

DIGITAL MAPPING OF FARMLAND CLASSES IN THREE LANDSCAPE IN MEXICO

G. Cruz-Cárdenas^{1,5*}, C.A. Ortiz-Solorio¹, E. Ojeda-Trejo¹, J.F. Martínez-Montoya², E.D. Sotelo-Ruiz³ and A.L. Licona-Vargas⁴

¹Colegio de Postgraduados, Campus Montecillo, Carretera México-Texcoco km 36.5, 56230, Texcoco, km 36.5, México, México. ²Colegio de Postgraduados, Campus San Luis Potosí, Iturbide 73, 78620, Salinas de Hidalgo, Salinas, SLP, México. ³Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, km 18.5 Carretera Los Reyes-Lechería, Col. Chapingo, 56230, Texcoco, México, México. ⁴Universidad Autónoma Chapingo, CRUO, km 3 Carretera Huatusco-Jalapa, 94100, Huatusco, Veracruz, México. ⁵Postdoctoral, Instituto de Biología, Departamento de Botánica, Universidad Nacional Autónoma de México, Tercer Circuito s/n, Ciudad Universitaria, 04510, Delegación Coyoacán, México D.F., México.

*Corresponding author: gcruz@colpos.mx

ABSTRACT

The cartography of farmland classes allows generating land maps, using a methodology based on local knowledge, rapidly and at low cost, and with a greater number of cartographic units than conventional soil surveys maps. 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.

Keywords: Map accuracy, IDW interpolator, soil sampling strategies.

INTRODUCTION

A farmland class (FLC) 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, Martínez (1993) and Ortiz (1999) mention that digital mapping of FLC cannot be done since the classes cannot be identified satisfactorily. 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 is 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 (technical maps) 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 (it is an exponent which determines the weight assigned to each of the observations) 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 $30p$ is taken, meaning 30 pixels times the number of bands or layers (p) 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, nevertheless Kravchenko and Bullock (1999) affirm 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.

MATERIALS AND METHODS

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%.

Cartography of the farmland classes

The FLC maps for each zone were generated through the methodology of Ortiz *et al.* (1990), which consists of 1) selecting a base map on which boundaries are drawn, the recommendation is to use with aerial photographs; 2) the farmers or

informants are selected of the study zone. It is important to mention that the informants can be divided into two groups; one for the cartography of FLC, and other for the characterization of the FLC, their problems, management techniques and even alternatives for improvement. Experience showed that the first group might be composed of two or three persons who are familiar with the entire area; 3) soil surveys around the area, accompanied by informants, with the corresponding aerial photograph in hand. The soil surveys walk around the area accompanied by informants, the following questions are made: Where does land class change? and how is it different from neighboring ones?

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.

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):

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

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

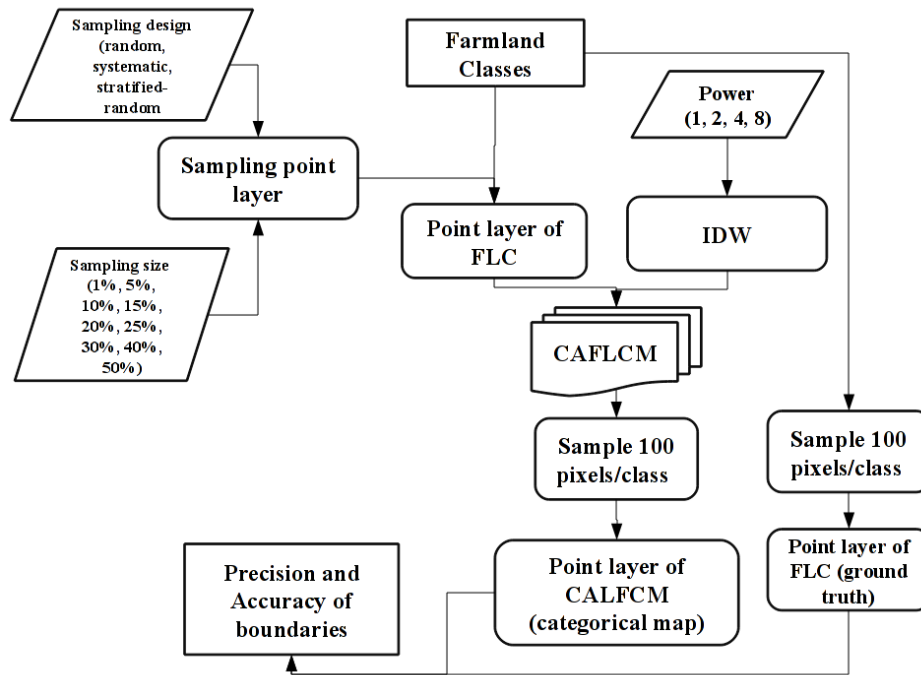


Figure 1. General scheme of methodology. Farmland Classes (FLC); Computer Assisted Farmland Classes Map (CAFLCM).

Table 1. Sampling density per plot size.

