos topográficos. Martínez *et al.* (2003) tuvieron el mismo problema al realizar la fotointerpretación para identificar estas clases de tierras porque comparten los mismos rasgos fotográficos y se confunden. El factor que las diferencia es la presencia o ausencia de infraestructuras de riego.

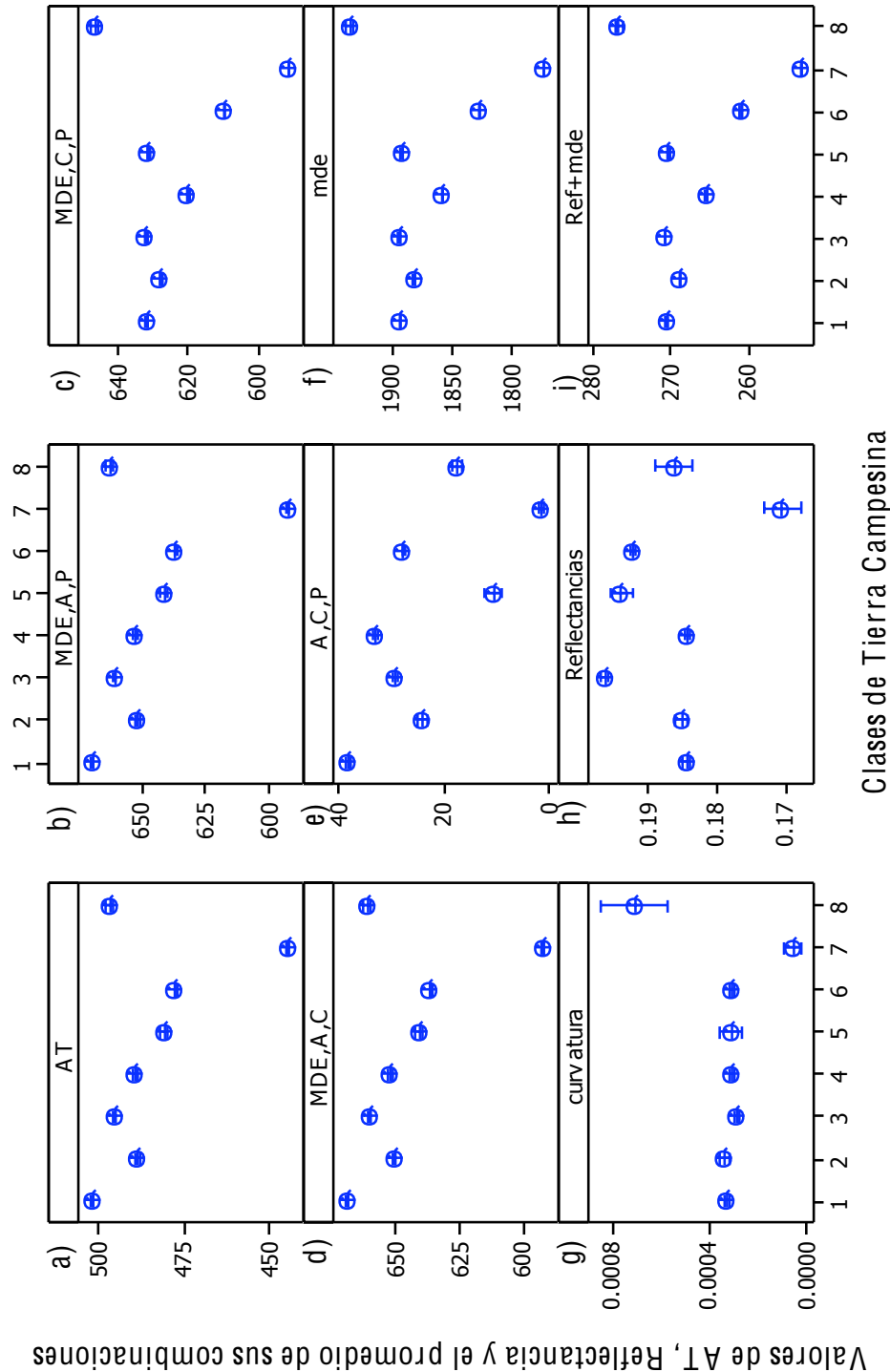


Figura 1. Gráficas de intervalos de confianza ($\alpha=0.05$) de los valores de los atributos topográficos (AT) y reflectancia (Ref), y el promedio de sus combinaciones de las ocho clases de tierra campesinas de la zona árida. MDE=mde= modelo digital de elevación; A = aspecto; C=Curvatura del perfil; P=Pendiente.

Zona templada

El mapa de mayor precisión (42%) y exactitud (14%) que se obtuvo en la zona templada empleando árboles de decisiones y atributos topográficos fue cuando sólo se utilizó el MDE. Caso contrario fue el mapa generado sólo con la curvatura del perfil (menos de 24% de confiabilidad).

La Figura 2 muestra que con MDE, curvatura y pendiente (Figura 2c) y el MDE (Figura 2f) como predictores sólo dos de las seis clases de tierra campesinas se traslapan (clase 4 y 6). Estas clases representan el 40% de la superficie total, con lo cual se puede predecir que los mapas sólo pueden alcanzar hasta el 60% de confiabilidad. Con la curvatura del perfil las seis clases se sobreponen (Figura 2g). Este mismo comportamiento se presentan cuando se emplean todos los atributos topográficos (Figura 6a), y las combinaciones entre aspecto, curvatura del perfil y pendiente (Figura 2e).

El empleo del clasificador de máxima verosimilitud con reflectancia y el MDE como predictores también generó el mapa de mayor confiabilidad en esta zona, con 62% de precisión y 16% de exactitud. El mapa de menor confiabilidad se obtuvo con la red neuronal artificial y las reflectancias de las siete bandas, con 18% de precisión y 0% de exactitud. La Figura 2i muestra que sólo en dos pares de clases de tierra campesinas existe sobreposición, 2 y 5 (Barro y Lama) y 4 y 6 (Cacahuatuda y Salina) respectivamente, cuando se utilizan los valores de reflectancias y la combinación de estas con el MDE. Por lo tanto, en este caso el clasificador es el que determina el porcentaje de confiabilidad de los mapas. Licona *et al.* (1993) empleando la fotointerpretación como técnica de mapeo, no tuvieron problemas para identificar estas clases de tierra. El principal rasgo que emplearon para discriminarlas fue la forma.

Zona tropical

La mayoría de atributos topográficos y sus combinaciones empleados para generación de mapas en la zona tropical tuvieron en general la misma precisión y exactitud. Lo anterior se debe a que los valores de las variables que se utilizaron para discriminar a las clases de tierra campesinas las describen en forma adecuada y no se traslapan (Figura 3a, 3b, 3c, 3d, 3e, 3f, 3g). Mismo comportamiento se presenta con el uso de reflectancias y el MDE (Figura 3f, 3h, 3i). Como en la zona templada el clasificador es el principal factor que interviene en la calidad de los mapas generados.

