**Incorporating Model and Reference Uncertainty into Accurate Assessments: A proposed generalization of current best practices**

Accurate and precise estimates of land-use and land cover (LULC) are critical for making wise management decisions. Probability sampling designs are statistically robust strategies for working with accuracy assessment data of LULC classifications, to produce bias-adjusted estimates of omission/commission error and class area. While today’s state of the art has excellent statistical treatment of errors, it does not yet fully incorporate the important aspect of variation in confidence within a study area. From both the perspective of a given LULC model and of reference data, the label at any given point may well be less than 100% certain. For example, imagine a setting where the LULC model estimates that a given plot is 63% likely to have been deforested; meanwhile, 3 of 4 independent interpreters called it deforested. Considering the potentially large impact on accuracy and area estimates of incorrectly labeled information, how could the uncertainties at that point be considered when estimating the overall accuracy of the classes and map, in a way that respects and builds on the careful work that forms the state of the art?

Here we propose an approach to treating varying confidence when making estimates from LULC classifications. It is designed to consider both model uncertainty and reference uncertainty in weighting a reference data point for an accuracy assessment. Thinking of each reference and/or model as being a vector of LULC probabilities at each sample point, we use a dominance measure to assign a confidence-driven weight. Scores range from unity, meaning perfectly certain it is a given class, to 0, representing complete uncertainty about a point. The resulting contingency table is analyzed using “effective points”, where uncertain points are valued as less powerful than those with higher-confidence labels. Once the power of each point is weighed, estimates of area can proceed using the established calculations to produce bias-adjusted values.

The resulting calculations have several appealing characteristics. First, they produce reasonable differentiation among points with different confidence. Points of low confidence are able to be neither trusted completely nor discarded as though they had no value. Second, points of different weights are combined into a familiar contingency table. Third, the method is agnostic to how uncertainties are determined, allowing users to define uncertainties in whatever way is best for a given project. Fourth, when there is 100% confidence in both the model and reference data, the calculations give the same result as current best practices.

In addition to explaining this generalization, we provide a working code repository that can be used in accuracy assessment estimation. Test data includes working examples of typical use cases with data from some best-practices research. By proposing a straightforward way to integrate uncertainty in reference and/or model data, we see this work as a step toward accuracy assessments that more fully reflect real settings where fair estimates are needed that weigh multiple data sources. We encourage others to try it out and work with us to evaluate the usefulness and robustness of this approach.