Geophysical Research Abstracts
Vol. 20, EGU2018-6323, 2018
EGU General Assembly 2018

A comparison of resampling methods for remote sensing classification and accuracy assessment

Mitchell Lyons (1), David Keith (1,2), Stuart Phinn (3), Tanya Mason (1), and Jane Elith (4)

(1) University of New South Wales, Australia (mitchell.lyons@gmail.com), (2) NSW Office of Environment and Heritage, Australia, (3) University of Queensland, Australia, (4) University of Melbourne, Australia

Maps that categorise the landscape into discrete units are a cornerstone of many scientific, management and conservation activities. The accuracy of these maps is often critical in determining both the quality of the map as well as being directly used in subsequent applications of the map data. Variance and uncertainty are critical components of accuracy, yet commonly reported accuracy metrics often do not provide this information. Various resampling frameworks have been proposed to reconcile this issue, but have had limited uptake. In this work, we compare the traditional approach of a single split of data into a training set (for classification) and test set (for accuracy assessment), to a resampling framework where the classification and accuracy assessment are repeated many times. Using a vegetation mapping example and two common classifiers (maximum likelihood and random forest), we investigate uncertainty in mapped area estimates and accuracy assessment metrics (overall, kappa, user, producer, entropy, purity, quantity/allocation disagreement). Input data were repeatedly split into training and test sets of various designs via bootstrapping, Monte Carlo cross-validation (67:33 and 80:20 split ratios) and k-fold (5-fold) cross-validation. Additionally, within the cross-validation, four stratification designs were tested: simple random, block hold-out, stratification by class, and stratification by both class and space. A classification was performed on every split for each combination of sampling design, creating sampling distributions for mapped areas and accuracy metrics. We found that regardless of resampling design, a single split of data into training and test sets results in a large variance in estimates of accuracy and mapped area. In the worst case, overall accuracy varied between ∼40-80% in one resampling design, due only to random variation in partitioning into training and test sets. On the other hand, we found that essentially any resampling design provides an accurate estimate of error, and unlike a single iteration, provides confidence intervals that are informative about the performance and uncertainty of the classification. Importantly, we show that these confidence intervals commonly encompassed the magnitudes of increase or decrease in accuracy that are often cited in literature as justification for methodological or sampling design choices. We make recommendations about what resampling design to use and how it could be implemented. We also show how a resampling approach can be used to generate spatially continuous estimates of mapping accuracy and uncertainty.