Positional and thematic tolerance operators for the intercomparable accuracy measures of land use/ land-cover base-maps

Stephane Couturier
CentroGeo -Centre for Geography and Geomatics Research
(Centro de Investigación en Geografía y Geomática 'Ing. Jorge L. Tamayo' A.C.)
Mexico City, Mexico
Stephane.Couturier@CentroGeo.org.mx

LAGE –Laboratory of Geospatial Analysis
Instituto de Geografía
Universidad Nacional Autónoma de México (UNAM)
Mexico City, Mexico

***Abstract*—Accuracy assessments at regional scale are generally tailored to give one measure of error (either global or per class), which may vary substantially with the uncertainty contained in the assessment process. In fact, procedures are generally incorporated in the assessment to implicitly set a degree of tolerance regarding positional and/or thematic aspects, at which the map is evaluated. This research focusses on the adoption of a standardized response design for map accuracy assessments, enabling the comparison between agreement definitions among maps, with parametrized, moveable, degrees of tolerance. To this end, a formalization based on fuzzy GIS-based tolerance operators is described, whose continuous and discrete-field parameters are able to approximate major (positional and thematic) aspects of published regional accuracy assessment designs. The operators are applied to two cartographic datasets in Mexico, facilitating the comparison with the accuracy of the international cartography.**

***Keywords:** accuracy assessment; fuzzy map comparison; reference maplet; agreement definition.*

I. INTRODUCTION

The accuracy assessment of a map aims at estimating errors regarding its spatial and thematic representation with respect to reality. The accuracy assessment design is usually composed of the sampling process, the response design, and the synthesis of the evaluation (see Stehman & Czaplewski, 1998). The response design includes a protocol for reference data collection, and an agreement definition for the comparison of the reference data with the map.

It is well recognized that the reference material is also subject to error. Strategies have been employed in order to take into account reference material vagueness in the accuracy measures. For example, Stehman et al. (2003) employ, for the assessment of the 1992 National Land-Cover Data (NLCD) of the USA, two agreement definitions, leading to two sets of accuracy estimates (double accuracy assessments). Noticeably, the difference in per-class accuracy estimates between the two sets may be far greater than the confidence interval of the estimates themselves, which suggests a difficulty in comparing single accuracy measures of a map with the accuracy of others maps, measured with a different assessment design.

This research aims at describing accuracy assessment designs with enhanced potential for inter-comparison of the derived accuracy measures. Two features of the response design are proposed as key features to ease the inter-comparison of map accuracy. The first feature concerns the spatial extent of reference data collection. Reference maplets (Stoms, 1996) have been suggested as useful for one stage cluster sampling (Stehman, 2009). Although admittedly a constraint on the budget of the accuracy assessment, the reference maplets can be used with two stage sampling designs, where the secondary units are points, and where the maplet is used to explore spatial aspects of accuracy (Couturier et al., 2009). The other feature is the formulation or approximation of agreement definitions in terms of degrees of tolerance, conceived in three aspects/ dimensions: positional, thematic and temporal. The method sets accordingly the formulation of the reference data collection and of the agreement definition. Next, the utility of this design formulation is illustrated in the case of the accuracy of two forest cover cartographic datasets in Mexico. The aim was to compare the forest class accuracy indices with accuracy measures reported in international cartography.

II. CHARACTERISTICS OF CARTOGRAPHIC DATASETS

A. Datasets for Mexico

One cartographic dataset is the Landsat-based 2000 Mexican National Forest Inventory (NFI) map, assessed in the Candelaria watershed, the Cuitzeo lake watershed, and the Tancitaro region by Couturier et al. (in press). Another cartographic dataset is the MODIS- derived land cover map (MOD12Q1) of year 2004, assessed at the national level by Couturier (in press). The emphasis of this study was the accuracy of the maps for forest classes.

Sponsors for this research are DGAPA (PAPIIT project IN113609-2) and CentroGeo.

B. International Datasets

The international datasets include the 1992 NLCD of the USA assessed (Stehman et al., 2003) in New England (NE), New York/New Jersey (NY/NJ), and the Mid-Atlantic region (Mid-Atl), the 2000 EOSD map in Canada assessed in the Vancouver Island (Wulder et al., 2007), the European 2000 CORINE map assessed in Portugal (Caetano et al., 2006), the GLC 2000 map assessed in South- SouthEast Asia (SSE Asia, Stibig et al., 2007). The MOD12Q1 dataset has been assessed in Costa Rica (Kalacska et al., 2008), and globally by the MODIS team (cross-validation: MODIS, 2003). The attributes of all datasets were aggregated to the 16 classes IGBP legend. A (possibly strong) assumption is that the variability of forest class definitions across cartographic datasets will not affect the comparison of accuracy measures.

III. METHODS

In this paper the response design is the only step of the accuracy assessment that we describe. The sampling design and synthesis of the evaluation for the accuracy measures of the Mexican datasets are described in Couturier et al. (in press) and Couturier (in press). The response design supposes the selection of points as sampling units and is based on a reference maplet construction and on the application of two tolerance operators to the reference maplet. The latter step is assumed equivalent to setting agreement definitions. The agreement tests are then processed in the form of a fuzzy map comparison, in the sense of Hagen (2003).

A. Reference Maplet Construction

The reference material for the accuracy assessment of the NFI 2000 was a set of aerial photographs, used as PSUs (Couturier et al., 2009). The reference maplet construction consisted in interpreting, with a linguistic fuzzy representation, the full selected PSUs, and correcting the polygons on screen over the georectified Landsat scenes which supported the production of the NFI 2000 map. The reference material for the accuracy assessment of the MOD12Q1 2004 product was based on two datasets: the local scale database of the National Inventory of Forests and Soils (INFyS) in Mexico, derived from a spatially exhaustive, cyclic campaign of ground visits across years 2003 to 2007 (see Couturier, in press); and the regional scale Landsat-derived 2003 LULC cartography of the National Institute for Statistics, Geography and Informatics (INEGI) in Mexico. The construction of the reference maplet consisted in a fuzzy union between the local and regional datasets.

B. The Minimum Tolerance Vector

The classification map *C* and the set of reference maplets *R* can be represented by the fuzzy vectors V_C and V_R, where all points contain degrees of membership μk to each class k of a common legend, as follows:

$$V_M = \begin{pmatrix} \mu_{M,1} \\ \mu_{M,2} \\ \\ \mu_{M,K} \end{pmatrix} \qquad (1)$$

For map *M* being either *C* or *R*. K is the total number of classes in the common legend. We name C' the foreign map to be compared with, with known accuracy measure and agreement definition. Three phases are considered. Phase 1 consists in setting the membership values of V_R to a maximum of 2 (usually the strictest agreement definition across map assessments in terms of thematic tolerance, e.g. consider primary and secondary labels). The accuracy measure characterizing phase 1 would be the fuzzy intersect of V_C with the resulting '*hard*' V_R ($V_{R,hard}$) over the universe of selected sampling units:

$$V_{strict} = V_{C'} \cap V_{R,hard} \qquad (2)$$

C. Increasing Positional Tolerance

Phase 2 consists in increasing the membership values of $V_{R,hard}$ in the sample sites, according to a fuzzy union with neighbouring elements of $V_{R,hard}$:

$$V_R^{positional} = V_{R,hard} \oplus \tau_{pos} \qquad (3)$$

τ_{pos} being a distance value for positional tolerance. The fuzzy union is made with all instances of $V_{R,hard}$ within a shift smaller or equal to τ_{pos} in any direction, as suggested by Hagen (2003) with a neighbouring function. The accuracy measure characterizing phase 2 is a fuzzy intersect of V_C with $V_{R,positional}$, τ_{pos} being set to the tolerance observed in the agreement definition of the C' assessment. The analysis of the spatial aspect of each international assessment yielded a distance value (e.g 30m or 1 km), in terms of positional tolerance, as defined in this work.

