Phytocoenologia Vol. 48 (2018), Issue 2, 171–179
Published online December 2017

Special Issue Classification Approaches
Research Paper

Carolina Vegetation Survey: an initiative to improve regional implementation of the U.S. National Vegetation Classification

Robert K. Peet*, Kyle A. Palmquist, Thomas R. Wentworth, Michael P. Schafale, Alan S. Weakley & Michael T. Lee

Abstract

Purpose: The purpose of the Carolina Vegetation Survey (CVS) is to provide a framework for characterization of natural plant communities throughout North and South Carolina and adjacent US states. The resulting classification supports scientific interpretation of vegetation pattern, biodiversity inventory, biodiversity monitoring, conservation efforts, and identification of restoration targets. **Application of the approach**: The CVS classification approach will lead to a synthetic treatment of the vegetation of the Carolinas. Although regional in its scope, the approach is generalizable to other geographic regions. It will support further development of the US National Vegetation Classification (USNVC), providing a model for similar work in other regions, thereby leading to more rapid improvement and application of the USNVC. **Main features and protocols**: Our protocols were developed for use with a large database of vegetation-plot records inventoried using a consistent, published methodology. Plot sizes typically range from 100 to 1000 m², although data from smaller subplots are also collected. Each record has a full list of vascular plant species and includes cover-class estimates and tallies of woody stems. Species concepts and nomenclature are regularly updated to a consistent standard. Supporting data include soil chemical and physical properties and other site attributes. Class definition procedures employ node-based agglomerative hierarchical algorithms, informed by ordination procedures and by *a priori* assignment of records to vegetation classes. **Advantages and limitations**: Classification protocols draw on widely-used, well-established procedures and algorithms. Typological resolution aims to conform to one or more of the lower levels of the USNVC hierarchy. A limitation is that most plots were located using preferential sampling, which has the potential for incorporating selection biases. However, this approach captures rare or unanticipated types that would otherwise be missed. To date CVS data collection has been restricted to natural communities and consequently cannot inform classification of semi-natural or cultural vegetation.

Keywords: association; biodiversity; classification protocol; natural vegetation; North Carolina; South Carolina; United States National Vegetation Classification; vegetation classification.

Nomenclature: Weakley (2015).

Abbreviations: CVS = Carolina Vegetation Survey; FGDC = Federal Geographic Data Committee; FL = Florida; GA = Georgia; NC = North Carolina; SC = South Carolina; TN = Tennessee; USNVC = United States National Vegetation Classification; VA = Virginia; WV = West Virginia.

Submitted: 19 October 2016; first decision: 20 March 2017; accepted: 28 April 2017

Co-ordinating Editor: Miquel De Cáceres

Introduction

In 1988, a group of North Carolina ecologists, representing several universities, government agencies, and non-profit organizations, formed the Carolina Vegetation Survey (CVS). These founding members were inspired by the remarkable diversity of natural communities in North Carolina (NC), South Carolina (SC), and surrounding southeastern US states, and they were also concerned that the region's natural heritage was rapidly eroding under the combined pressures of population growth and economic development. The initial intent of

*Corresponding author's address: University of North Carolina, Chapel Hill, NC, United States; peet@unc.edu. Complete addresses of all authors can be found at the bottom of the paper.

DOI: 10.1127/phyto/2017/0168

www.borntraeger-cramer.de
0340-269X/2017/0168 $ 4.50

the organizers of CVS was to develop and share a deep understanding of the pattern of vegetation across the region's diverse landscapes, which range from isolated barrier islands along the Atlantic Coast to peaks in excess of 2,000 m in the southern Appalachian Mountains. Although CVS was established through the individual initiatives of its founding members, it has at times been supported by government agencies with an interest in particular applications, including the U.S. Forest Service, the U.S. Geological Survey, the NC Department of Mitigation Services (formerly NC Ecosystem Enhancement Program), and the NC Natural Heritage Program.

Vegetation classification has always been a core interest of CVS because of its need for a robust framework for characterization of natural communities throughout the Carolinas and adjacent states. One goal immediately embraced was development of a rigorous, plot-based classification of natural communities for the region. Almost simultaneously with establishment of CVS in 1988, two of the founding members published a draft classification of NC's natural communities to guide inventory of important natural areas for conservation (Schafale & Weakley 1990). This preliminary classification was based on literature and personal experience and has served as a starting point for the CVS classification initiative. It also informed the original development of the US National Vegetation Classification (USNVC), which has since evolved to conform to the EcoVeg approach (Faber-Langendoen et al. 2014) and its global implementation.

