

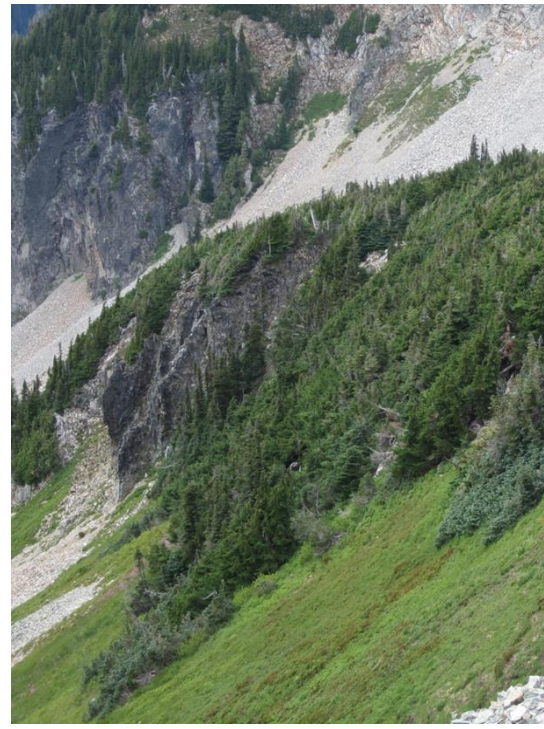
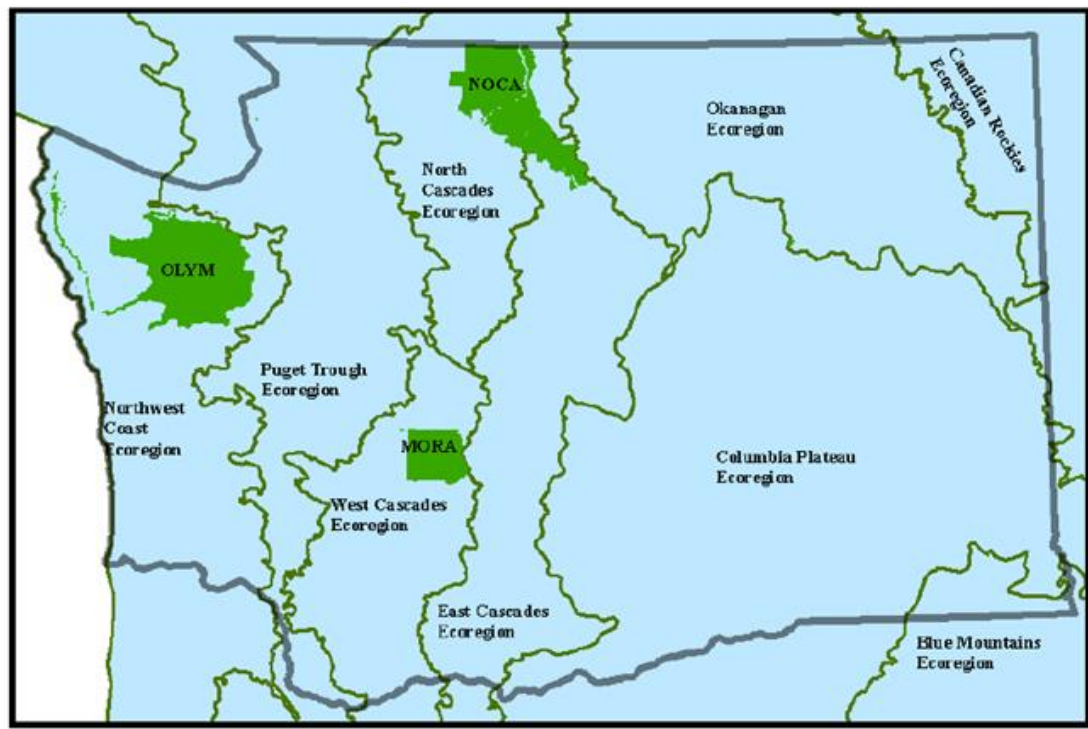
## A Data-Driven Method for Assembling Map Classes from Vegetation Associations in WA State National Parks

Rachel Brunner (Portland State University), Eric Nielsen (Portland State University), Catharine Copass (National Park Service)

### Background

The large national parks of Washington State are large, rugged, and ecologically complicated; exhaustive field mapping of plant communities would be prohibitively expensive. Therefore, our plant community maps rely on modeling to extend field data across remote areas. We model with predictor layers derived (in-house) from satellite data, air photos, lidar, and climate data.