D. Increasing Thematic Tolerance

Finally, phase 3 consists in adding more thematic tolerance to the reference map; $V_{R,positional}$ is softened, incorporating the membership functions of V_R, up to a membership value observed in the C' map assessment. This membership value is called the thematic tolerance (τ_{th}) of the assessement, and V_R is truncated to $V_{R,th}$ with maximum membership value τ_{th}. In this case study, since most assessments tended to apply a thematic tolerance of 2, τ_{th} was set equal to 2. Phase 3 yields the accuracy vector with enhanced compatibility with the C' accuracy measures:

$$V_R^{pos+them} = V_R^{positional} \cup V_{R,th} \qquad (4)$$

Eventually, in order to allow higher degrees of membership, a fuzzy union is previously processed with additional reference datasets available (Couturier, in press).

E. Consideration for Temporal Tolerance

Reference datasets are not exactly synchronous with the material for map production. A third dimension worth stating in the accuracy comparison process is the delay date of the reference material with respect to the map. This delay is another dimension of tolerance in accuracy assessments, since some uncertainty is implied, with generally increasing uncertainty when the delay is increased. However, purposely moving this 'tolerance degree' makes little sense *a priori*.

IV. RESULTS

Tables 1 and 2 present forest class accuracy indices for international cartography (as reported in the literature and assimilated in our framework) and accuracy indices calculated for the Mexican cartography, at various levels of positional tolerance. Table 1 makes emphasis on user's accuracy indices of the IFN 2000. Table 2 deals with the producer's accuracy indices of the MOD12Q1 product in Mexico and of some 500m - 1km resolution international cartography. At first glance, a slight difference in positional tolerance may result in a significant change in accuracy measure (compared with confidence intervals), especially for the NLCD high resolution cartography. For this reason, it is difficult to compare accuracy beween the CORINE 2000 dataset and the North American cartography, for example.

A. NFI 2000 Cartography

Keeping in mind the highly differing characteristics of the datasets and our possible misinterpretation of classes and assessment methods, there seems to be a general agreement for the accuracy of the overall forest class, especifically looking at it with a 30m positional tolerance. NLCD Mid-Atl lower accuracy can be related to the lower thematic tolerance at which the map has been effectively evaluated (only 1 membership value). The GLC2000 lower accuracy value is likely related to the lower spatial resolution of the imagery for map production and the greater area span of the assessment with respect to other cartography assessments.

There is much more accuracy variability for individual forest classes, possibly reflecting many factors, including class definitions (not necessarily accuracy *per se*). With this caution, some tendencies still appear: Much evergreen needleleaf forest seems well mapped in Canada and erroneously mapped in Mexico. Mixed needleleaf-broadleaf forests seem better mapped in Mexico than in the USA.

Accuracy values reported for CORINE 2000 seem much higher than those of the American continent, although some misunderstanding of the agreement definition for CORINE 2000 may result in a biased comparison. Besides, working with polygons of the map for the delineation of reference material, although a practical strategy, may affect the spatial independence of the reference data with respect to the map. This assessment technique may thus cause an optimistic bias in accuracy measurements.

TABLE II. PRODUCER'S ACCURACY FOR MOD12Q1 AND OTHER INTERNATIONAL CARTOGRAPHY

Producer's Accuracy	Map:	MOD12Q1			CORINE 2000	GLC 2000
	Positional Tolerance (m) :	*Mexico*	*Costa Rica*	*Cross Valid.*	*Portugal*	*SSE Asia*
Evergreen neadleleaf forest	*0*	27 +-2				
	100	27 +-2			84 +-15	
	500	38 +-3				
	1000	46 +-3		60 +-2		
Mixed (neadle-broadleaf) forest	*0*					
	100				69 +-13	
	500					
	1000			61+-2		
Deciduous broadleaf forest	*0*	22 +-1	35			
	100	24 +-1			83 +-5	
	500	32 +-2				
	1000	39 +-2		34 +-3		
Evergreen broadleaf (tropical) forest	*0*	78 +-1	88			
	100	80 +-1				
	500	84 +-1				
	1000	89 +-1		90 +-1		
Overall forest	*0*	63 +-2	76			
	100	65 +-2				
	500	68 +-2				78 +-6
	1000	72 +-2		89 +-1		

TABLE I. USER'S ACCURACY FOR INTERNATIONAL CARTOGRAPHY AND THE MEXICAN NFI 2000 MAP

User's Accuracy	Map:	NLCD 1992 (USA)			EOSD Canada	CORINE 2000	GLC 2000	NFI 2000 Mexico		
	Positional Tolerance (m) :	*NE*	*NY/NJ*	*Mid-Atl*	*Vancouver*	*Portugal*	*SSE Asia*	*Tancitaro*	*Cuitzeo*	*Candelaria*
Evergreen neadleleaf forest	*0*	41 +-5	84 +-4	33 +-7	85 +-4			34 +-5		
	30	38 +-5	92 +-4	39 +- 7	86 +-4			34 +-5		
	100					60 +-15		35 +-4		
	500							37 +-4		
Mixed (neadle-broadleaf) forest	*0*	37 +-5	44 +-4	21 +-6				52 +-5	60 +-8	
	30	39 +-5	55 +-4	32 +-6				54 +-5	63 +-8	
	100					73+-13		60 +-4	66 +-9	
	500							66 +-3	72 +-9	
Deciduous broadleaf forest	*0*	48 +-8	57 +-3	40 +-9					49 +-10	
	30	48+-8	61 +-3	47 +-9					50 +-10	
	100					94 +-3			54 +-10	
	500								64 +-11	
Evergreen broadleaf (tropical) forest	*0*									87 +-9
	30									87 +-9
	100									87 +-9
	500									89 +-8
Overall forest	*0*	82 +-2	80 +-2	65 +-4	73 +-3			79 +-2	78 +-5	86 +-6
	30	82 +-2	84 +-2	70 +-4	77 +-3			79 +-2	78 +-5	86 +-6
	100							80 +-2	82 +-5	87 +-5
	500						72 +-6	83 +-2	86 +-5	88 +-5

B. MOD12Q1 Cartography

A similar tendency of accuracy characterizes MOD12Q1 in Mexico and Costa Rica, agreeing with the tendency registered in the cross-validation study. Nonetheless, the cross-validation absolute accuracy values are higher for all classes but the deciduous broadleaf forest class, probably because of the optimitic bias inherent to cross-validation with homogeneous sites. Results in Costa Rica are close to results in Mexico at higher positional tolerance. Maybe some misunderstanding of the agreement definition in the Costa Rica dataset explains this systematic shift in accuracy measures.

V. Discussion on the Method

The fuzzy map comparison scheme has been used in previous research (Couturier et al., 2009) in order to estimate the contribution of positional and thematic vagueness of reference material in the accuracy indices of a single map. Yet, the idea of a set of tolerance degrees may be considered elsewhere as an imprecise way to approximate the accuracy of maps. Ideally, uncertainty in the map as well as in the reference maplet could be represented by a set of quantitative fuzzy figures. In this case, Binaghi et al. (1999) constructed a fuzzy error matrix where the accuracy of the map is estimated by a single set of measures. In this sense, Stehman (2009) identified sampling designs which would include reference maplets and where the fuzzy map comparison formulated by Hagen (2003) would accomodate reference data uncertainty.

However, in all operational accuracy assessments so far, only linguistic fuzzy definitions (with discrete membership functions) have been employed for thematic uncertainty, perhaps because linguistics (still?) better characterize land cover reference materials, especially visual classification of nominal categories. In that case, the notion of thematic tolerance degrees with discrete membership values is thought to be well adapted, and associates well with a positional tolerance degree for estimating map accuracy at inter-comparable scales.

VI. Conclusions

Much variability is registered in accuracy measures, even for a single map. The response design as formulated in this research may reduce the difficulty in comparing accuracy measures in spite of this intrinsic variability. The parametrization embedded in the operators of the response design, is, admittedly, not a substitute for the full description of the response design, which contains much more complex information. Besides, the design may imply increased costs for the use of purpose- acquired, spatially extended reference maplets, with respect to traditional point-like reference data collection. The assignment of degrees of tolerance to accuracy assessments could appear as a lower precision method than an assessment with a single measure. However, multiple measures have been the operational practice in the remote sensing community (e.g. Powell et al. 2004) and not single (not even quantitative fuzzy) measures.

In the coming years, with the increasing variety of possible reference data, global environmental agreements may require countries to meet common accuracy criteria. The two examples of this study illustrated the practicality of accuracy scenarii, obtained through flexible degrees of tolerance, for the inter-comparison of cartographic data.

Acknowledgements

I am indebted to J.-F. Mas for his thoughtful advice on multiscale data management and to the CONAFOR and INEGI technical teams for providing the INFyS and Series III datasets, respectively.

References

Binaghi, E., P.A. Brivio, P. Ghezzi, and A. Rampini (1999). A fuzzy set-based accuracy assessment of soft classification, *Pattern Recognition Letters,* 20, 935-948.

Caetano, M., F. Mata, and S. Freire, 2006, Accuracy assessment of the Portuguese CORINE Land Cover map, VII[th] International Symposium on Spatial Accuracy Assessment in *Natural Resources and Environmental Sciences 'Accuracy 2006'*, Lisboa, 5-7 julio 2006.

Couturier S., J.-F. Mas, G. Cuevas, J. Benítez, A. Vega-Guzmán, and V. Coria-Tapia (2009). An accuracy index with positional and thematic fuzzy bounds for land cover/land use maps. *Photogrammetric Engineering and Remote Sensing.* 75 (7), 789-805.

Couturier, S., J.-F. Mas, E. López-Granados, J. Benítez, V. Coria-Tapia, and A. Vega-Guzmán (in press). Accuracy assessment of the Mexican National Forest Inventory map: a study in four eco-geographical areas. *Singapore Journal of Tropical Geography.*

Couturier, S. (in press). A fuzzy-based method for the regional validation of global maps: the case of MODIS- derived phenological classes in a mega- diverse zone. *International Journal of Remote Sensing.*

Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science.* 17 (3), 235-249.

Kalacska, M., G.A. Sánchez-Azofeifa, B. Rivard, J.C. Calvo- Alvarado, and M. Quesada (2008). Baseline assessment for environmental services payments from satellite imagery: A case study from Costa Rica and Mexico. *Journal of Environmental Management.* 88, 348-359

MODIS (2003). Validation of the consistent year 2003 V003 MODIS land cover product. Available online at: http://geography.bu.edu/ landcover/ userguidelc/consistent.htm (accessed May 2009).

Powell, R.L., N. Matzke, C. de Souza, M. Clark, I. Numata, L.L. Hess, and D.A. Roberts (2004). Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon, *Remote Sensing of Environment.* 90, 21-234.

Stehman, S.V., and R.L. Czaplewski (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment.* 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. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing.* 30 (20), 5243-5272.

Stibig, H.J., A.S. Belward, P.S. Roy, and U. Rosalina-Wasrin (2007). A land-cover map for South and Southeast Asia derived from SPOT-VEGETATION data. *Journal of Biogeography.* 34, 625-637.

Stoms DM (1996). Validating large land cover databases with maplets. *Geocarto International.* 11(2), 87-95.

Wulder, M.A. (2007). Validation of a large area land cover product using purpose- acquired airborne video. *Remote Sensing of Environment.* 106, 480-491.