

## Variance of similar neighbors compared to random imputation

Kenneth B. Pierce

PNW Research Lab, USDA Forest Service, Corvallis, USA,  
3200 SW Jefferson Way, Corvallis, OR 97330, [ken.pierce@oregonstate.edu](mailto:ken.pierce@oregonstate.edu): 541 750-7393

Janet L. Ohmann

PNW Research Lab, USDA Forest Service, Corvallis, USA,  
3200 SW Jefferson Way, Corvallis, OR 97330, [janet.ohmann@oregonstate.edu](mailto:janet.ohmann@oregonstate.edu): 541 750-7487

**Abstract:** Nearest-neighbor methods are based on the premise that similarity in completely sampled covariates can be a good predictor of similarity in sampled variates. For instance, plots with similar environmental and spectral characteristics should have similar vegetation. At the very least, potential imputation targets should be less variable than random targets. We set out to test these assumptions by examining the distribution of the variances drawn from the nearest five neighbors for an unsampled location and 5 randomly sampled neighbors. We also built subsets of our plots based on the number of neighbors available within a predefined distance in multivariate gradient-space in order to see if well sampled locations perform better than poorly sampled locations. We performed this analysis in three locations known to provide varying results using imputation: coastal Oregon, eastern Washington and the central Sierra Nevada. The results varied markedly between sites and, at times, in non-intuitive ways. The results from restricting the targets based on sampling sufficiency often yielded better results for poorly sampled areas than those areas well sampled in the plot inventory. We discuss these results in the context of implications for forest inventory sampling frameworks.