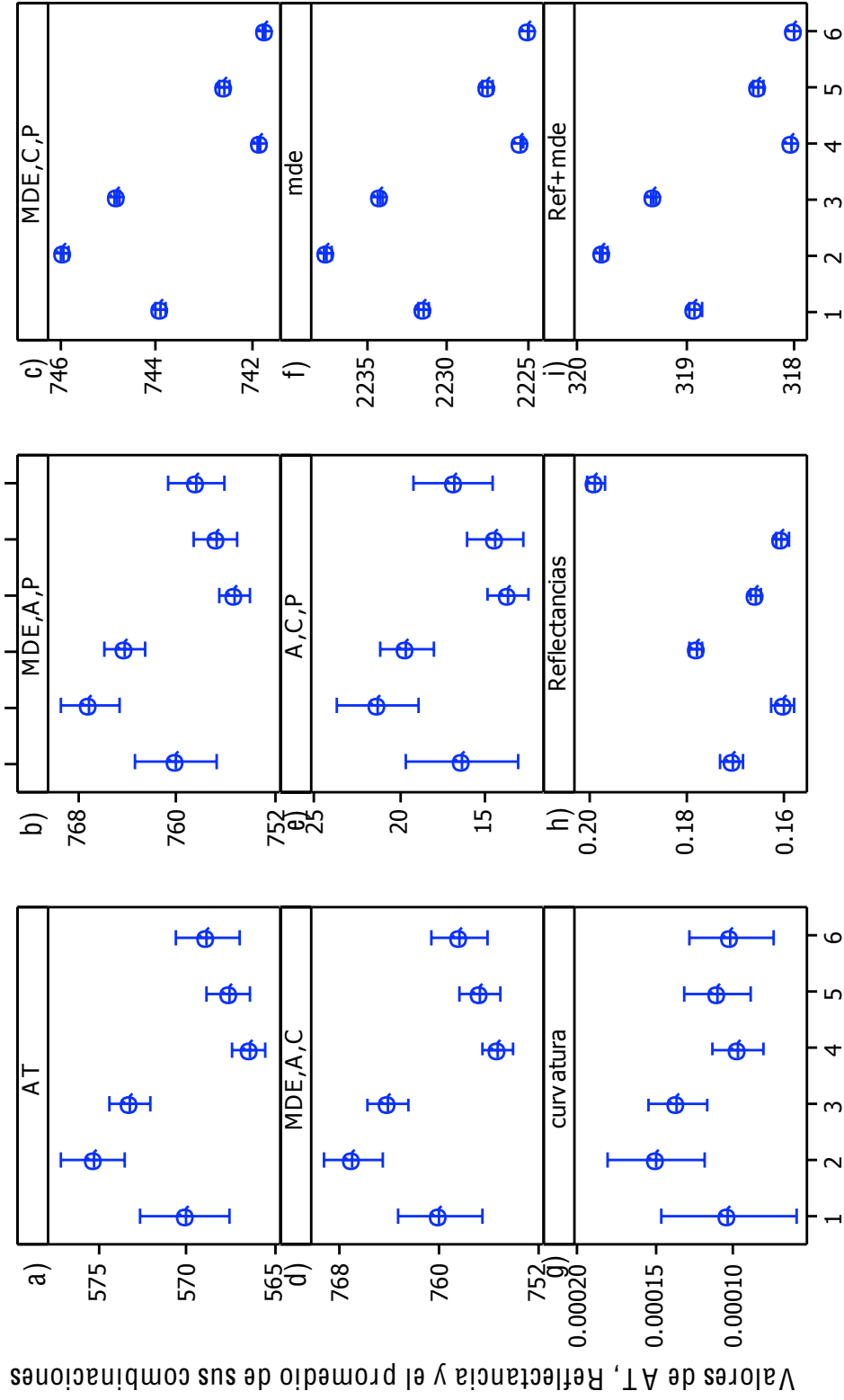


Figura 2. Gráficas de intervalos de confianza ($\alpha=0.05$) de los valores de los atributos topográficos (AT) y reflectancia (Ref) y el promedio de sus combinaciones de las seis clases de tierra campesinas de la zona Templada. MDE=mde= modelo digital de elevación; A = aspecto; C=Curvatura del perfil; P=Pendiente.

Con lo que respecta al uso del interpolador inverso de la distancia (IDW) en combinación con el tamaño de la muestra para el entrenamiento, el diseño espacial de la misma y el valor del poder del interpolador, se encontró que este interpolador demostró ser más eficiente que los dos algoritmos de inteligencia artificial y máxima verosimilitud porque genera mapas de clases de tierra campesinas de más de 60% de confiabilidad en las tres zonas de estudio, además, de emplear menos puntos de entrenamiento. Por ejemplo, con árboles de decisión, MDE, curvatura del perfil y pendiente se obtuvo un mapa en la zona árida de 39% y 6% de precisión y exactitud respectivamente. Con máxima verosimilitud, reflectancia y MDE el mapa tuvo 42% de precisión y 6% de exactitud. En ambos casos el tamaño de la muestra para entrenar a los algoritmos fue de 78% con respecto al número de parcelas que serían muestreadas. En contraste con lo anterior, el mapa que se generó con IDW tiene más de 70% de confiabilidad, usando sólo el 25% como tamaño de muestra. El mismo caso se presentó en la zona templada, ya que mientras que con máxima verosimilitud, reflectancia y MDE se generó un mapa de 62% de precisión y 16% de exactitud empleando más del 116% de tamaño de muestra, con IDW y un tamaño de muestra del 10%, el mapa generado tuvo 80% y 72% de precisión y exactitud respectivamente. Así mismo, mientras que los mapas obtenidos en la zona tropical con RNA y máxima verosimilitud tuvieron 84% de precisión y 30% exactitud con un tamaño de muestra de 146%, los que se generaron con IDW y un tamaño de muestra del 10%, el mapa generado tuvo 80% y 72% de precisión y exactitud respectivamente. Así mismo, mientras que los mapas obtenidos en la zona tropical con RNA y máxima verosimilitud tuvieron 84% de precisión y 30% exactitud con un tamaño de muestra de 146%, los que se generaron con IDW y un tamaño de muestra sólo del 10%, tuvieron una precisión similar (85%) pero con mayor exactitud (68%).

