OnLine Journal of Earth Sciences 1 (3): 127-135, 2007
ISSN: 1991-7708

Accuracy Assessment of Land Cover Maps in Sub-tropical Countries: A Sampling Design for the Mexican National Forest Inventory Map.Θ

[1]Stéphane Couturier, [2]Jean-François Mas, [1]Álvaro Vega and [1]Valdemar Tapia
[1]Geography Institute, Universidad Nacional Autónoma de México (UNAM),
Ciudad Universitaria, Coyoacán 04510, México DF, México
[2]Geography Institute, Universidad Nacional Autónoma de México (UNAM),
Unidad Académica Morelia, Antigua Carretera a Pátzcuaro nº 8701 Col.
Ex-hacienda de San José de la Huerta, CP 58190, Morelia, México

Abstract: There is no record so far in the literature of a comprehensive method to assess the accuracy of detailed regional scale Land Use/Land Cover (LULC) maps in the sub-tropical belt. The elevated bio-diversity and the presence of highly fragmented classes hamper the use of sampling designs commonly employed in previous assessments of mainly temperate zones. A sampling design for assessing the accuracy of the Mexican National Forest Inventory (NFI) map at community level is presented. A pilot study was conducted on the Cuitzeo Lake watershed region covering 400,000 ha of the 2000 Landsat-derived map. Various sampling designs were tested in order to find a trade-off between operational costs, a good spatial distribution of the sample and the inclusion of all sparsely distributed classes ('rare classes'). A two-stage sampling design where the selection of primary sampling units was done under separate schemes for commonly and sparsely distributed classes, showed best characteristics. Issues regarding the assessment strategy are devised.

Key words: Double sampling, rare class, proportional sampling, landsat, stereoscopic interpretation, aerial photography

INTRODUCTION

The accuracy assessment of maps consists in estimating their reliability and making an account of errors. Assessing the accuracy of Land-Use and Land-Cover (LULC) maps is a common procedure in geo-science disciplines, as a means, for example, of testing automatic classification methods on satellite imagery. For regional scale LULC maps, the abundance and distribution of categories (or classes) over the large extension of the map, confronted with tight budget constraints, add complexity to accuracy assessments. Only relatively recently have comprehensive accuracy assessments, with estimates for each class, been built and applied to regional or continental LULC maps; Examples are accuracy assessments of Landsat-derived 1992 LULC maps conducted in the United States of America (USA) by Laba *et al*. (2002) and Wickham *et al*. (2004) and the programmed accuracy assessment of the Canadian forest inventory of year 2000 (Wulder *et al*., 2006).

The National Forest Inventory (NFI) 2000 map of Mexico, sponsored by SEMARNAP, the Ministry of the Environment in Mexico, was conceived and jointly coordinated by the Institute of Geography of UNAM (the National Autonomous University of Mexico) and INEGI, the National Institute of Statistics, Geography and Informatics in Mexico. The cartography was generated in the Geography Institute at UNAM based on visual interpretation at scale 1:250 000 of Landsat imagery of year 2000 (Mas *et al*., 2002). The aim of the 2000 NFI cartographic project was to update previous LULC maps, especially the 1993 cartography known as the 'serie II' INEGI maps, based on visual interpretation of Landsat images acquired in year 1993. A main trait of the NFI project was to propose a hierarchical classification scheme as a convenient LULC standard in Mexico for future remote-sensing based cartography. Four aggregation levels were considered; the finest taxonomic resolution (denominated community level, including sub-communities) consisted in 75 classes (Palacio-Prieto et al., 2000), a much larger number, for example, than the 21 classes which constitute the legend of fine taxonomic resolution LULC maps (Wickham *et al*., 2004) in the USA. The taxonomic resolution of the NFI at community

Corresponding Author: Stéphane Couturier, Geography Institute, Universidad Nacional Autónoma de México (UNAM), Ciudad Universitaria, Coyoacán 04510, México DF, México

level is comparable to the subclass level of the National Vegetation Classification System (NVCS) of the USA (FGDC, 1997). However, the Mexican NFI map was conceived at much coarser spatial resolution (1:250 000 scale) than the LULC projects in the USA (spatial resolution close to the Landsat pixel). The differences in taxonomic diversity and spatial resolution, mentioned above, are both likely to have an incidence on the way the accuracy of the NFI map can be assessed.

