Mapping forest species and plant communities with Gradient Nearest Neighbor imputation

Janet L. Ohmann,[1] Kenneth B. Pierce,[2] and Matthew J. Gregory[3]

Abstract: Gradient Nearest Neighbor (GNN) imputation had its genesis within a large, interdisciplinary research project that aimed to model the effects of alternative forest policies on socio-economic and biophysical responses across a large, multi-ownership region. To initialize the current landscape for simulation modeling, the team of scientists required vegetation information that was spatially complete (i.e., mapped), consistent across all ownerships, of fine spatial resolution (25-m pixels), contained detailed attributes on forest species composition and structure (i.e., a 'tree list') for each pixel, and reasonably represented both fine-scale heterogeneity and landscape pattern. The vegetation map needed to support stand projection models, landscape simulation models, and various response models for wildlife habitat, other biodiversity measures, and timber production. The GNN flavor of imputation was born in particular from the need to accurately map species composition. Methods underlying existing imputation methods assumed a linear relationship between species (response) variables and environmental (explanatory) variables, whereas species responses to environmental gradients are most often gaussian. Furthermore, in large regions where gradients are long and species turnover is high, data matrices of species relative abundance on plots are quite sparse (contain many zeroes). Many multivariate methods have been shown to perform poorly on these sparse matrices. For these reasons, we adapted nearest-neighbor imputation to use canonical correspondence analysis (CCA) as the underlying statistical model and distance measure. CCA has a long track record and acceptance by ecologists seeking to quantify relationships between biotic communities and environmental gradients, but has rarely been used for prediction in general, nor for mapping in particular, and is unfamiliar to many statisticians. We have now applied GNN in five contrasting forested ecoregions in Washington, Oregon, and California, in studies focused on policy analysis, natural resource planning, fuel conditions and fire risk, and conservation planning. I will present and compare results of spatial predictions of individual tree species and forest vegetation types from these ecoregions, including several measures of map accuracy. I will discuss the various advantages and limitations of GNN for mapping species and communities, based on what we've learned so far, and suggest possible areas for future research and collaboration.

[1] Pacific Northwest Research Station, USDA Forest Service, Forestry Sciences Lab, 3200 SW Jefferson Way, Corvallis, OR 97331. E-mail: janet.ohmann@oregonstate.edu.

[2] Pacific Northwest Research Station, USDA Forest Service, Forestry Sciences Lab, 3200 SW Jefferson Way, Corvallis, OR 97331. E-mail: ken.pierce@oregonstate.edu.

[3] Department of Forest Science, Oregon State University, Richardson Hall 321, Corvallis, OR 97331. E-mail: matt.gregory@oregonstate.edu.