Long Database Report

Vegetation-plot database of the Carolina Vegetation Survey

Robert K. Peet, Michael T. Lee, M. Forbes Boyle, Thomas R. Wentworth, Michael P. Schafale & Alan S. Weakley

Abstract: The Carolina Vegetation Survey (CVS) is a multi-institutional research program designed to document the composition of the natural vegetation of the Carolinas and adjacent states. CVS maintains a system of databases (GIVD ID NA-US-006) that contain data from over 8,200 vegetation plots containing records of in excess of 3,500 species and spanning over 600 vegetation types recognized in the US National Vegetation Classification Standard (NVCS). Over 5,300 of these plots were collected using the level-5 CVS protocol, which provides documentation of composition across a range of spatial scales from 0.01 to at least 100 and often 1,000 m². As such, the CVS database contains the largest set of multi-scale vegetation plot data yet assembled. In addition, the standard CVS protocol includes tallies of woody stems by diameter and detailed documentation of soil and other environmental attributes. The Carolina Vegetation Survey Database system provides a flexible data archive built on the VegBank data model and is designed to store a wide range of vegetation-plot data, from restoration plots tracking the success of planted stems to plots with a complete inventory of both tree stems and plant cover values. Concept-based taxonomy is implemented for both community types and plant taxa, reducing the ambiguity that is inherent in using names only, as the meaning of a name can vary with taxonomic authority. The plots in the CVS archive have been used in numerous publications to document compositional variation in vegetation of the southeastern US and to address a range of broader questions such as patterns in species richness, patterns of species specialization versus generalization, species-area relationships, design of targets for ecological restoration, and documentation of long-term trends in vegetation composition.

Keywords: community type; concept-based taxonomy; scale; Southeastern United States; species-area.

Nomenclature: Weakley (2010).

Abbreviations: CVS = Carolina Vegetation Survey; DBH = diameter at breast height (1.37 m); FGDC= U.S. Federal Geographic Data Committee; NPS = U.S. National Park Service; NVCS= U.S. National Vegetation Classification Standard.

Received: 23 August 2011 – Accepted: 22 December 2011 – Co-ordinating Editor: Gabriela Lopez-Gonzalez.

Introduction

The Carolina Vegetation Survey (CVS) is a collaborative, multi-institutional program established in 1988 to document and disseminate information on the composition and status of the natural vegetation of the Carolinas for purposes of biodiversity inventory, monitoring of environmental impacts, and assessment of conservation status. CVS seeks to identify and prioritize natural communities for conservation, inform the development of a regional classification of these communities, provide accurate templates for restoration, and provide a baseline for assessment of change in natural vegetation of North and South Carolina. CVS archives and distributes information for classification, description, and restoration of the natural vegetation communities of the Carolinas.

CVS maintains a system of databases that house data from vegetation plots collected across the southeastern portion of the United States, with a particular focus on North and South Carolina (Fig. 1). The data come from multiple sources. One particularly important source is provided by "pulse" events that draw together numerous volunteers for a week-long intensive sampling activity where we focus on capturing the range of variation in vegetation within one geographic region. Volunteers get to meet and work with other enthusiasts as well as the most knowledgeable experts from across the region and are generally provided accommodations and a souvenir T-shirt (Fig. 2). There have been one or two (and occasionally three) pulse events every year since 1988, and in excess of 900 persons have participated in at least one of these events. In addition to pulse events, which focus on a region, we have had a significant number of graduate student projects focused on a particular vegetation type across its range of variation, often for vegetation types not amenable to volunteers (e.g. delicate mountain wetlands, dangerous mountain cliffs). We also include in the CVS database vegetation plot data collected by partner organizations, such as the U.S. National Park Service, the U.S. Forest Service, and NatureServe. In this paper we summarize the CVS field protocol, the CVS database system, the available CVS vegetation data, and some of the many applications that have been made of these data. It is our hope that other groups might benefit from learning about the CVS approach, through use of our database tools, or use of the data that we have collected.

Carolina Vegetation Survey

Scope: Southeastern states of the United States, with particular emphasis on North Carolina and South Carolina.

Status: completed and continuing   **Period:** 1976-2012

Database manager(s): Robert Peet (peet@unc.edu); Michael Lee (michael.lee@unc.edu)

Owner: University of North Carolina

Web address: http://cvs.bio.unc.edu

Availability: free upon request   **Online upload:** no   **Online search:** no

Database format(s): MS Access   **Export format(s):** [NA]

Publication: [NA]

Plot type(s): normal plots; nested plots; time series   **Plot-size range:** 0.01-1000 m²

Non-overlapping plots: 8,283   **Estimate of existing plots:** [NA]   **Completeness:** [NA]

Total plot observations: 180,732   **Number of sources:** 10   **Valid taxa:** 3,568

Countries: US: 100.0%

Forest: [NA] — **Non-forest:** [NA]

Guilds: all vascular plants: 100%

Environmental data: altitude: 67%; slope aspect: 69%; slope inclination: 71%; surface cover other than plants (open soil, litter, bare rock etc.): 43%; soil pH: 56%; other soil attributes: 56%