As now envisioned, the CVS vegetation classification is intended to be used as a framework for characterizing vegetation, organizing further ecological research, identifying research needs, and guiding conservation of biodiversity at the community and ecosystem levels. In addition, CVS and associated researchers are exploring environmental correlates of vegetation and flora, ecological behavior of individual species, and spatial and temporal patterns of species richness and composition in plant communities (e.g. Peet et al. 2014; Palmquist et al. 2015). The NC Natural Heritage Program, one of the collaborating institutions, uses a state-wide natural community classification (Schafale 2012) to guide biodiversity conservation planning, and CVS contributes to advancement of that effort. Since the inception of CVS, the USNVC has been developed and advanced as the national standard for vegetation classification in the US. Currently, a major goal of CVS is to contribute to and refine the USNVC vegetation types for the southeastern United States using quantitative analysis of vegetation plot data and in the process provide a model for how ecologists in other regions could similarly inform and improve the USNVC. The specific products of CVS have been intended from the beginning to include a comprehensive book or books on the natural vegetation of the Carolinas and a series of journal articles on specific subsets of vegetation and specific ecological topics of interest.

To achieve these goals, CVS has been collecting vegetation-plot records since 1988 following a consistent and detailed protocol (Peet et al. 1998, 2012a). Plot size is flexible, typically in the range of 100 m^2 to 1000 m^2, depending on the nature of the vegetation and the stand characteristics. Each record has a full list of vascular plant species with cover-class estimates plus tallies of woody stems. Species concepts and nomenclature are regularly updated to a consistent standard (currently Weakley 2015). Supporting data include geocoordinates, soil chemical and physical properties, and other site attributes. Data have been collected at annual collaborative events collectively involving over 1,100 volunteers, and by graduate students and ecological professionals as components of their specific research projects. Plots are permanently marked and most are located on public conservation lands. Thus, most plots are available for future resampling, though this is not part of the primary focus of CVS.

The CVS geographic focus is NC and SC with some extension into adjacent states to capture the range of variation of recognized types. The focal vegetation spans the range of natural, non-ruderal terrestrial vegetation, including emergent wetlands, with infrequent extensions into submerged aquatic vegetation and ruderal vegetation. The ecological scope of a classification covering such a large region is necessarily broad. In the southern Appalachian Mountains, for example, CVS has vegetation-plot records that span the range from fertile, protected valley bottoms that support large-statured, mixed-mesophytic forests, to extremely exposed, high-elevation rock outcrops that support sparse herbaceous vegetation. On barrier islands of the Maritime Fringe, vegetation ranges from well-developed maritime forests to sparse herbaceous vegetation of dunes and salt flats. Plots were selected to represent the most natural remaining examples of vegetation across the study area. Ruderal, heavily altered, and exotic-dominated stands were generally avoided, though the level of alteration of vegetation in CVS plots varies depending on the remaining vegetation available.

CVS has endeavored to make available both its protocol and research findings to the scientific community. The sampling protocol is detailed in Peet et al. (1998, 2012a) and has been widely adopted by other researchers. For example, the vegetation monitoring protocol of the NEON program (http://www.neonscience.org/) is in large part modeled after the CVS protocol (Barnett 2014), as is the sampling protocol of the Cumberland Piedmont Network of the US National Park Service (https://irma.nps.gov/DataStore/Reference/Profile/2192468; accessed 17 Oct. 2016). Several ecologists have published vegetation research that illustrates application of the CVS approach to various natural communities and geographic regions (e.g. Newell et al. 1999; Carr et al. 2010; Palmquist et al. 2015). The CVS database also provides a rich resource for individuals interested in addressing broader scientific questions related to plant ecology. For example, there is a growing awareness among vegetation scientists that scale of observation strongly influences perception of ecological pattern and process, and for this reason it is important to inventory vegetation at multiple spatial scales (Shmida & Wilson 1985; Stohlgren et al. 1995; Peet et al. 1998; Dengler 2009). The CVS protocol generates species occurrence data at 6 spatial scales on a logarithmic scale from 0.01 m^2 to 1000 m^2. Moreover, ongoing efforts by CVS have generated what is by far the largest dataset currently available that contains observation of species co-occurrence of all vascular plants over a broad range of spatial scales.

Application of this approach

Most plot data in the CVS database have been collected since 1988 and adhere to the CVS protocol (Peet et al. 1998, 2012a). In the interest of complete and comprehensive coverage we have included additional datasets collected from the Carolinas since 1975 that conform to the USNVC standards (Jennings et al. 2009). In addition, to allow examination and description of communities across their entire geographic range, as mandated by the USNVC standards, we have incorporated plot data collected in Virginia (VA), West Virginia (WV), Tennessee (TN), Georgia (GA), and Florida (FL). In some cases we have resampled plots to document successional change (e.g. Taverna et al. 2005; Israel 2012) or the impact of disturbance events (e.g. Reilly et al. 2005a, 2005b) or management practices (e.g. Palmquist et al. 2014, 2015). As of October 2016, our database contained 19500 plot observations from our target states, including 7317 from NC, 1392 from SC, 4944 from VA, 4302 from WV, 500 from GA, 574 from TN and 471 from FL.