Figure 1. Location of Mount Rainier (MORA), North Cascades (NOCA), and Olympic (OLYM) National Parks with respect to Washington State ecoregions

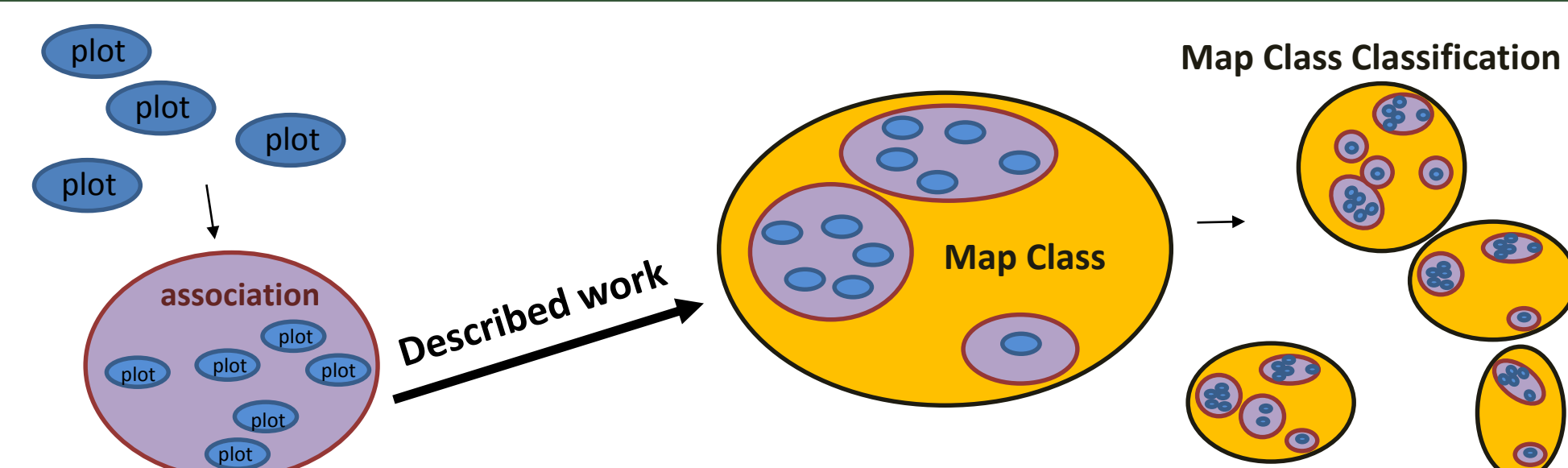


Training data were collected at more than 6000 field plots representing 294 distinct vegetation associations—far too many to map. National Park Service methods suggest mapping at higher levels of the vegetation hierarchy, such as alliances or even groups (Natureserve 2010). Based on this approach, we initially lumped associations into 40 alliance and group level map classes.

Unfortunately, this hierarchical lumping did not produce well distinguished map classes. On average, the map class an association was assigned to received just 0.5% more model votes than its second best fit. Floristics were also not well separated: associations were only 3.6% more similar to their assigned class than to their second best fit.

Accurate modeled maps require map classes to be distinguishable on the ground *and* in predictor data. We worked to group associations in a way that maximized the separability of map classes in mapping while maintain field identification (floristic separation) and structure from the United States National Vegetation Classification (USNVC).

