

Approximating Prediction Uncertainty for Random Forest Regression Models

J. W. Coulston^{a,*}, C. E. Blinn^b, V. A. Thomas^b, R. H. Wynne^b

^a Forest Inventory and Analysis, USDA-Forest Service, Knoxville, TN, USA – jcoulston@fs.fed.us

^b Dept. of Forest Resources & Environmental Conservation, Virginia Tech, Blacksburg, VA, USA – cblinn@vt.edu

Commission VI, WG VI/4

KEY WORDS: Estimation, Land Cover, Statistics, Modelling, Mapping, Prediction, Landsat

ABSTRACT:

The use of machine learning approaches such as random forests has increased for the spatial modeling and mapping of continuous variables. These include, for example, percent tree canopy cover and forest biomass modeling. Random forest is a non-parametric ensemble approach and unlike traditional regression approaches, there is no direct quantification of prediction error. Understanding prediction uncertainty is essential particularly when using model-based continuous maps as inputs to other modeling applications such as fire modeling. In addition, understanding prediction uncertainty is important for discriminating significant change from modeling error. In this work we are concerned with the prediction uncertainty when predicting a value for a new observation. We develop a non-parametric approach to quantify prediction uncertainty and demonstrate the method for a random forest model of forest tree density.

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.