Our primary goal is a comprehensive classification and associated publications that treat all natural vegetation of the Carolinas. We further intend that this classification will contribute to both the USNVC (FGDC 2008; Jennings et al. 2009; Faber-Langendoen et al. this volume) and the NC Natural Heritage Program community classification (Schafale 2012). To date, we and our collaborators have generated multiple publications and theses on subsets of the vegetation of the Carolinas such as: forests (Newell & Peet 1998; Newell et al. 1999), bogs and fens (Wichmann 2009), rock outcrops (Wiser et al. 1996), and river floodplains (Brown & Peet 2003) of the Blue Ridge Mountains; river floodplains (Matthews et al. 2011), upland forests (e.g. Taverna et al. 2005; Israel 2012), and non-alluvial wetlands (Seymour 2011) of the Piedmont region; and river floodplains (Faestel 2012), maritime forests (Wentworth et al. 1992), and longleaf pine (*Pinus palustris*) vegetation (Peet 2006; Palmquist et al. 2016) of the Coastal Plain and Coastal Fringe.

The CVS dataset allows for many applications in addition to classification. For example, these data have resulted in several novel studies in which species richness has been examined across a range of spatial scales (e.g. Brown & Peet 2003; Fridley et al. 2005; Peet et al. 2014, Palmquist et al. 2015), and also several studies that have explored the effect of soil properties on species composition (e.g. Newell & Peet 1998; Newell et al. 1999; Peet et al. 2003, 2014; Palmquist et al. 2015). The dataset has also allowed comparisons with similar datasets from other parts of the world to address a variety of questions, such as the degree of specialization of North American versus European trees (Manthey et al. 2011), or the extent of exchange of exotic species between two regions and assessment of the habitats that are most vulnerable in these regions (Kalusová et al. 2014, 2015).

CVS plot data are maintained with a set of four Microsoft Access databases largely conforming to the VegBank data model (Peet et al. 2012a, 2012b). One Access database (CVS Archive) contains all plot records and tracks changes in species and community determinations. A somewhat simpler and denormalized database (CVS Analysis) is used by most researchers, who connect to it via a third database to view, query, and export data (CVS Viewer). A fourth database (CVS Entry) is used to facilitate data entry and ensure data quality and consistency. To ensure long-term maintenance, all plot data are stored in VegBank, the vegetation plot archive of the Ecological Society of America (Peet et al. 2012b). Finally, the results of our classification efforts are disseminated via the CVS website (http://cvs.bio.unc.edu).

As a demonstration of the CVS approach, we are developing a comprehensive treatment of *Pinus palustris* dominated vegetation of the Coastal Plain from southeastern VA southward, including occurrences in NC, SC, GA, and FL. Intensive plot-based data collection for this treatment began in the late 1980s and was completed in 2015. A preliminary assessment was published by Peet (2006) and a comprehensive treatment of the xeric types constitutes the first publication in the Proceedings of the USNVC (Palmquist et al. 2016), a peer-reviewed platform for additions to or revisions of USNVC types (Faber-Langendoen et al. this volume).

Main features of the classification approach

The primary classification units recognized by the Carolina Vegetation Survey are associations in the sense of the USNVC (Jennings et al. 2009; Faber-Langendoen et al. 2014), which are roughly equivalent to associations as recognized in the Braun-Blanquet approach. Entitation and description of these types from CVS data allows for revision and improved delineation of existing USNVC associations, or specification of new associations. Proposals for revision and refinement of the USNVC can also address alliances, the next level higher in the USNVC hierarchy. In addition, these units inform the ongoing development of the NC Natural Heritage Program classification of natural communities, which largely, though not entirely, maps onto the USNVC.

Although the CVS units fit within the formal USNVC hierarchy, CVS is also in the process of developing an alternative structure that is more intuitive to the regional user community and can be used to organize publications and websites. The current draft of this structure has four levels above the association corresponding first to geographic region, and then, variously, to physiognomy, environmental setting, and sometimes dominant taxa (e.g. 1. Mountains, 2. Montane upland forests, 3. Montane acid mesic forests, and finally 4. Acidic cove forests). An alternative system employed by the NC Natural Heritage Program has two tiers above the association (e.g. Uplands, Montane cove forests). Although a few associations could potentially be placed in more than one of these alternative organizational categories, this is rare in that regional boundaries are typically consistent with significant changes in environment and the available species pool.

The CVS classification process is consistent across all vegetation types, although the details represent an evolving process. The typical sequence, largely consistent with the recommendations of Peet & Roberts (2013) and as applied by Palmquist et al. (2016), is to identify a target set of communities (often a USNVC Group, the level above alliance; see Faber-Langendoen 2014: Table 2), and then collect in a dataset all plots that might belong to this set. The data are then harmonized in terms of format and taxonomy. Numerical analytic techniques are used to develop tentative clusters and interpret them in terms of site variables. Problematic plots are considered as to whether they should be moved between clusters or excluded from the larger set. The analysis is rerun and the results examined in the context of the current USNVC associations and alliances in an effort to minimize changes in established types, but when necessary still allowing establishment of new types. When the types and their relationships to extant USNVC associations and alliances have been finalized, summary tables are generated describing the types. The last step is preparation of a formal proposal for consideration for adoption by the USNVC (e.g. Palmquist et al. 2016).

Classification protocols

Ecological scope and typological resolution

The primary classification focus of CVS is revision and documentation of the USNVC associations and alliances that occur in the Carolinas and surrounding states. Because of the large size and heterogeneity of the data set, current analysis techniques do not perform well when applied simultaneously to the entire data set. Our detailed analyses typically focus on one USNVC Group or a small number of Groups at one time.