2. Conclusiones

La aplicación de técnicas que comúnmente se utilizan en la cartografía digital de suelos, fueron empleadas para la generación de mapas de clases de tierra campesinas y los resultados fueron satisfactorios, en las tres zonas de estudio, en la medida que se logró incrementar la confiabilidad de los mapas que se habían generado en estudios previos. Sin embargo, la calidad de los mapas siguió siendo baja en las zonas árida y templada, ya que la precisión fue inferior a 62% y la exactitud a 16%. Se demostró que el modelo digital de elevación es el

único atributo topográfico necesario en la cartografía digital de clases de tierra campesinas, pero la combinación de este atributo con los valores de reflectancia de seis bandas del sensor ETM+ del satélite Landsat (espectro visible e infrarrojo) aumentó la confiabilidad de los mapas. Las redes neuronales artificiales y el árbol de decisiones, conocidos como de inteligencia artificial, no fueron los mejores clasificadores a pesar de contar con la estructura más compleja y que en la literatura, especialmente en revistas de alto impacto científico, se reportan como los algoritmos que generan mapas del 80% al 96% de confiabilidad. En este contexto, el mejor clasificador fue máxima verosimilitud, entre redes neuronales artificiales, árboles de decisiones, paralepípedo y mínima distancia (clasificadores que emplean variables auxiliares o predictores) porque con estos se crearon los mapas más confiables en las tres zonas de estudio.

A diferencia de lo anterior, al utilizar el interpolador inverso de la distancia se generaron mapas de mayor calidad en comparación con los mejores mapas, de cada una de las zonas, generados con el modelo digital de elevación, reflectancia y máxima verosimilitud, a pesar de que IDW es un modelo que sólo se había utilizado, hasta el momento, para interpolar variables continuas en el mapeo digital de suelos (McBratney *et al.* 2003). Otras ventajas que se tienen utilizando IDW son: que no necesita los predictores o variables de entrada para realizar la modelación, con lo cual se reduce el tiempo de configuración y el costo, especialmente al no emplear imágenes de satélite; el interpolador no es complejo y se encuentra en la mayoría de los programas de sistemas de información geográfica; requiere menor cantidad de sitios de entrenamiento.

Relacionado con lo anterior, también se concluye que el mejor diseño espacial de muestreo fue el sistemático y el tamaño de muestra de 15% a 20% con respecto al número total de parcelas. Cabe mencionar, que este punto es muy importante porque, para que el clasificador sea entrenado correctamente, es necesario que ésta información sea confiable y por lo tanto es necesario que la fuente sea de campo.