Performance measure(s): cover: 100%; measurements like diameter or height of trees: 64%

Geographic localisation: GPS coordinates (precision 25 m or less): 31%; point coordinates less precise than GPS, up to 1 km: 66%; small grid (not coarser than 10 km): 1%; political units or only on a coarser scale (>10 km): 2%

Sampling periods: 1970-1979: 5.0%; 1980-1989: 6.0%; 1990-1999: 40.0%; 2000-2009: 45.0%; 2010-2019: 3.0%

Information as of 2012-07-21; further details and future updates available from http://www.givd.info/ID/NA-US-006

The CVS plot protocol

The original CVS protocol

Vegetation sampling strategy and field methodology are influenced by vegetation type, the purpose and scale of the study, and the available financial resources. We developed the CVS protocol to be sufficiently flexible that it would be broadly applicable and still provide consistent and compatible data across many studies (Peet et al. 1998). We imposed several requirements when developing the protocol. The protocol had to be applicable to most vegetation types and appropriate for diverse applications in terms of sampling intensity and time commitment. The protocol also needed to be scale transgressive and not be tied to a single scale of observation, as we had shown that the correlations between vegetation and environment vary dramatically with scale of observation (Reed et al. 1993). Finally the protocol had to be appropriate for long-term studies, compatible with other methodologies, and easy to learn and use.

The CVS protocol evolved from the plot methodology developed by Whittaker for his field work in the Siskiyou Mountains (Whittaker 1960). The basic Whittaker plot is 20 m × 50 m (1000 m²) and contains a set of 25 1-m² subplots. Two of us had previously used variants of this method extensively (Peet 1981, Peet & Christensen 1980, Wentworth 1981). The subsequent development by Whittaker (Whittaker et al. 1979, Shmida 1984) of a plot method with multiple scales of observation also appealed to us and influenced our design.

The CVS protocol provides for flexibility in size and shape of the vegetation plot, while retaining a requirement for collection of specific core data that assure comparability across all plots. The key to this flexibility is a modular approach to plot layout, wherein all measurements are made in plots comprised of one or more 100 m² modules (100 m² or 1 "are" = 0.01 hectare). The module size and shape were chosen to provide a convenient building block for larger plots and because a body of data already exists for plots of multiples of this size. In effect, the methodology defines most spatial heterogeneity in vegetation at scales below 10 m × 10 m as an expression of within-community pattern. Where vegetation patterns are long and narrow, module shape may be modified, with a shape of 5 m × 20 m being recommended if this will better represent the vegetation than the standard 10 m × 10 m module.