Spatial grain

CVS plots consist of from 1 to 10 modules, each 100 m^2 in area, with cover class values reported from 1 to 4 of these modules and for the entire plot, as well as species lists for a range of smaller subplot sizes. Plot data derived from non-CVS sources generally range from 100 to 1000 m^2 and have cover data that apply only to the entire plot. For a particular project we select a range of spatial grains that maximizes the plots available, yet assures some consistency. Typically, numerical classification is performed on data collected from plots ranging in size from 100 to 1000 m^2. Where possible, we summarize composition and diversity at a standard size, such as 100 m^2.

Primary vegetation attributes

We complete entitation based on both abundance and presence-absence data. Most commonly we use CVS cover class codes (1-10; see Peet et al. 1998) as our preferred metric of abundance because this provides a balanced representation of sparse and common species. We then assess the differences between the abundance and presence-absence clustering solutions. We have higher confidence in solutions where there is agreement between these results.

Constraining attributes

USNVC Groups are not defined based on floristic composition, but rather follow the EcoVeg approach and reflect variation in vegetation with respect to geography, physiognomy and environment (Faber-Langendoen 2014). To circumscribe plots to be analyzed for a particular project, we initially select all plots that were assigned to associations within a USNVC Group or Groups. These initial association assignments are based on expert interpretation of the vegetation, environmental setting, and geographic location of each plot (not numerical analysis) and represent temporary, initial assignments. We then add marginal plots (plots that have floristic affinities to the Group or Groups in question, but may have initially been assigned to an association in another Group) to ensure inclusion of all plots potentially relevant to the scope of the project. Because of the nature of USNVC Groups, we typically use physiographic region (Coastal

Fringe, Coastal Plain, Piedmont, Mountains), hydrology, and soil attributes to help define the set of plots to analyze for the focal USNVC Group(s) in question. We then proceed with numerical classification of floristic data to derive associations, which we then characterize in terms of typical geography, hydrology, and physical setting to aid in later assignment of new plots.

Properties of class definition procedures

CVS types are initially defined through numerical clustering, which leads to extensive class definitions (a list of plot records belonging to each class) with the associated plots being reported in the classification publications. The next step is to generate summary statistics and describe a central concept for each association, each of which is in turn integrated into the text-based descriptions of types in the USNVC database, and which is available at http://usnvc.org. In short, the original extensive definitions are used to create descriptions of associations that future users can employ to identify vegetation observed or recorded at other sites.

Associations developed by CVS constitute a hybrid of numerical and expert-based units. Central to the CVS approach is the use of numerical clustering methods to develop potential classification units. However, units derived from numerical analysis are then compared against the extant types in the USNVC in an effort to achieve consistency in degree of homogeneity within types and the degree of differences between types. There is also an effort to preserve as much as possible of the original classification so as to not be disruptive to ongoing applications of the USNVC.

Summary of plot-based definition procedures

Acquisition of plot data. CVS has systematically and on an almost annual basis since 1988 collected high-quality plot data using the CVS protocol (Peet et al. 1998, 2012a). Because these plots are subjectively located to represent the floristic, geographic and environmental range of remaining high-quality natural vegetation, there is inevitably some selection bias in plot location, but this approach assures that we capture far more of the unusual and rare types than would be the case with either random or stratified random sampling. In addition, we supplement our plot data with plot data from projects conducted by other research groups. We pay close attention to these externally collected data to identify possible differences in taxonomic resolution or floristic completeness and exclude plots of questionable consistency with CVS-collected plots.

Preparation of plot data. One of four databases that comprises the CVS database is the data entry tool, which has tables that replicate all data sheets for ease of entry. After data are entered using the entry tool, both transcription error-checking and logical error checking are conducted. After error-checking, data are migrated from the entry tool into the CVS archive database.

To minimize the degree of difference in the resolution of taxonomic names between years and field observers, we standardize these names prior to analysis. As an initial step, taxonomic names need to be standardized to current nomenclature, typically following Weakley's flora (currently Weakley 2015). Observations of unknown taxa, ambiguous taxa, hybrid taxa, non-vascular plant taxa, and family- and higher-level taxa are removed. We create complexes for one or more species or genera that cannot be (or have not been) consistently distinguished from one another (e.g. *Bulbostylis* [*ciliatifolia* + *coarctata*]). When there are observations identified to species within a genus (e.g. *Agalinis aphylla*), but also observations whose highest level of resolution is to genus (*Agalinis* sp.), we usually choose to remove the observations for genus-level taxa, except where a high percentage of occurrences is recorded only at the genus level, in which case all occurrences are treated at the genus level. Similarly, if there are many species-complex identifications relative to species-level identifications (e.g. *Antennaria* [*parlinii* + *plantaginifolia*], n=9; *A. parlinii,* n=1; and *A. plantaginifolia,* n=3), we generally lump the species-level taxa into the multi-species complex, dropping any ambiguous genus-level identifications.

Because some plots provide species' cover values within individual vertical strata and others only for the plot as a whole, we combine species cover values spread across multiple strata into a single plot cover value using the equation recommended by Jennings et al. (2009).