### Methods

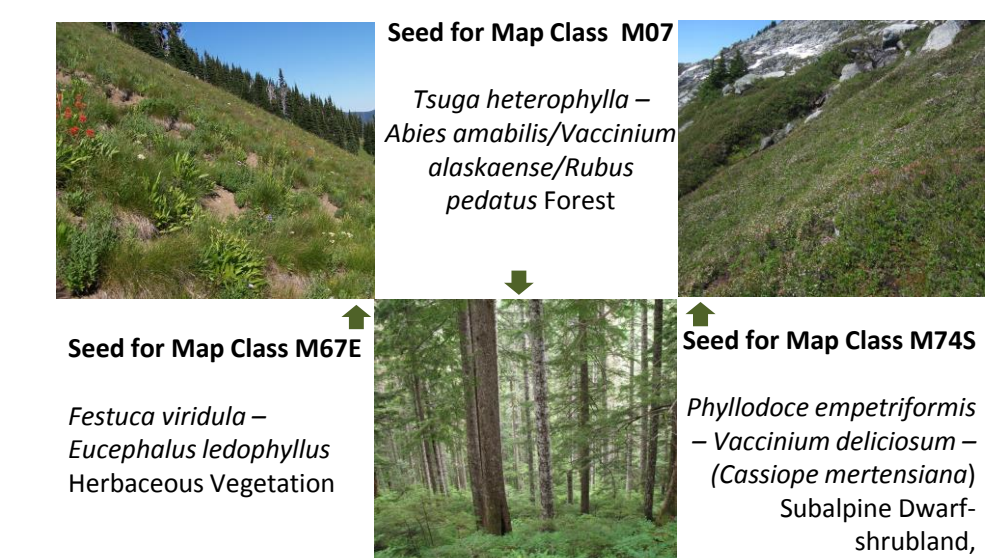


We used a data-driven approach to refine the map classification. Confusion between map classes is realized on the ground, so we based our analysis of confusion on the characteristics of our 6000+ training plots rather than simple association summaries. We optimized the classification to minimize confusion in the field (floristic similarity between map classes) and confusion in the modeling process (which depresses mapability and thus map accuracy).

We define **modeling confusion** as the percent difference in 'model votes' for random forests models between each pair of associations. We define **floristic similarity** as the Euclidian distance between each plot and the floristic centroid of each association.

For both floristic similarity and modeling confusion, we aimed to **maximize** the plot-level margin of victory (model votes or distance in the assigned class compared to next best fit). Summarized plot-level margin of victory scores were used to assess the consequences of each assignment on classification-wide floristic similarity and model confusion.

Our classification is based on user-expected plant community concepts (often alliances). We seeded map classes with these distinct associations, then grew each class in a stepwise fashion to create the classification. After each assignment, the geometric mean of plot-level margin of victories was re-calculated across all associations and map classes.



Analyses used R (3.3.2, 'vegclust' package), MS Excel, and ArcGIS (10.3)

### Map Classification

### Results

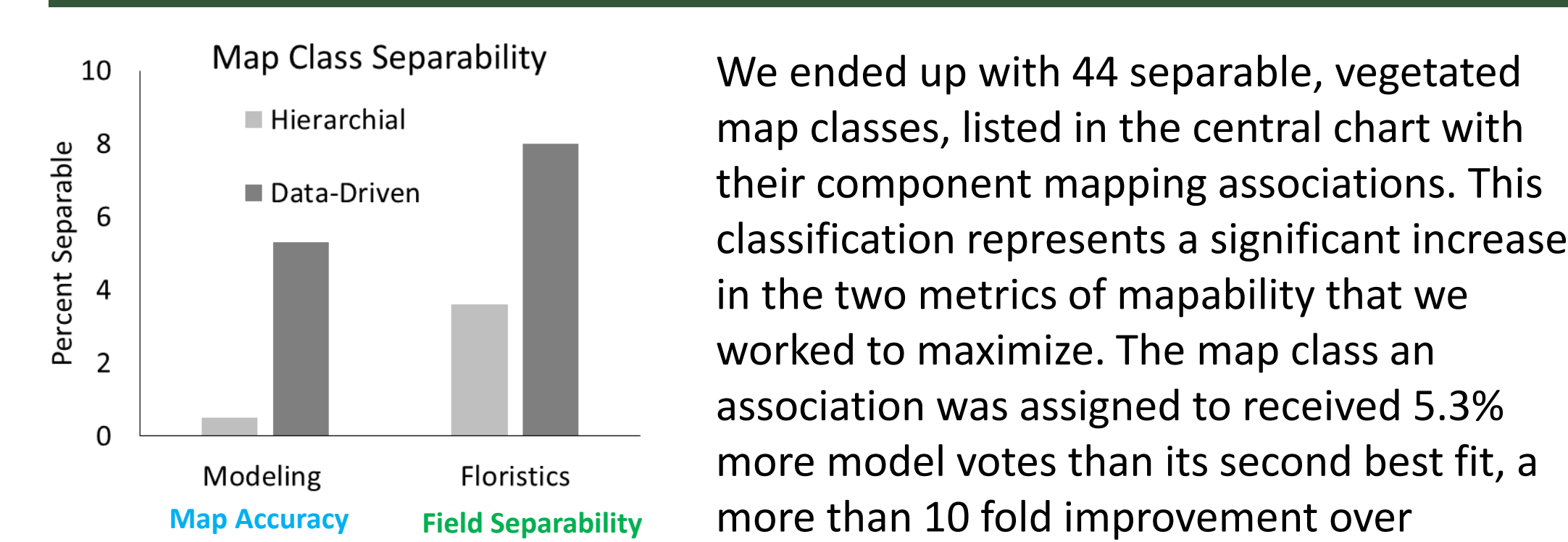


Figure 2. Comparison of map class separability by classification method. Our percent separability metrics are the geometric mean of the margin of victory across all associations and map classes.

We ended up with 44 separable, vegetated map classes, listed in the central chart with their component mapping associations. This classification represents a significant increase in the two metrics of mapability that we worked to maximize. The map class an association was assigned to received 5.3% more model votes than its second best fit, a more than 10 fold improvement over modeling based on NVC hierarchy. Floristic separability also improved markedly: floristics were 8.0% more similar to their assigned class than to their second best fit (up from 3.6%).

Our revised classification also highlights and maintains the key ecological communities that local ecologists and NPS staff expect in these maps. To reflect these expert opinions and expectations, our final classification is a compromise between a pure data-driven agglomeration process and previous methods based on manual assignment based on user expectations and local ecology alone.

We retained parts of the hierarchical classification that were important to managers, in some cases despite an alternative that increased both measures of separability. For example, we retained the alpine heather map class (M74S) as distinct from the alpine sparsely vegetated map class (M63) because of requests from NPS staff.

These three large national parks cover four ecoregions (Figure 1) and span steep elevation (0 – 14,411 feet) and climate gradients (66 – 660 cm annual rainfall). Despite these differences, the parks have similar montane and subalpine forests and National Park Service Inventory and Monitoring efforts are organized at the network scale (North Coast and Cascades Network). We created a consistent map class classification, where all map classes are defined similarly at all parks. Two thirds of the map classes are common enough to be mapped at more than two parks, while one third are unique to just one park (e.g. hypermaritime map classes at Olympic National Park). If a map class was not represented by enough plots to effectively map at a particular park, it was lumped with the next most similar type. This lumping is not expected to have a significant effect on the final maps as it affects under 1% of field plots.

### Conclusions

We created a classification that maximizes separability without sacrificing user experience. Our data-driven approach maintained well-supported NVC hierarchical groupings, increased achievable map accuracy and increased consistency of map class field identification. Our extensive field dataset allowed us to ask the data which associations were most confused. To meet user needs, we seeded map classes with the robust categories (endmembers) that were important to map users, and manually maintained a consistent crosswalk between parks, except where data were insufficient.

Our work showed the benefit of specifically testing and maximizing separability. We suggest that these techniques can help USNVC Hierarchy-based vegetation mapping efforts using modeling or photo-interpretation meet desired accuracy targets by specifically prioritizing separability in the mapping process. Our compromise was toward map classes which do not adhere strictly to USNVC hierarchy but we believe the gains in accuracy and field utility will be valued by the map users.

### Acknowledgments

This work was generously supported by the National Park Service Inventory and Monitoring Program.  
 Crawford, R. C., C. B. Chappell, C. C. Thompson, and F. J. Rocchio. 2009. Vegetation classification of Mount Rainier, North Cascades, and Olympic National Parks. Natural Resource Technical Report NPS/NCCN/NRTR—2009/211. National Park Service, Fort Collins, Colorado.  
 NatureServe. 2010. NCCN alliance descriptions: forested and a subset of non-forested alliances from Mount Rainier, North Cascades, and Olympic National Parks. Interim Report, NatureServe, Arlington, VA.

The central chart represents the resultant crosswalk between mapping associations and map classes. This crosswalk balances user expectation, floristic similarity and model separability. Final modeling, map class prevalence review and accuracy assessments are not yet completed. Map classes may be combined based on accuracy assessment results and map class names are subject to change.

Example plot photos are shown above entries for some map classes. Each photo represents a different component association. Bolded mapping associations are already part of the USNVC, all others are preliminary or provisional and stem from extensive field work and ongoing classification efforts (Crawford et al. 2009).