A preliminary assessment had been conducted immediately after the elaboration of the NFI map, based on the interpretation of aerial photographs. However, confidence levels could be reported only for a few classes, abundant in the homogeneous northern part of the country (Mas *et al*., 2002). The research described in this study is motivated by the conception of a full assessment design capable of estimating confidence levels of the total amount of classes and for the entire country. The high number of classes at community level (75) implies the presence of many sparsely distributed classes (rare classes) on the map. This situation complicates a probabilistic sampling scheme (*sensu* Stehman Czaplewski, 1998) which combines both cost-efficiency and statistical validity. For example, the watershed of the Cuitzeo lake, state of Michoacán, Mexico (thereafter the Cuitzeo study area, Fig. 1) is extended on 400,000 hectares and contains 21 classes of the NFI map at community level, including 14 rare classes. This study area illustrates the complexity of the NFI map. The objective of this research, was to find a sampling design where the entire set of rare classes present on typical sub-tropical cartography can be sampled under reasonable cost constraints; for this purpose, we propose novel sampling strategies. In this study, we applied a few strategies on the Cuitzeo study area, some of which had been successfully applied to the LULC maps in the USA, and we compared their performance with our novel sampling strategies.

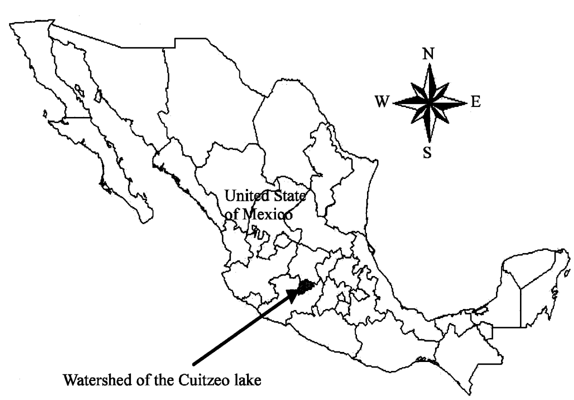

Fig. 1: Location map of the watershed of the Cuitzeo lake

In the methods section, we present the entire accuracy assessment, with an emphasis on simulations of the various sampling designs. The comparison of their performance, in the results section, allowed us to analyze the rationale behind an optimal design and select one design for the evaluation of the map in the pilot area. The derivation of the confusion matrix and accuracy indices is then detailed in the subsection on the synthesis of the evaluation, taking into account the characteristics of the sampling design. Perspectives of the sampling design are finally devised.

MATERIALS AND METHODS

Data from the cuitzeo study area: The Cuitzeo lake, second in size in Mexico, is located on the transversal volcanic axis in central western Mexico. The closed watershed of the Cuitzeo lake (area under investigation) is located to the North of the Michoacán state (Fig. 1), and is covered with temperate sub-humid or tropical dry vegetation. The original vegetation was mainly forests of oak, oak-pine (mixed) and pine towards the South-Southwest and tropical scrubland towards the North-Northeast. Major land uses in year 2000 according to the NFI are annual rain-fed agriculture, irrigated agriculture and induced grassland for pasture (Table 1). Apart from the diversity of LULC classes, a key motivation for selecting this pilot area was the availability of detailed reference data exploitable in the process of independently verifying the 2000 Landsat-based NFI. A complete coverage of 244 black and white aerial photographs on paper at scale 1: 37 000 was acquired in 1999 over the Cuitzeo area, therefore quasi synchronous with the Landsat imagery. The following reference data were also gathered and processed:

- Twelve topographic INEGI maps at 1:50 000 were scanned.
- One colour composite (bands 541) of a Landsat image mosaic of year 2000 was prepared over the watershed, georectified against the INEGI 1:50 000 topographic maps.

Accuracy assessment strategy: The algorithm adopted for the accuracy assessment, represented in Fig. 2, is organized around three steps:

Sampling design: Selection of representative reference sites.

Verification design: Description of reference sites, and labeling of the reference sites by means of interpretation of aerial photograph.

Table 1: Distribution of classes of the National Forest Inventory map in the Cuitzeo lake watershed

Class code	Class name	Formation	Area fraction	Area (km2)
100	Irrigated crop	Cropland	0.1411	564.97
110	Hygrophilous crop		0.0048	19.04
200	Perennial crop		0.0021	8.27
210	Annual crop		0.2356	943.14
300	Forest plantation		0.0071	28.24
410	Fir forest	Temperate Forest	0.0037	14.72
420	Pine forest		0.0041	16.32
421	Pine forest and Sec Veg[a]		0.0036	14.31
510	Oak-pine forest		0.0958	383.34
511	Oak-pine forest and Sec Veg[a]		0.0325	130.29
600	Oak forest		0.0232	92.88
601	Oak forest and Sec Veg[a]		0.0553	221.54
700	Cloud forest	Tropical forest	0.0029	11.73
920	Sub-trop scrubland	Scrubland	0.0194	77.58
921	Sub-trop scrubland and Sec Veg[a]		0.0768	307.25
1000	Mezquital		0.0004	1.51
1200	Chaparral[b]			
1330	Induced grassland	Grassland	0.1594	638.04
1410	Hygrophilous grassland	Hygrophilous Veg	0.0209	83.50
1510	Halophilous vegetation		0.0069	27.78
1600	No apparent vegetation cover[b]	Other Cover		
1700	Human settlement		0.0250	100.02
1800	Water	Water	0.0796	318.75
All			1.0000	4003.23

[a]Secondary Vegetation (secondary vegetation, herbaceous and/or shrub-like vegetation, according to NFI classification scheme). [b]This class was not present on the map but appeared as a reference label

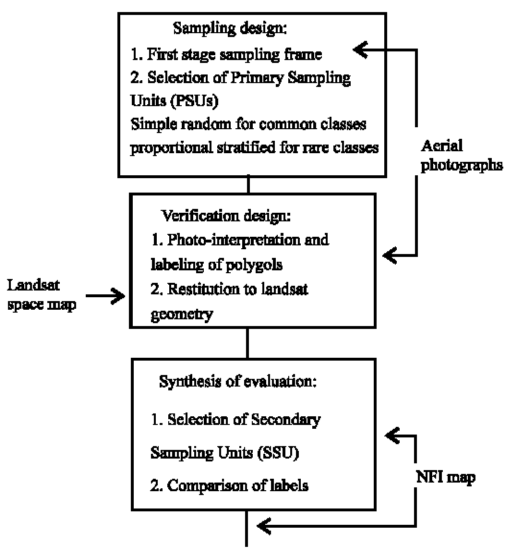

Fig. 2: Algorithm for the accuracy assessment

Synthesis of the evaluation: Comparison between the information contained in the reference site and the information contained in the map, construction of a confusion matrix and computation of assessment indices.

Sampling designs: The sampling design of a regional accuracy assessment aims at estimating accuracy for each class on the map. The selection of reference sites must be such that a statistically representative number of sites covers each mapped class listed in Table 2. Five strategies were contemplated and tested: A first design consisted in a simple random sampling (Simple Sampling, or SS), stratified by class, whereby each class was sampled randomly over the entire territory. Four other designs consisted in two stage, or double, sampling (Double Sampling, or DS). A double sampling design is operated in two steps, first with the selection of Primary Sampling Units (PSUs), and then with the selection of Secondary Sampling Units (SSUs). Conveniently, a PSU corresponds to the effective area of an aerial photograph (photographic frame). The SSUs are dots of a regular grid within the selected PSUs.

Figure 3 shows the complete set of PSUs, which was constructed from the grid of all photographic center points, geo-referenced onto the INEGI topographic maps at 1:50,000. The space map represented in Fig. 3 is derived from a mosaic of four Landsat images of year 2000, which were the basis for the NFI map production (Mas *et al.*, 2002). These Landsat images had also been geo-referenced to the INEGI topographic maps at 1:50,000. A photographic frame pertained to the set of PSUs if its center point was included in the watershed. In this way, the random character of the selection was preserved at the edge of the watershed (Zhu *et al.*, 2000).Two hundred and eighteen photographic frames formed the complete set of PSUs (Fig. 3). The four Double Sampling (DS) designs are distinguished by the selection mode of a PSU, which in turn defines the probability of inclusion of its surface in the sample. For example, a simple random selection is associated with a uniform (constant) inclusion probability among all PSUs

Table 2: Evaluation of five sampling designs on the National Forest Inventory map in the watershed of the Cuitzeo lake

	Total number of selected PSUs[a]	Number of classes with non-representative sample size (confidence interval superior to 15%)	Average number of SSUs[b] per PSU[a] among rare classes	Average number of SSUs[b] per PSU[a] among common classes
SS[c]	167-193	0	14-20	9-13
DS[d] 1	54	3-8	21-25	14-21
DS[d] 2	54	2-4	12-23	19-25
DS[d] 3	54	0-4	20-33	26-36
DS[d] 4	54	0	17-23	16-22

[a]Primary Sampling Unit, [b]Secundary Sampling Unit, [c]Simple (one-stage) Sampling, [d]Double (two-stage) Sampling, Ten simulations of five sampling designs are applied: SS is a simple random, stratified by class sampling. DS1 consists in a simple random selection of Primary Sampling Units (PSU). DS2 consists in a random, stratified by class selection of PSUs, DS3 is a proportional stratified selection, and DS4 treats separately common classes (to which DS1 is applied) and rare classes (to which DS3 is applied)

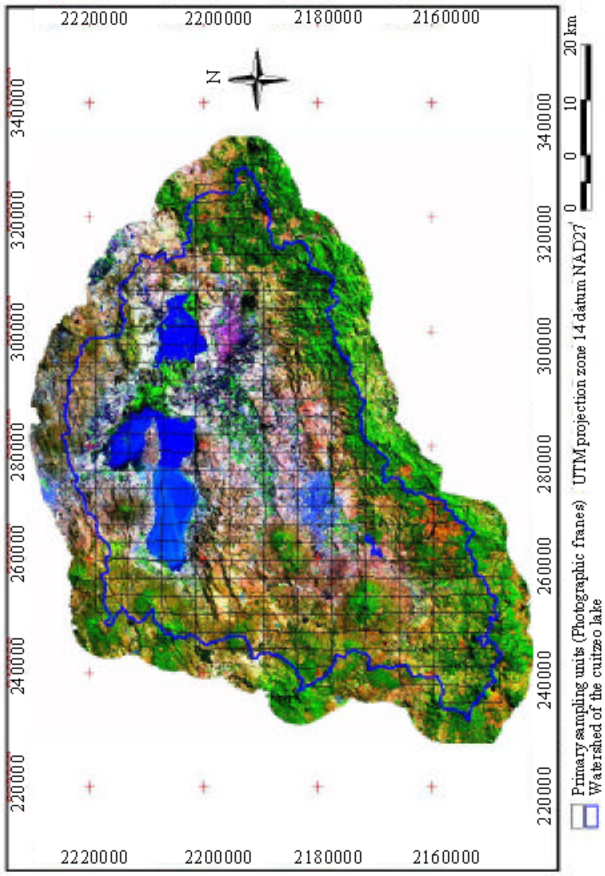

Fig. 3: Space map of the watershed of the Cuitzeo lake

of the complete set. The computation of the accuracy index is governed by the probability of inclusion of each sampled point (selected SSU), as detailed in the section on the synthesis of the evaluation.

The first Double Sampling design (DS1) is defined by the simple random selection of the PSUs, as in Laba *et al.* (2002). DS2 is characterized by the random, stratified by class, selection of PSUs, as in Stehman *et al.* (2003). DS3 is defined by a proportional, stratified by class, selection of PSUs. For the latter design, not applied in previous published research, the probability of inclusion of a PSU is proportional to the abundance of the class in the PSU. The abundance of a class is its surface fraction, obtainable via easy attribute computation in a GIS. Then, in order to ease accuracy calculations, the probability of inclusion of SSUs at the second stage is defined as being inversely proportional to the abundance of the class in the PSU. Proportional sampling is a known statistical technique (Cochran, 1977) and some characteristics of its application to map accuracy assessment are devised in

Stehman *et al.* (2000). However, DS3 has never been applied in published studies, maybe because it was not necessary for maps with classification systems of mainly temperate countries. Finally, an entirely novel, hybrid design (DS4) includes a simple random selection of PSUs (as in DS1) for common classes (area fraction above 5%, 7 classes in Cuitzeo), and a proportional stratified selection of PSUs (as in DS3) for rare classes (area fraction below 5%, 14 classes in Cuitzeo). After selection of the PSUs, the sample size of SSUs was fixed at 100 per mapped class, a value widely adopted in similar assessments (Stehman and Czaplewski, 1998).