Grouping plot records. An important first step in grouping plots into types is to determine the set of plots to be analyzed that represent a given USNVC Group (or similar subset of the USNVC). We use several approaches to select an overly inclusive starting set of plots, including presence of typical dominant and indicator species and the previous subjective or numeric assignment of individual plots to existing USNVC associations within the Group(s) of interest. After data preparation, initial clustering is performed on this data set, and a combination of statistical indicators and expert judgment is used to remove outliers and peripheral plots. Several iterations of this approach are often used to refine the data set (now a Consistent Classification Section) for further analysis at the association level. This refinement of the data set represents a clarification of the conceptual boundaries of the Group(s) being analyzed.

The second step is to conduct entitation on the set of plots identified in the step above to identify clusters that represent potential USNVC associations. Typically, we calculate a Sørenson dissimilarity matrix from the threecolumn vegetation dataset (plot, species, cover-class code) and then use agglomerative, hierarchical clustering with flexible-group linkage (β = -0.25) (see Peet & Roberts 2013) on both abundance data and presence-absence data. We chose this dual approach because species abundance across plots can be affected by external factors other than environmental conditions (e.g. fire suppression and land-use history). As such, presence-absence may give us a clearer picture of species-environmental relationships. Because we are revising an existing classification hierarchy, we identify the extant number of associations for the focal USNVC Group(s) and set this as our initial cluster number. However, we also run hierarchical clustering with a range of cluster numbers to quantify both finer-scale and broader-scale patterns in the data. Collectively, we seek agreement between the abundance and presence-absence clustering for all results from different cluster numbers. In addition, we use silhouette width and the optpart function (R package optpart, Roberts 2015, Roberts 2016) to assess cluster validity and reassign plots to better-fit clusters (Peet & Roberts 2013). While we, to some degree, are still investigating the optimal methods for analysis and classification at the association level, the protocol described here is the result of substantial testing and refinement, and variants have been used in several publications (e.g. Carr et al. 2010; Matthews et al. 2011; Palmquist et al. 2016) and theses (e.g. Wichmann 2009; Seymour 2011; Faestel 2012).

Evaluation of vegetation types. We next determine the interpretability of the clusters using non-metric multidimensional scaling (NMS; see Peet & Roberts 2013) ordination to visualize the homogeneity of plots within each cluster. We run NMS for 200 iterations with 200 random starts for all plots to explore differences among the alliances and for all associations within each alliance. We also use NMS to explore the environmental and geographic differences between our clusters and to determine where associations fall with respect to environmental and geographic gradients. We highlight plots in the NMS ordination according to cluster identity and overlay environmental and site attributes and species richness at multiple spatial scales to identify which edaphic and geographic factors are related to compositional differences. In addition, we report Pearson's correlation coefficients (r) and their associated R^2 values between the first three NMS ordination axes and all environmental and geographic variables.

Characterization of vegetation types. To describe the floristics of each association and identify compositional differences between types, we generate constancy tables for all clusters, which include average cover % and constancy for each species in each cluster (*sensu* Matthews et al. 2011). In addition, we identify those species that are prevalents (i.e. those N species with the highest constancy, where N is the mean number of species in a standard plot area, typically 100 m^2). We also use the Murdoch Preference Function (R package optpart, function murdoch; Roberts 2016) to identify indicator taxa for each cluster. To summarize the topographic, edaphic, and species richness gradients across associations, we provide boxplots of the environmental attributes described above and species richness values across multiple spatial scales. Finally, we map our newly defined types to extant USNVC associations to indicate whether the association concept is equal to a previous USNVC concept, is approximately equal to a previous type, is greater than but includes a previous type, is less than but is included in a previous type, or whether the association does not overlap a seemingly similar established USNVC concept (for examples see Carr et al. 2010; Matthews et al. 2011; Palmquist et al. 2016).

Advantages and limitations of the approach

One of the significant challenges confronted in improving and revising the USNVC is that we are working with a classification system that has been in use for some years. New standards adopted in 2008 mandating use of plot data and quantitative analyses are being retroactively applied (Faber-Langendoen 2014 this volume). The existing USNVC units vary in their origin, level of clarity and focus, and amount of data, analysis, and experience behind them. Some are very general concepts of probable vegetation, some are the result of local quantitative analysis with uncertain applicability beyond the specific area studied, and some have been well tested by use and experience, while others have not. Nevertheless, the USNVC is in widespread use for the multiple purposes that our work aims to promote (e.g. ecological characterization, inventory, biodiversity conservation). Consequently, we perceive that it is important to improve the USNVC content without unnecessary disruption to the investment already made in using it. What we are doing is analogous to repairing a car as it is speeding down the highway. Although challenging, our classification approach yields units that are both regionally consistent and backed by publicly available plot data, allowing revision of our types if and when additional data become available.