Although a CVS plot may consist of any number of modules, the standard CVS plot contains 10 modules of 10 m × 10 m (each 100 m²) arranged as a 20 m × 50 m (1000 m²) plot (Fig. 3). Within this plot 4 modules are defined as intensive modules. Ten permanent stakes are inserted, 6 along the 50 m center line and 4 at the ends of two 20 m cross lines located at 10 and 30 m on the center line. In each intensive module two sets of nested quadrats are established for recording species presence. These quadrats are 0.01, 0.1 1.0 and 10 m² in area, and are configured so as to be permanently locatable via a corner stake. When only one intensive module is recorded, there are typically four sets of nested quadrats. Species presence is also recorded for the entire module and a cover value is recorded following the 1–10 CVS cover scale (<0.1%, 0.1–1, 1–2, 2–5, 5–10, 10–25, 25–50, 50–75, 75–95, 95–100%). After the intensive modules have been completed, the remaining 6 modules are surveyed for additional species, and then cover is estimated for each species for the whole plot by vertical strata, again using the CVS cover scale. Woody stems are recorded by species and diameter size classes for the individual intensive modules and also for the remainder of the plot. Soils are usually collected for the intensive modules and analyzed for chemical and physical attributes.
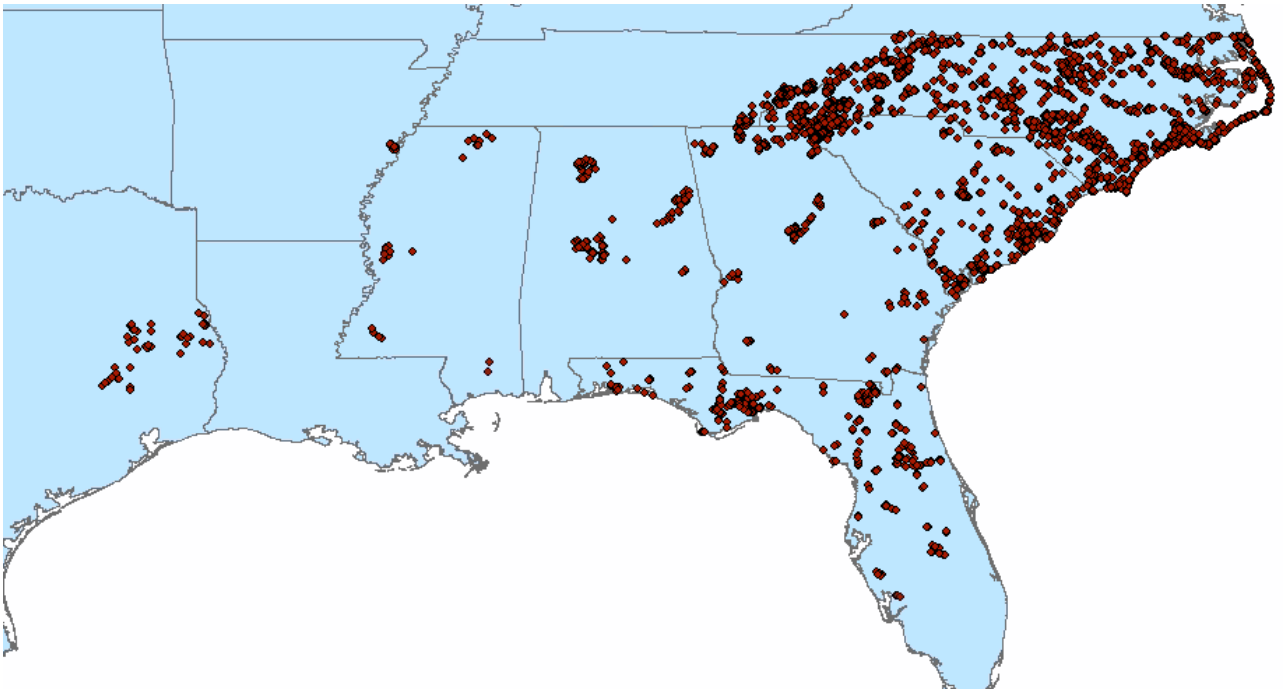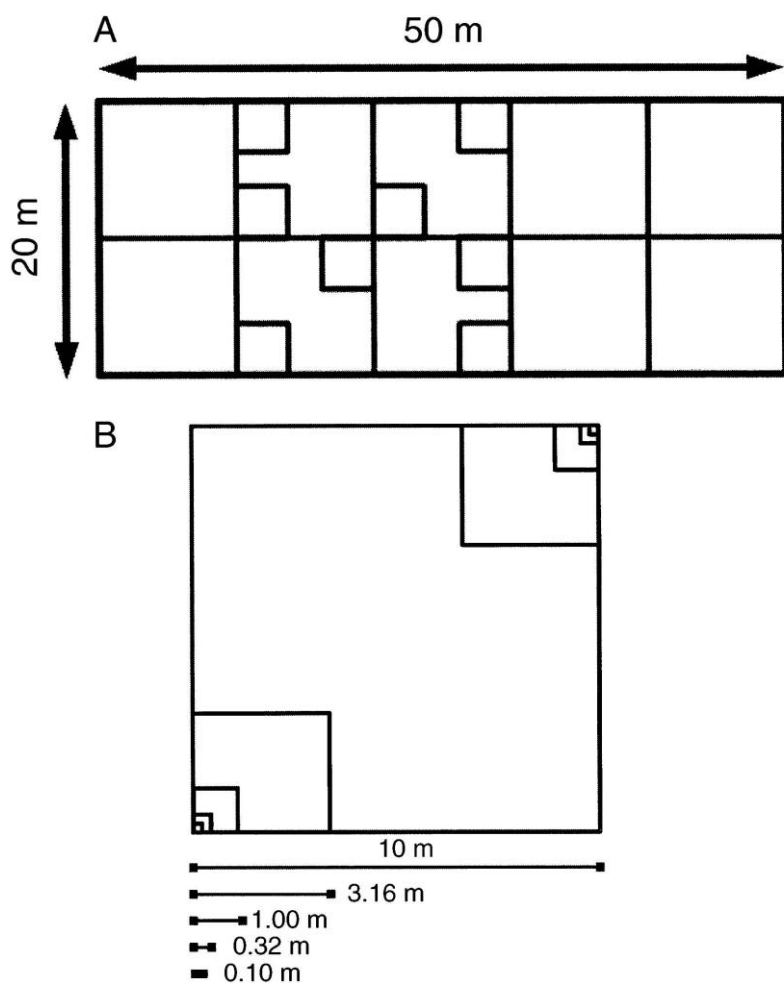
Fig. 1: Map of the 8,466 vegetation plots in the CVS database.

Fig. 2: The Carolina Vegetation Survey works late into the night processing plants collected while sampling vegetation plots. CVS typically holds two week-long plot sampling events each year. Participants are, from the top center counter-clockwise M.F. Boyle, M.P. Schafale, T.R. Wentworth, B.L. Wichmann, D. Welch, S.D. Seymour, M. Faestel, R.K. Peet, J. Brubaker, and J.M. Gramling (Photo: D. Blevins).

Fig. 3: Vegetation plot design of the Carolina Vegetation Survey (Peet et al. 1998). (A) A standard 20 m × 50 m plot composed of 10 modules of 10 m x 10 m, with four intensive modules, each containing two sets of nested subplots. (B) An intensive module, 10 m × 10 m, with nested subplots sized 0.01, 0.1, 1 and 10 m².

Level-3: Community occurrence plots.

Level-3 plots are used to document the overall abundance and vertical distribution of leaf area cover of the more common species in a plot. