Designing a reference validation database for accuracy assessment of land cover

Stephen Stehman
College of Environmental Science & Forestry
State University of New York
Syracuse, NY, USA
svstehma@syr.edu

Pontus Olofsson, Curtis Woodcock, Mark Friedl,
Adam Sibley, Jared Newell,
and Damien Sulla-Menashe
Department of Geography and Environment
Boston University Boston, MA,
USA
olofsson@bu.edu

Martin Herold
Laboratory of Geo-Information Science and Remote Sensing
Wageningen University
Wageningen, NL
Martin.Herold@wur.nl

Abstract—The cost efficiency and consistency of accuracy assessments of global and regional land-cover maps would be greatly enhanced by creating a global reference validation database applicable to a variety land-cover maps. This validation database would have at its core a common underlying rigorous sampling design and a consistent protocol for interpreting the reference (ground condition) classification against which the map classification would be compared. The basic sampling design proposed for collecting reference data is a cluster design stratified based on Köppen climate classes and population density. Some features of a sample of 500 clusters selected by this design are described.

Keywords: sampling design; stratified random sampling; error matrix

I. INTRODUCTION

The accuracy of global and regional land-cover products must be documented to allow a user to evaluate the utility of a map for a particular application. Subjecting every land-cover product to a separate, independent accuracy assessment is expensive and inefficient, and the variety of different accuracy assessment protocols implemented hinders comparisons of accuracy among maps. A coordinated, global land cover validation database would be a cost efficient approach to validating multiple large-area land-cover products and would enhance our ability to compare accuracy among different maps.

The methodology for constructing a global validation reference database is based on a rigorous probability sample of reference land-cover data that could be used to assess accuracy of different land-cover products. This global reference database would serve as a baseline sample that could be augmented with additional sample data to tailor an assessment to the objectives specific to a particular land-cover product. Because the proposed global validation database is constructed from a common underlying sampling design and a consistent response design protocol, additional reference data collected to augment the initial sample can be readily merged into the common database for application to other land-cover maps as long as the sampling and response design protocols of the global validation database are adhered to. The global validation methodology described addresses accuracy of one-point-in-time maps, but the general strategy can be extended to provide a global validation database for assessing accuracy of land-cover change.

II. RESPONSE DESIGN AND ANALYSIS

The basic spatial assessment unit of the proposed global validation database is a 5-km x 5-km block. Each block will constitute a “cluster” of pixels for any particular land-cover map subject to validation. Within each sampled cluster, a map of the reference land cover will be interpreted from high resolution satellite imagery (e.g. QuickBird) according to the United Nation’s Food and Agriculture Organization’s (UN FAO) Land Cover Classification System (LCCS). The basis of the accuracy assessment is the comparison, within each sample block, of the reference “map” to the land cover of the map being evaluated. If the reference land cover is converted to the pixel resolution of the land-cover map being evaluated, a traditional per-pixel accuracy assessment can be obtained. That is, the global validation sample can be used to estimate an error matrix and accuracy parameters associated with an error matrix such as overall, user’s, and producer’s accuracies (Stehman and Foody, 2009). Alternatively, the proportion of area covered by each land cover type based on the reference classification could be compared to the corresponding area proportions derived from the map classification. In this latter approach, the analysis would focus on estimating measures of agreement for the continuous area proportions. That is, the parameters summarizing accuracy may include mean absolute deviation, root mean square error, and correlation, with accuracy quantified by these parameters being specific to the support of the 5-km blocks.

III. SAMPLING DESIGN

To create the frame for implementing the sampling design, the land surface of the earth is partitioned into 5-km x 5-km blocks. The sampling design is then constructed to satisfy the definition of a probability sampling design, provide adequate sample sizes for rare land-cover classes, and allow flexibility to easily augment the sample to add sample clusters in particular regions or strata to tailor the assessment to a particular map. Stratified random sampling achieves all three of these major desirable design criteria in a simple and
straightforward fashion. Moreover, stratified sampling is a familiar probability sampling design implemented for accuracy assessment for the purpose of increasing the sample size for rare classes (Stehman, 2009).

Because the global validation database is designed to be useful for a broad array of global land-cover products, stratifying by any single land-cover map would diminish the advantage of stratification when evaluating other maps. Therefore, the strata are constructed from a more general-purpose perspective using a combination of Köppen climate classes and population density. The rationale for stratification is still to provide a mechanism for allocating a larger sample size to rare land-cover classes. The strata based on the Köppen climate classes and population density are effective for this purpose to the extent that increasing the sample allocation within a certain stratum will increase the sample size of all land-cover classes associated with that stratum. Clearly these more generalized strata will sacrifice some of the targeted sample size control attained by stratifying based on individual land-cover classes, but these general strata render the base validation sample more broadly useful to all global land-cover products.

The Köppen climate system classifies the world into five major climatic groups based on annual and monthly air temperature and precipitation. Each of these five major groups is further divided into subgroups based on seasonal patterns of temperature and precipitation. The climate groups are intended to correspond to vegetation groups (Trewartha, 1968) and thus to natural land-cover types. The initial stratification is based on an updated version of the Köppen system (Peel et al., 2007) in which 32 classes are identified at 0.1° resolution. The 32 climatic groups are then collapsed to 13 by merging entire classes (e.g. “Desert, hot” with “Desert, cold”).

To incorporate the impact of human activity on land-cover complexity, population density is added to the stratification. A global population density map from 2000 (CIESIN, 2005) at a resolution of 2.5 arc-minutes (4.6 km at the equator) is used for this purpose. A threshold of 5 persons per km² is used to separate between populated and unpopulated areas, and some of the climate-based strata are then further partitioned into “populated” and “unpopulated” strata. An additional stratum is created of all areas having a population density of 1000 persons per km². The 21 strata resulting from this process are displayed in Figure 1.

IV. GLOBAL VALIDATION SAMPLE

An initial base global validation sample of 500 blocks was selected. The sample size of 500 blocks is assumed practically manageable, yet large enough that the estimates of overall accuracy and some user’s accuracies will be reasonably precise. The 500 sample clusters are allocated to the 21 strata so that complex, more ambiguous (but not necessarily rare) land-cover classes receive a relatively high sample size. Heterogeneous landscapes with a mix of different classes such as mosaics of cropland and natural vegetation and built-up areas are considered complex whereas homogeneous land covers such as deserts and stable forests are not. Land-cover classes that are more likely to be misclassified are of higher interest.

Because a stratum may include several land-cover classes, increasing the sample size in a stratum will increase the sample size of several land-cover classes. For example, increasing the sample size in the urban stratum would boost the sample size of land-cover classes that tend to be proximate to urban areas such as cropland and cropland mosaic classes in addition to increasing the sample size of the urban class itself. The global distribution of the sample is shown in Figure 2.

The distribution of land cover within the base sample of 500 blocks is examined for the 17 classes of the MODIS LC IGBP classification (Figure 3). The proportion of area of each land-cover class for the full map is compared to the proportion of area of each class in the sample. The stratified design is constructed to increase the sample proportion in several classes of high interest. The most notable classes in which the sample proportion is increased relative to the population proportion are urban, cropland, and mixed deciduous forest. Open shrubland and barren/sparse vegetation are the two most common classes mapped, and the sample area proportions of these two classes are reduced considerably relative to their area proportions in the full map. This is a desirable outcome of the stratification.

V. CONCLUSIONS

The stratified random sampling design underlying the base validation sample achieves the desirable criteria specified for the design. The stratified random design allows for augmenting the base sample of 500 blocks to address objectives specific to a particular land-cover product. For example, it is straightforward to increase the sample size within any of the 21 strata to target specific land-cover types or specific regions (e.g. a country), while still maintaining the probability sampling character of the design. It should also be possible to merge the base validation sample with other probability samples to provide an approach for taking advantage of existing accuracy assessment reference data in those cases where the existing data meet the response design requirements of the global validation database. The specific methods for creating a global validation database applicable to assess accuracy of a broad variety of land-cover maps continue to be developed. The response design protocol needs to be completed and tested, and full details of the analysis and format for reporting accuracy results are still being developed.

REFERENCES


Figure 1. Stratification based on Köppen climate classes and population density.

Figure 2. Global distribution of 500 sample clusters.

Figure 3. Distribution of land cover for the map (blue) and 500 sample blocks (red) for the MODIS LC IGBP global land-cover map.