Our focus for analysis has been on individual USNVC Groups, with characterization and entitation of the narrower vegetation units contained therein. As noted above, we need to identify the subset of plots that represent each USNVC Group (Consistent Classification Section sensu De Cáceres et al. 2015) before group entitation. However, that partitioning has proven somewhat challenging for two reasons. First, vegetation is continuous and some plots or clusters of plots do not fit perfectly into a single USNVC Group, but rather span the boundary between, or have characteristics of, two Groups. As such, we sometimes find it necessary to include a set of plots in the analyses for multiple Groups, with final assignment based on relative similarity to these groups, and on our knowledge of the vegetation. Second, the physical environmental properties (e.g. soil moisture or texture) that are the basis of some USNVC Groups may not prove to have as strong an influence on vegetation clustering as does geography. In such a case, the Group has to be defined by assignment of branches of the larger dendrogram in piecemeal fashion. This has occurred in Groups having diverse flora with strong geographic species turnover, and may represent a challenge for other workers in similar situations.

Like most plot classification approaches in the literature, our approach begins with non-quantitative, *a priori* assignment of individual plots to existing associations, followed by unsupervised classification of plot data. The resulting quantitatively defined sets of plots are mapped onto existing USNVC associations, based on their initial, *a priori* assignments. Successful *a priori* assignment requires extensive experience with the USNVC, and, where that was lacking, plots were often given problematic assignments. We have tempered this approach by extensive reexamination of *a priori* assignments using both quantitative measures based on compositional similarity and subjective assessment based on personal experience with the vegetation and the existing descriptions of the associations. A further challenge is that the quantitatively defined units often do not map cleanly onto existing associations. This discordance may indicate a need to correct the boundaries of associations, but it may simply be the result of continuous variation in vegetation. In the latter case, we may temper quantitative results by applying our experience with the vegetation. However, it is possible that an approach that begins with supervised classification, or an approach that starts with more carefully defined concepts of existing associations, would yield greater consistency with the existing associations of the USNVC.

One additional limitation inherent in most plot-based approaches, ours included, is the degree to which the plots are intended to represent natural vegetation. The USNVC mandates that associations be based on existing floristics, but our goals for classification, and for most intended uses of the USNVC, benefit from having units that represent natural, unaltered vegetation of particular ecological settings. We have attempted to sample the most natural vegetation remaining for each type and region, but plots inevitably vary in the degree to which they have been altered as a direct or indirect consequence of post-European human activity (e.g. lumbering, introduction of exotic species, altered herbivore populations, altered disturbance regimes, and climate change). At times, despite our best efforts to sample high-quality sites, effects of human-mediated alteration of the environment and vegetation create a stronger signal in the quantitative analysis than the underlying ecological processes that the classification is meant to reflect. As an example, in *Pinus palustris* dominated vegetation of the Coastal Plain, vegetation structure, especially shrub cover relative to herb cover, varies substantially with fire history. We found plots from different environments and regions grouping together in our quantitative analyses, apparently solely because they had high cover of common, wide-spread shrub species and had reduced cover of ecologically diagnostic species owing to reduced fire frequency. In such cases, it is necessary to apply expert knowledge and judgment to delete plots representing such degraded sites. In addition, vegetation types that are well defined in the USNVC are, nevertheless, occasionally not well characterized by the plot data, owing to alterations in all of the remaining examples. In such cases, some description is better than none, but it is important to be clear to the user what the specific data represent.

Despite the challenges and limitations we describe above, the Carolina Vegetation Survey initiative has resulted in a large, multi-scale vegetation plot dataset and a steadily improving classification of the vegetation of the Carolinas consistent with the USNVC. The dataset will continue to provide opportunities to document and refine the USNVC, while providing a platform for basic and applied science beyond the scope of vegetation classification. Our activities provide a model for how a diverse set of professionals can collaboratively revise or otherwise improve the USNVC at regional to subcontinental scales. Finally, the CVS approach enhances collaboration between a diverse group of stakeholders, including the general public and students, and in the process increases their awareness of environmental issues, threats to biodiversity, and the value of vegetation classification.

Author contributions

All authors contributed to the design of the Carolina Vegetation Survey, collection of plot data, and analysis of those data. R.K.P. & M.T.L. designed and maintained the CVS database and attended to quality control. All authors contributed to the design of this paper, and R.K.P., K.A.P., T.R.W. & M.P.S. contributed portions of the original text. All authors contributed to revision and refinement of the text.

Acknowledgements

We are pleased to thank the 1100+ persons who have participated in the collection of the vegetation plots in the Carolina Vegetation Survey database. Collection of these data and development of the database were supported by the U.S. Forest Service, the U.S.D.A. Natural Resources Conservation Service, the NC Department of Mitigation Services (formerly NC Ecosystem Enhancement Program), the NC Natural Heritage Program, and numerous other groups, programs and persons. Analysis of the data was supported by the above organizations plus the U.S. Geological Survey and the U.S. National Park Service.

References

Barnett, D.T. 2014. *TOS Science design for plant diversity*. National Ecological Observatory Network. NEON.DOC. 000912. http://data.neonscience.org/documents/ 10179/1723439/NEON.DOC.000912vA/.

Brown, R.L. & Peet, R.K. 2003. Diversity and invasibility of southern Appalachian plant communities. *Ecology* 84: 32–39.