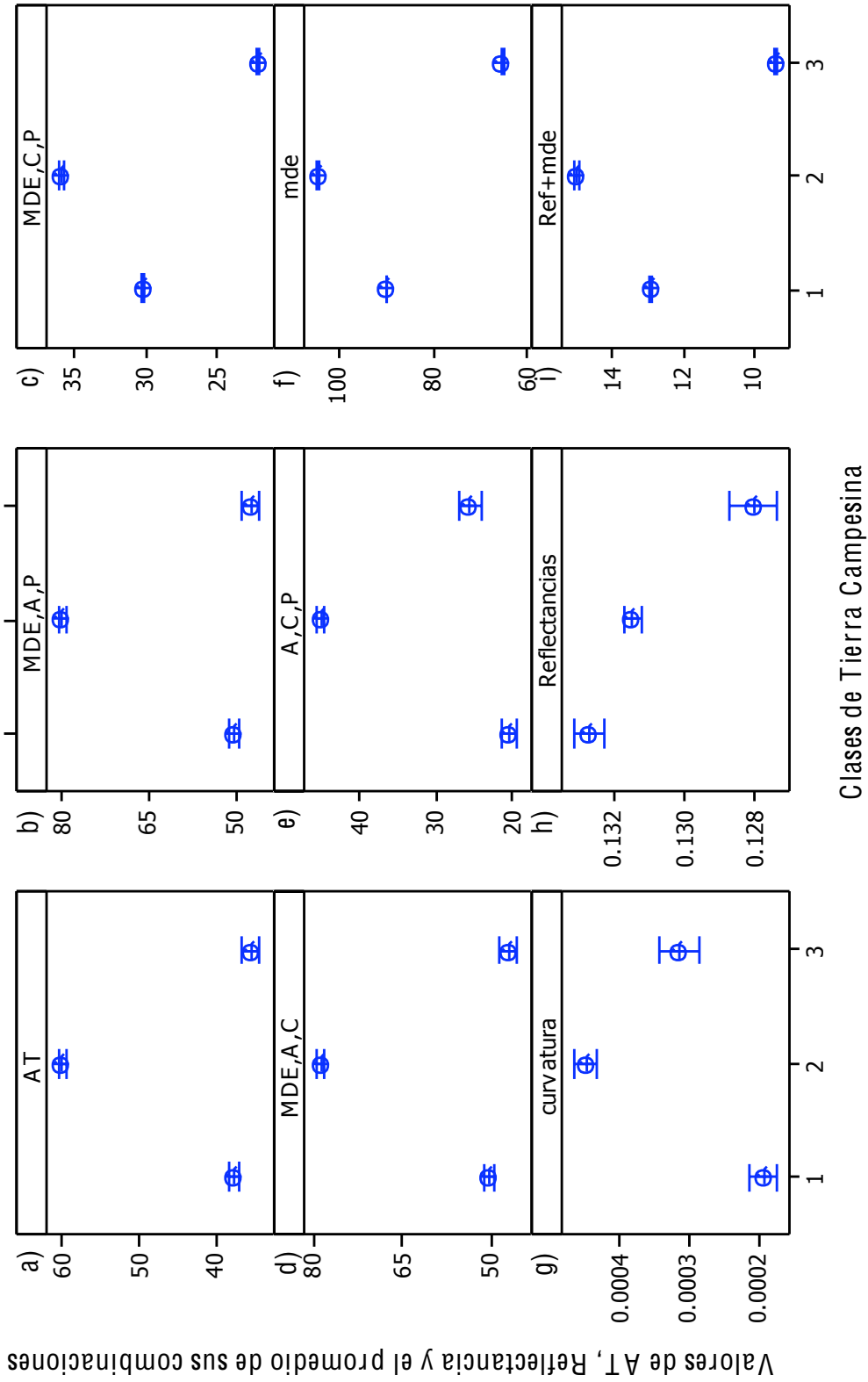


Figura 3. Gráficas de intervalos de confianza ($\alpha=0.05$) de los valores de los atributos topográficos (AT) y reflectancia (Ref), y el promedio de sus combinaciones de las tres clases de tierra campesinas de la zona tropical. MDE=mde= modelo digital de elevación; A = aspecto; C=Curvatura del perfil; P=Pendiente.

3. Recomendaciones (metodología propuesta)

A continuación se describe la metodología que se propone a partir de esta investigación, la cual es sencilla y fácil de desarrollar como se muestra a continuación. Además, se presenta el código de programación en R (2009: <http://www.R-project.org>.) software que se encuentra disponible en línea. Es necesario destacar que, para que el mapeo digital de clases de tierra campesinas sea eficiente, se deben de tomar en cuenta aspectos del tamaño de muestra y su diseño espacial, porque con base en estos factores se determina en parte la confiabilidad de los mapas. Por otra parte, el algoritmo que se propone es el IDW, ya que con él se obtuvieron los mapas de mayor confiabilidad en las tres áreas de estudio.