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

is the average walking speed of a person (km hr⁻¹), in this case 5, PS is the average size of the plot (km²).

$$ST = (TP + T)(SS - 1) \quad (4)$$

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.

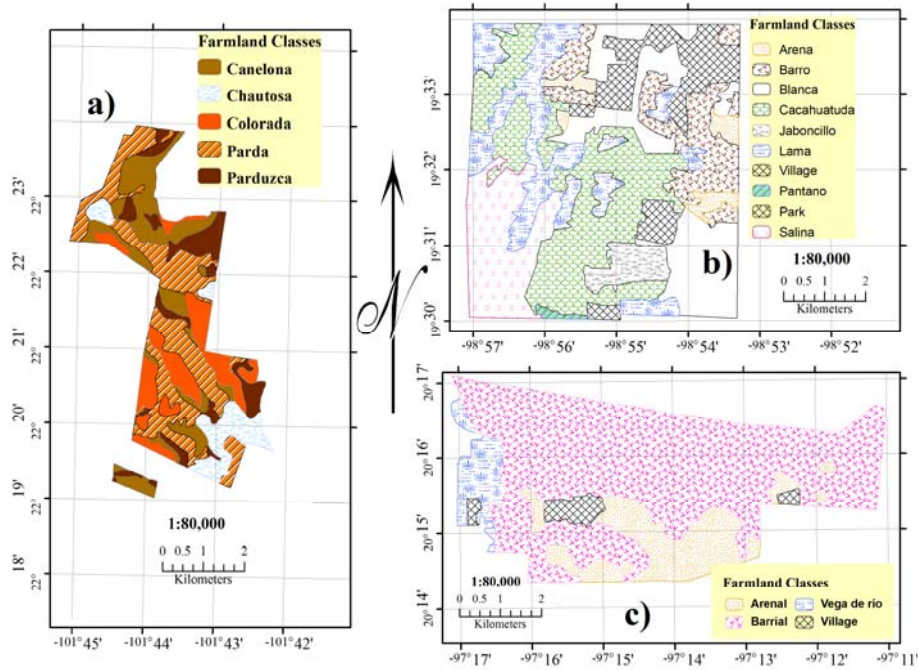


Figure 2. Farmland Classes maps a) Villa Hidalgo, Zacatecas; b) Texcoco, Mexico; c) Papantla Veracruz.

obtained are similar to those of a 50% sample size (11325). In Texcoco, 15% of the total pixels (7716) were 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