Carr, S.C., Robertson, K.M. & Peet, R.K. 2010. A vegetation classification of fire-dependent pinelands of Florida. *Castanea* 75: 153–189.

De Cáceres, M., Chytrý, M., Agrillo, E., Attorre, F., Botta-Dukát, Z., Capelo, J., Czúcz, B., Dengler, J., Ewald, J. (…) & Wiser, S.K. 2015. A comparative framework for broad-scale plot-based vegetation classification. *Applied Vegetation Science* 18: 543–560.

Dengler, J. 2009. A flexible multi-scale approach for standardised recording of plant species richness patterns. *Ecological Indicators* 9: 1169–1178.

Faber-Langendoen, Baldwin, D.K., Peet, R.K., Meidinger, D., Muldavin, E., Keeler-Wolf, T. & Josse, C. 2017. The EcoVeg Approach in the Americas: U.S., Canadian and International Vegetation Classifications. *Phytocoenologia* (this issue).

Faber-Langendoen, D., Keeler-Wolf, T., Meidinger, D., Tart, D., Hoagland, B., Josse, C., Navarro, G., Ponomarenko, S., Saucier, J.-P., Weakley, A. & Comer, P. 2014. EcoVeg: a new approach to vegetation description and classification. *Ecological Monographs* 84: 533–61.

Faestel, M.A. 2012. *Classification and description of alluvial plant communities of the North Carolina Coastal Plain*. M.S. Thesis, University of North Carolina, Chapel Hill, NC.

FGDC (Federal Geographic Data Committee). 2008. *National Vegetation Classification standard, version 2.* FGDC-STD-005-2008. Vegetation Subcommittee, U.S. Geological Survey, Reston, Virginia, USA.

Fridley, J.D., Peet, R.K., Wentworth, T.R. & White, P.S. 2005. Connecting fine- and broad-scale patterns of species diversity: species-area relationships of southeastern U.S. flora. *Ecology* 86: 1172–1177.

Israel, K.A. 2012. *Vegetation change in the Duke Forest, 1977–2010*. M.S. Thesis, University of North Carolina, Chapel Hill, NC.

Jennings, M.D., Faber-Langendoen, D., Loucks, O.L., Peet, R.K. & Roberts, D. 2009. Standards for associations and alliances of the U.S. National Vegetation Classification. *Ecological Monographs* 79: 173–199.

Kalusová, V., Chytrý, M., Peet, R.K. & Wentworth, T.R. 2014. Alien species pool influences the level of habitat invasion in intercontinental exchange of alien plants. *Global Ecology and Biogeography* 23:1366–1375.

Kalusová, V., Chytrý, M., Peet, R.K. & Wentworth, T.R. 2015. Intercontinental comparison of habitat levels of invasion between temperate North America and Europe. *Ecology* 96: 3363–3373.

Manthey M., Fridley, J.D. & Peet, R.K. 2011. Niche expansion from past species extinctions? A comparative assessment of habitat generalists and specialists in the tree floras of Southeastern North America and Southeastern Europe. *Journal of Biogeography* 38: 840–853.

Matthews, E.M., Peet, R.K. & Weakley, A.S. 2011. Classification and description of alluvial plant communities of the Piedmont region, North Carolina, U.S.A. *Applied Vegetation Science* 14: 485–505.

Newell, C.L. & Peet, R.K. 1998. Vegetation of Linville Gorge Wilderness, North Carolina. *Castanea* 63: 275–322.

Newell, C.L., Peet, R.K., Ulrey, C.J., Wentworth, T.R., Patterson, K.D. & McLeod, D.E. 1999. Geographic variation in forest distribution across five landscapes in the Southern Appalachian Mountains of North and South Carolina. In: Eckerlin, R.P. (ed.), *Proceedings of the Appalachian Biogeography Symposium*. Virginia Museum of Natural History Special Publication Number 7, pp. 19–33.

Palmquist, K.A., Peet, R.K. & Carr, S.C. 2016. Xeric Longleaf Pine Vegetation of the Atlantic and East Gulf Coast Coastal Plain: an Evaluation and Revision of Associations within the U.S. National Vegetation Classification. *Proceedings of the U.S. National Vegetation Classification* 1: 1–70. (http://50.87.248.75/~usnvcorg/proceedings/index.php).

Palmquist, K.A., Peet, R.K. & Mitchell, S.R. 2015. Scale-dependent responses of longleaf pine vegetation to fire frequency and environmental context across two decades. *Journal of Ecology* 103: 998–1008.

Palmquist, K.A., Peet, R.K. & Weakley, A.S. 2014. Changes in plant species richness following reduced fire frequency and drought in one of the most species-rich savannas in North America. *Journal of Vegetation Science* 25:1426–1437.

Peet, R.K. Ecological classification of longleaf pine woodlands. 2006. In: Jose, S., Jokela, E. & Miller, D. (eds.), *Longleaf pine ecosystems: ecology, management, and restoration*. Chapter 3, pp. 51–93. Springer, New York, NY.

Peet, R.K., Fridley, J.D. & Gramling, J.M. 2003. Variation in species richness and species pool size across pH gradients in forests of the Southern Blue Ridge Mountains. *Folia Geobotanica* 38: 391–401.