1. Se determina el tamaño promedio de parcela. Se puede hacer dividiendo el número total de parcelas entre la superficie total de la zona de estudio. Si no se cuenta con este dato, se puede realizar una exploración sobre imágenes de alta resolución espacial (como ortofotos) y medir la superficie de muestra de parcelas y con ello obtener un tamaño promedio.

```
#Cargar librerías que se emplearan en el análisis  
library(sp)  
library(spatstat)  
library(geoR)  
library(gstat)  
library(rgdal)  
#Se carga el contorno del area de estudio en formato ASCII  
CTC=readGDAL("D:/Zona_Templada/Simulacion_Temp/cttex.asc")
```

2. El tamaño promedio de parcelas se divide entre la superficie total del área de estudio con lo cual se obtiene el tamaño de muestra que representa el 100% de parcelas que serian muestreadas. De este tamaño de muestra se toma de 15 a 20% del total de puntos. Los puntos de la muestras se arreglan en forma sistemática en el espacio de la zona de estudio.

```
#Dividir el área entre tamaño de parcela 5734/2=2866 (100% parcelas muestreada)  
#516 puntos representan el 18% del total de puntos  
#se obtiene raíz de 516 para determinar el número de puntos por columna y fila  
#Se genera la malla de muestreo  
sistemico=rsyst(sqrt(516),win=as.owin(c(CTC@bbox[1],  
CTC@bbox[3],CTC@bbox[2],CTC@bbox[4])))
```

3. La capa de puntos se exporta y se sobrepone en una ortofoto (mapa base). Se va a campo a tomar la información de cada punto, es decir, a la clase de tierra que pertenece y se les etiqueta en la base de datos de la capa de puntos de muestreo.

```
#Se exporta la capa de puntos en formato Shapefile el cual después puede ser importado desde ILWIS
```

```

Xr=sistematico$x
Yr=sistematico$y
sisR= as.data.frame(cbind(Xr, Yr))
library(RPyGeo)
write.point.shapefile(sisR, "D:/Zona_Templada/Simulacion_Temp/sist_R",
x.field = "Xr", y.field = "Yr", id.field = NULL)

```

4. Se cargan los puntos de muestreo etiquetados a R y se emplea el interpolador inverso de la distancia con un poder 8 para realizar el mapeo digital de clases de tierra campesinas.

```

#Se carga la capa de puntos etiquetados con las CTC
#Cual debe de contener coordenadas (x,y) y el campo de CTC
p.CTC=read.csv("D:/Zona_Templada/Simulacion_Temp/pun_etique.csv")
#Convertir p.CTC a SpatialPointDateFrame
CTC.sp=p.CTC
coordinates(CTC.sp)=~x + y
###Delimitar area de estudio#####
xmin=CTC@bbox[1]
xmax=CTC@bbox[3]
ymin=CTC@bbox[2]
ymax=CTC@bbox[4]
#Resolucion
reso=30
#Hacer Grid donde se almacenan datos interpolados
grid.xy <- expand.grid(x = seq(xmin,xmax,by=reso),y= seq(ymax,ymin,by=-reso))
coordinates(grid.xy) <- ~x + y
gridded(grid.xy)=T
#Interpolación
IDW=krige(CTC~1, CTC.sp, grid.xy, set=list(idp=8))
MIDW=IDW["var1.pred"]
spplot(MIDW)
#Exportar a ILWIS para generar una salida de impresion
writeGDAL(MIDW, "D:/Zona_Templada/Simulacion_Temp/IDW_18.mpr", "ILWIS")

```

5. Este último punto es opcional ya que se realiza si desea conocer la calidad del mapa generado. Se crea otra capa de puntos, se va a campo y se identifica cada punto. La información de campo y del mapa se compara con una matriz de confusión para obtener la confiabilidad del mapa generado con IDW.

El código de programación funciona para cualquier de las tres regiones de estudio con características similares, sin embargo, se tienen que hacer algunas modificaciones para cada zona en particular.

- ✓ Se cambia el nombre del archivo ASCII al correspondiente.
- ✓ El tamaño de muestra con base al porcentaje del número total de parcelas.

4. Referencias

- LICONA, V.A., ORTIZ, S.C.A. y PÁJARO, H.D., 1993, El uso de la fotointerpretación en la cartografía de clases de tierra campesina. *Revista Geografía Agrícola*, 18, pp. 85-93.
- MACBRATNEY, B.A., MENDONÇA-SANTOS, L.M. and MINASNY, B., 2003, On digital soil mapping. *Geoderma*, 117, pp. 3-52.
- MARTÍNEZ, M.J.F., ORTIZ, S.C.A., PÁJARO, H.D., OJEDA, T. E. y VARGAS, P.E., 2003, Clave de fotointerpretación para la cartografía de tierras campesinas a escala regional. *Terra*, 21, pp. 301-309.
- R DEVELOPMENT CORE TEAM, 2009, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.