

QUANTIFYING ERROR IN VEGETATION MAPPING

Erica Serna

Drs. Hong He, Daniel Dey and John Fresen, Thesis Advisers

ABSTRACT

Understanding the current distribution and structure of forest vegetation is important for designing forest management plans and prioritizing restoration at landscape scales. This project provides information on Random Forest, a relatively new statistical package in the field of forestry, and patterns in mapping errors, a less explored field of study particularly in the forests of the Midwest United States. Vegetation maps can be made from classification and regression trees, such as Random Forest, by integrating environmental variables with vegetation information. An understanding of the accuracy of the maps is important because management plans and restoration efforts are more effective with accurate data. This study was done in forested regions in Minnesota with the purpose of 1) analyzing physiographic factors controlling tree species distribution; 2) mapping potential species distributions; 3) quantifying error in vegetation mapping; and 4) understanding map accuracy by evaluating minimum amounts of sample data necessary for reliable mapping. The results from Random Forest were found to be realistic ecologically and biologically. Also, tree species required records of 1-2 trees per 10,000 ha to produce accurate maps. Knowing the minimum amount of data points necessary for acceptable accuracy assists scientists mapping vegetation. This study demonstrates the effectiveness of Random Forest in vegetation mapping, which can be useful for future vegetation mapping.