Peet, R.K., Lee, M.T., Boyle, M.F., Wentworth, T.R., Schafale, M.P. & Weakley, A.S. 2012a. Vegetation-plot database of the Carolina Vegetation Survey. *Biodiversity and Ecology* 4: 243–253.

Peet, R.K., Lee, M.T., Jennings, M.D. & Faber-Langendoen, D. 2012b. VegBank – a permanent, open-access archive for vegetation-plot data. *Biodiversity and Ecology* 4: 233–241.

Peet, R.K., Palmquist, K.A. & Tessel, S.M. 2014. Herbaceous layer species richness of southeastern forests and woodlands: Patterns and causes. In: Gilliam, F.S. (ed.), *The herbaceous layer in forests of eastern North America*, Second edition. Chapter 10, pp. 255–276. Oxford University Press, Oxford, UK.

Peet, R.K. & Roberts, D.W. 2013. Classification of natural and semi-natural vegetation. In: Franklin, J. & van der Maarel, E. (eds.), *Vegetation Ecology*, Second edition. Chapter 2, pp. 28–70. Oxford University Press, New York, NY.

Peet, R.K., Wentworth, T.R. & White, P.S. 1998. A flexible, multipurpose method for recording vegetation composition and structure. *Castanea* 63: 262–274.

Reilly, M.J., Wimberly, M.C. & Newell, C.L. 2005a. Wildfire effects on plant species richness at multiple spatial scales in forest communities of the southern Appalachians. *Journal of Ecology* 94: 118–130.

Reilly, M.J., Wimberly, M.C. & Newell, C.L. 2005b. Wildfire effects on β-diversity and species turnover in a forested landscape. *Journal of Vegetation Science* 17: 447–454.

Roberts, D.W. 2015. Vegetation classification by two new iterative reallocation optimization algorithms. *Plant Ecology* 216: 741–758.

Roberts, D.W. 2016. *Optpart: Optimal partitioning of similarity relations*. R package version 2.3-0. https://CRAN.R-project.org/package=optpart.

Schafale, M.P. 2012. *Guide to the natural communities of North Carolina, fourth approximation*. North Carolina Natural Heritage Program, Department of Environment and Natural Resources, Raleigh, NC. 208 pp.

Schafale, M.P. & Weakley, A.S. 1990. *Classification of the natural communities of North Carolina, Third Approximation.* North Carolina Natural Heritage Program, Department of Environment and Natural Resources, Raleigh, NC. 321 pp.

Seymour, S.D. 2011. *Vegetation of non-alluvial wetlands of the southeastern Piedmont*. M.S. Thesis, University of North Carolina, Chapel Hill, NC.

Shmida, A. & Wilson, M.V. 1985. Biological determinants of species diversity. *Journal of Biogeography* 12: 1–20.

Stohlgren, T. J., Falkner, M.B. & Schell, L.D. 1995. A modified-Whittaker nested vegetation sampling method. *Vegetatio* 117: 113–121.

Taverna, K., Peet, R.K. & Phillips, L. 2005. Long-term change in ground-layer vegetation of deciduous forests of the North Carolina Piedmont, USA. *Journal of Ecology* 93: 202–213.

Weakley, A.S. 2015. *Flora of the Southern and Mid-Atlantic States*. University of North Carolina Herbarium, North Carolina Botanical Garden, University of North Carolina, Chapel Hill, NC. 1320 pp.

Wentworth, T.R., Schafale, M.P., Weakley, A.S., Peet, R.K., White, P.S. & Frost, C.C. 1992. A preliminary classification of North Carolina barrier island forests. In: C.A. Cole and K. Turner (eds.), *Barrier Island Ecology of the Mid-Atlantic Coast: A Symposium*, pp. 31–46. U.S. Department of Interior, National Park Service, Southeastern Regional Office, Atlanta, GA.

Wichmann, B.L. 2009. *Vegetation of geographically isolated montane non-alluvial wetlands of the southern Blue Ridge of North Carolina.* M.S. Thesis, North Carolina State University, Raleigh, NC.

Wiser, S.K., Peet, R.K. & White, P.S. 1996. High-elevation rock outcrop vegetation of the southern Appalachian Mountains. *Journal of Vegetation Science* 7: 703–722.

Author addresses

Peet, R. K. (Corresponding author, peet@unc.edu)[1], **Palmquist, K. A.** (kpalmqu1@uwyo.edu)[2], **Wentworth, T. R.** (twentwo@ncsu.edu)[3], **Schafale, M. P.** (michael.schafale@ncdcr.gov)[4], **Weakley, A. S.** (weakley@unc.edu)[5], **Lee, M. T.** (Michael_Lee@NatureServe.org)[1]

[1] Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, United States
[2] Department of Botany, University of Wyoming, Laramie, WY 82071, United States
[3] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695-7612, United States
[4] North Carolina Natural Heritage Program, 1651 Mail Service Center, Raleigh, NC 27699-1651, United States
[5] Herbarium (NCU), North Carolina Botanical Garden, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, United States