Under these conditions, it was possible to test the efficiency of the sampling designs. According to the binomial distribution theory, the confidence interval of the accuracy estimate depends on the sample size and on the reliability value (accuracy estimate) in the following manner (Snedecor and Cochran, 1967; Fitzpatrick-Lins, 1981):

$$d^2 = t^2\, p\, (1\text{-}p)/n \tag{1}$$

Where d is the standard deviation (or half the confidence interval) of the estimate, t is the standard deviate on the Gaussian curve (for example, t = 1.96 for a two-sided probability of error of 0.05), p is the reliability value, and n is the number of sampled points. The standard deviation of the accuracy estimate was calculated for each class, as an indicator of the efficiency of the sampling design.

Verification design: The verification design is based on the interpretation of aerial photographs. A first photo-interpreter was assigned to delineate homogeneous polygons at scale 1:125 000 (finer resolution than the NFI map, and suitable for photo-interpretation) on transparent slides over the photographic material. This delineation was done on the photographic frames corresponding to the selected PSUs. Next, another interpreter digitized the photo-interpreted polygons over the Landsat geo-rectified space map. This step was meant to minimize the interference of fictitious positional errors between the NFI map and the material of reference (Zhu *et al*., 2000). The comparison of numerous repeated trials ensured that a good control over positional errors was achieved. Indeed, the skill of restitution on screen was acquired by the second interpreter from previous research, where photo-interpreted maps had to be digitized on screen over ortho-photographs or scanned topographic maps. This exercise usually serves the purpose of transferring stereo-pair interpretation work (a valuable traditional skill of physical geographers and technical foresters) onto a digital geo-referenced GIS.

Each sample point (selected SSU) was located on the colour composite of the geo-rectified Landsat mosaic on screen. The location of the point was then visually transferred onto the interpreted transparent slide over the photographic material, with the aid of the digitized polygons on screen. Finally, a confidence rating was associated with the photo-interpretation:

4. Very high
3. High
2. Average
1. Low

Prior to all the work described, a third interpreter had independently built a LULC map at 1:50,000 of the entire watershed based on the same set of 244 aerial photographs, and with the aid of intensive field work (López *et al*., 2006). When the confidence rating of the first interpreter was 1 or 2, his interpretation was confirmed or corrected according to the LULC map at 1:50,000.

Synthesis of the evaluation: In all sample points (selected SSUs), a comparison between the label of the reference material and the label of the NFI map was made via GIS, allowing a spatial tolerance of 500 m around the SSU in order to take into account the scale of the map (1:250 000). The derivation of the global accuracy index is detailed here for the evaluation with sampling design DS4, selected for the present analysis.

For any class k, a double sampling was applied. Following the Bayesian rule, the probability of inclusion of a sample point p_{2k} is a multiplicative function of the inclusion probability p_{1k} of the PSU it pertains to, and of the inclusion probability of the SSU once the PSU has been selected $p_{2|1}$ (conditional inclusion probability):

$$p_{2k} = p_{2|1} * p_{1k} \tag{2}$$

In the case of rare classes, the proportional stratified sampling is constructed in such a way that Eq. 2 simplifies into the inclusion probability of a simple stratified sampling (Stehman et al., 2000): $p_k = f_k$, where f_k is the frequency of class k on the map. In the case of common classes, the first stage probability p_{1k} is p_{1k}= K/N, where K is the fixed number of randomly selected PSUs for common classes and N is the total number of PSUs. $p_{2|1}$, the conditional probability, is equal to n_k/N'_k where n_k is the sample size and N'_k is the number of SSUs of class k in the K selected PSUs. Therefore, following Eq. 2, differential weights of f_k * K/N were attributed to common classes when global estimates were computed.