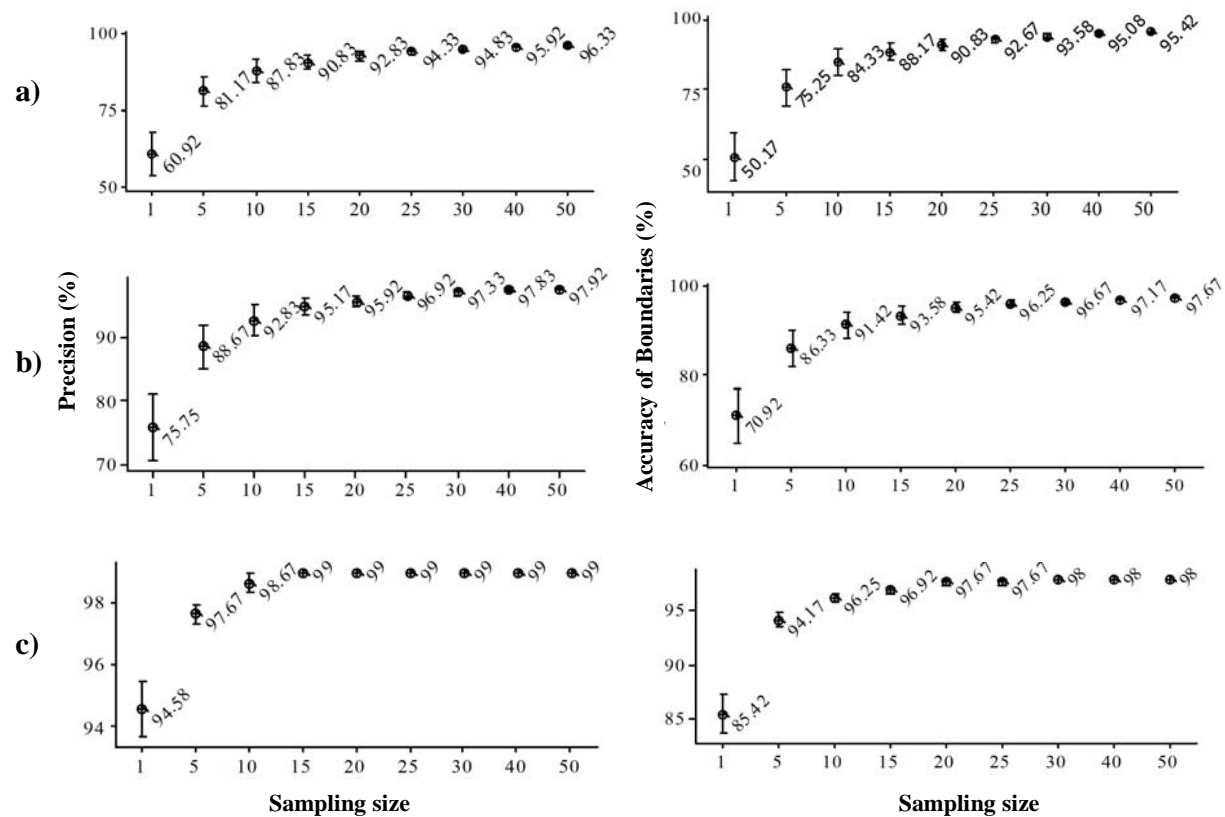


Figure 4. Precision and accuracy based on sample size. a) Villa Hidalgo, Zacatecas; b) Texcoco, Mexico; c) Papantla, Veracruz.

Power 4, followed by Power 1, 3, and 2 because the grid size for soil sampling was varied and this influenced the Power values.

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 3.

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 sample size varies because the recommendation of Foody and Mathur (2006) did not take into account the plot size. Small plot requires more than a larger sample size and vice versa. In addition, Foody and Mathur (2006) used support vector machine to classify, the sampling scheme was random and used the remote sensing data as input variables.

Table 3. 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

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%

plan was the systematic scheme and Power 8 for all three zones because maps were obtained the most precision and accuracy.

The IDW does not use predictors or input variables for modeling compared algorithms commonly used in digital mapping soils like support vector machine, artificial neural network, decision tree, thus reducing setup time and cost, especially because the IDW does not use remote sensing data, the interpolation is not complex and the IDW is in most programs of geographic information systems.

REFERENCES

- Barrera-Bassols, N., Zinck, A.J., Van Ranst, E. 2006.** Local soil classification and comparison of indigenous and technical soil maps in a Mesoamerican community using spatial analysis. *Geoderma* 135, 140-162.
- Carré, F., MacBratney, B.A., Minasny, B. 2007.** Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. *Geoderma* 141, 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. *Photogramm. Eng. Rem. S.* 54, 593-600.
- Congalton, G.R. 1991.** A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* 37, 35-46.
- Cruz, C. G., Ortiz, C.A., Gutiérrez, C. Ma. del C., Villegas, M.A. 2008.** Las clases de tierras citrícolas del ejido Pueblillo, Papantla, Veracruz. *Terra Latinoam.* 26, 11-19.
- Ericksen, P.J., Ardon, M. 2003.** Similarities and differences between farmer and scientist views on soil quality issues in central Honduras. *Geoderma* 111, 233-245.
- Foody, M.G., 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 Sens. Environ.* 103, 179-189.
- Foody, M.G., Mathur, A., Sanchez-Hernandez, C., Boyd, S.D. 2006.** Training set size requirements for the classification of a specific class. *Remote Sens. Environ.* 104, 1-14.
- García, E. 1988.** Modificaciones al sistema de clasificación climática de Köppen. Ind, México, 220 p.
- Gotway, C.A., Ferguson, R.B., Hergert, G.W., Peterson, T. A. 1996.** Comparison of Kriging and Inverse Distance Methods for mapping soils parameters. *Soil Sci. Soc. Am. J.* 60, 1237-1247.
- Grinand, C., Arrouays, D., Laroche, B., Martin, P.M. 2008.** Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma* 143, 180-190.
- Hengl, T., Rossiter, D., Stein, A. 2003.** Soil Sampling strategies for spatial prediction by correlation with auxiliary maps. *Aust. J. Soil Res.* 41, 1403-1422.
- Instituto Nacional de Estadística, Geografía e Informática (INEGI). 1988.** Atlas nacional del medio físico. Aguascalientes, México, 224 p.
- Kravchenko, A., Bullock, G.D. 1999.** Comparative Study of Interpolation Methods for Mapping Soil Properties. *Soil Sci. Soc. Am. J.* 91, 393-400.
- Licona-Vargas, A.L., Ortiz-Solorio, C.A., Gutiérrez-Castorena Ma. Del C., Manzo-Ramos, F. 2006.** Clasificación local de tierras y tecnología del policultivo café-plátano para velillo-sombra en comunidades cafetaleras. *Terra Latinoam.* 24, 1-7.
- Lleverino, G.E., Ortiz, C.A., Gutiérrez, C. Ma. del C. 2000.** Calidad de los mapas de suelos en el ejido de Atenco, estado de México. *Terra* 18, 103-113.
- Lloyd, D.C. 2007.** Local Models for Spatial Analysis. CRC Press, Boca Raton, FL, 219 p.
- Martínez, M.J.F. 1993.** El uso de Fotografías Aéreas e Imágenes de Satélite en la Cartografía de Tierras. Master thesis, Colegio de Postgraduados, México, 164 p.
- Moran, J.C., Bui, N.E. 2002.** Spatial data mining for enhanced soil map modeling. *Int. J. Geogr. Inf. Sci.* 16, 533-549.