RESULTS AND DISCUSSION

With the target sample size of 100 per mapped class, we simulated 10 distributions of sample points for every one of the five sampling designs and examined the results. Some characteristics of the distributions are synthesized in Table 2, in order to evaluate the efficiency of each design.

The strategy of Simple Sampling (SS) yielded a dispersion o f the sample points over more than 85% of the PSUs (Table 2). As in other regional accuracy assessments, simple random sampling stratified by class is judged cost-prohibitive (here, in terms of purchase of photography and photo-interpretation labour). Accounting for the costs of operation, the reasonable number of photographic frames was fixed to a maximum of one fourth of the total coverage of the pilot area (one fourth of 218, or 54 photographic frames). Under this condition, the problem of non-statistically representative sample sizes was likely to occur. We required that the confidence interval for any class (Eq. 1) be less than 15% in order to consider the sample as statistically representative for this class.

The simulated distributions under DS1 left at least three (usually rare) classes with insufficient sample sizes (Table 2). This design is known for its good capacity of sample dispersion, counterbalancing the clustering effect of the double sampling. For this reason, Stehman *et al.* (2003) use it for the accuracy assessment of the National Land-Cover Data (NLCD) 1992 en EUA. However, the design was discarded for the pilot study on the Mexican NFI because it cannot handle satisfactorily the high taxonomic diversity (large number of rare classes) in the classification system.

Under DS2, the selection of PSUs is separately operated for each class, and all PSUs where class k is present have the same probability of inclusion for class k. In such a design, if a PSU only contains one point (SSU) of class k, this PSU has equal probability of being selected than another PSU that contains 50 points (SSUs) of class k. If class k is highly fragmented, a high number of PSUs is therefore, necessary in order to statistically cover class k. For this reason, simulations showed that more than 80 PSUs were necessary to cover the 21 classes, a quantity which overrides the admitted 54 photographic frames. Although less problematic than DS1 according to the number of underrepresented classes (Table 2), the DS2 design remains inadequate in this pilot study of the NFI accuracy assessment, presumably because of the presence of highly fragmented classes. In addition, the computation of inclusion probabilities in the second stage of the design proved complicated to handle via automatic GIS procedures.

By contrast, proportional sampling in DS3 maintains a low level of complexity in the calculations of accuracy indices (Stehman *et al.*, 2000) because all sample points of class k have equal probabilities of inclusion, whatever the PSU they pertain to. However, as in DS2, the selection of PSUs also operates separately for each one of the 21 classes.

With the constraint of 54 photographic frames, preliminary simulations incited us to limit PSUs to three per class. We observed that three PSUs were generally sufficient to sample a majority of rare classes, and approach their real distribution pattern. However, the limitation to three PSUs for common classes generated much spatial clumping of the sample, with respect to the real distribution pattern. These observations are illustrated in the last two columns of Table 2: The average amount of Sample Points (SSUs) per PSU raises to 26-36 in comparison with much lower values in simulations of previous designs. In the presence of much spatial clumping, the precision of the estimates is likely to be affected by the spatial autocorrelation of errors.

DS4 was designed with the objective of benefiting from the good spatial dispersion of DS1 (a characteristic which was exploited for the LULC map assessment in the USA) for common classes and from the advantages of DS3 (inclusion of all classes and easy estimate computation) for rare classes. Effectively, the selection of a group of PSUs for the entire set of 7 common classes permitted a dispersion of the sample comparable to the results of DS1 (Table 2). Overall, visual inspection of the results indicated that DS4 yielded the best spatial distribution of samples among double sampling strategies. At the same time, the number of PSUs could be extended to more than three for those rare classes which did not obtain a sufficient sample size in the simulations of DS3, maintaining the total number of PSUs at 54. In this way, all classes were guaranteed a statistically valid sample size. A sequence of ArcView and Visual Basic (for Excel) routines, operating on attributes of the sample points, was developed for proportional sampling in both stages. The Fig. 4 shows the set of selected PSUs as the result of a simulation using DS4. The Primary Sampling Units were selected using the hybrid sampling design. The hybrid design is based on simple random sampling of PSUs for common classes and on proportional stratified sampling of PSUs for rare classes. The National Forest Inventory map is in the background. The Fig. 5 shows the dispersion of sample points (selected SSUs) in the area of the Cuitzeo Eastern lagoon.

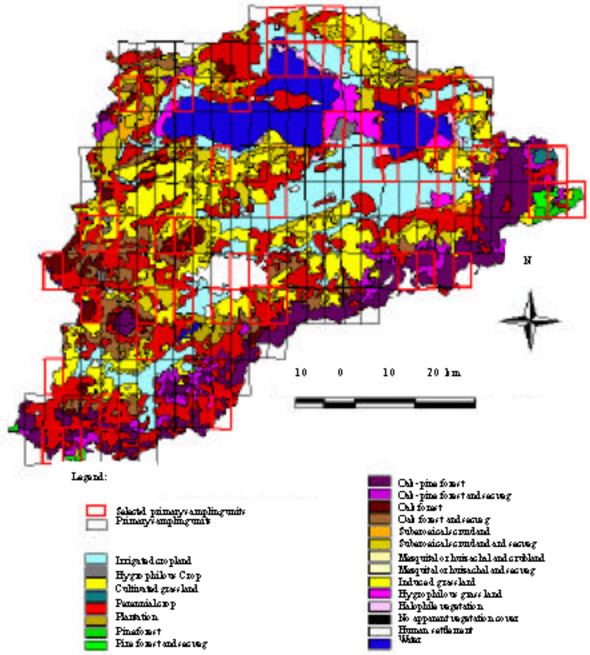

Fig. 4: Selected Primary Sampling Units on the watershed of the Cuitzeo lake

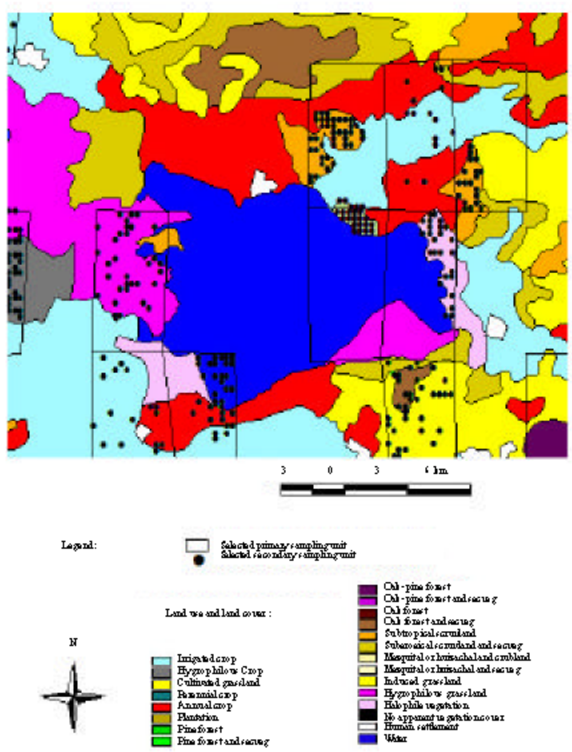

Fig. 5: Selected Secondary Sampling Units in the area of the Eastern Cuitzeo Lagoon

CONCLUSION

No comprehensive method has been published on the accuracy assessment of regional scale LULC maps in sub-tropical countries. We present a hybrid sampling design for the evaluation of LULC maps with high taxonomic diversity and highly fragmented landscapes. This and other four probability sampling designs-some of them previously employed for mainly temperate countries-were tested for the accuracy assessment of the Mexican National Forest Inventory (NFI) map at community level. Each sampling design was evaluated on the watershed of the Cuitzeo lake, a pilot study where a high number of classes is represented. As expected, the Double Sampling (DS) designs achieved much more control over the spatial distribution of the sample than a simple, random stratified sampling, thereby permitting the control over costs of operation. With fixed operational costs, the only design that systematically provided statistically representative estimates for all classes was the hybrid design (DS4). Additionally, the hybrid design achieved a spatial dispersion of the sample similar to the dispersion achieved by DS1, with simple random selection of Primary Sampling Units (PSUs). DS1 is known for generating a good dispersion of the sample in regional map assessments. For this reason, DS1 was successfully applied in the accuracy assessment of the National Land Cover Data (NLCD) in the USA (Stehman *et al*., 2003). However, DS1 was discarded in our pilot study because it was not able to handle the high number of rare classes of the NFI, a characteristic which the NFI also has at national scale. Instead, the hybrid design maintains simple random selection of PSUs for common classes, but applies a proportional stratified selection of PSUs for rare classes. This way, DS4 cumulates the advantage of a good sample dispersion for common classes, and the advantages of a sufficient sample size and easy estimate calculation for rare classes.

The entire accuracy assessment strategy, including the hybrid sampling design, permitted the evaluation of the pilot area within a probability sampling scheme. For all sampling designs, the dispersion of sample points remained low, due to the relatively small geographic extension of the pilot area, given the coarse scale (1:250 000) of the map and the high number of classes. For this reason, the average number of sample points per PSU remained relatively high for all designs. However, the hybrid sampling design and more generally the entire assessment design presented here is immediately transferable to more extended study areas, with typically available data and skills in Mexico. Specifically, its application to national scale is possible, using additional strata such as a systematic photographic coverage along flight lines (material initially planned in the NFI project) and administrative states.

ACKNOWLEDGEMENT

This research was done during the PhD thesis of the principal author, who receives a grant from the CONACYT public institution under scholarship number 205000.

REFERENCES

Cochran, W.G., 1977. Sampling Techniques, (3rd Edn.), John Wiley and Sons, New York, pp: 428.

FGDC, 1997. FGDC vegetation classification and information standard, June 3, 1996 draft. Reston, VA: Federal Geographic Data Committee, Vegetation Subcommittee (FGDC-VS), FGDC Secretariat.

Fitzpatrick-Lins, K., 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map, Photogrammetric Engineering and Remote Sensing, 47: 343-351.

Mas, J.F., A. Velázquez, J.L. Palacio-Prieto, G. Bocco, A. Peralta and J. Prado, 2002. Assessing forest resources in Mexico: Wall-to-wall land use/cover mapping. Photogrammetric Engineering and Remote Sensing 68: 966-969.

Laba, M., S.K. Gregory and J. Braden *et al.*, 2002. Conventional and fuzzy accuracy assessment of the New York Gap Analysis Project land cover map. Remote Sensing of Environ., 81: 443-455.

López, E., G. Bocco, M. Mendoza, A. Velázquez and R. Aguirre-Rivera, 2006. Peasant emigration and land-use change at the watershed level: A GIS-based approach in central Mexico, Agricultural Systems 90: 62-78.

Palacio-Prieto, J.L., G. Bocco, A. Velázquez, J.F. Mas, F. Takaki-Takaki, A. Victoria and L. Luna-González *et al.*, 2000. La condición actual de los recursos forestales en México: Resultados del Inventario Forestal Nacional, Investigaciones Geográficas, 43: 183-202.

Snedecor, G.W. and W.F. Cochran, 1967. Statistical methods, State University Press, Ames, Iowa, pp: 728.

Stehman, S.V. and R.L. Czaplewski, 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sensing of Environ., 64: 331-344.

Stehman, S.V., J.D. Wickham, J.H. Smith and L. Yang 2003. Thematic accuracy of the 1992 National Land-Cover Data for the eastern United-States: Statistical methodology and regional results. Remote Sensing of Environment 86: 500-516.

Stehman, S.V., J.D. Wickham, L. Yang and J.H. Smith, 2000. Assessing the Accuracy of Large-Area Land Cover Maps: Experiences from the Multi-Resolution Land-Cover Characteristics (MRLC) Project, 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Amsterdam, pp: 601-608.

Wickham, J.D., S.V. Stehman, J.H. Smith and L. Yang 2004. Thematic accuracy of the 1992 National Land-Cover Data for the western United-States. Remote Sensing of Environ., 91: 452-468.

Wulder, M.A., S.F. Franklin, J.C. White, J. Linke and S. Magnussen, 2006. An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data, Int. J. Remote Sensing, 27: 663-683.

Zhu, Z., L. Yang S.V. Stehman and R.L. Czaplewski, 2000. Accuracy Assessment for the U.S. Geological Survey Regional Land-Cover Mapping Program: New York and New Jersey Region, Photogrammetric Engineering and Remote Sensing, 66: 1425-1435.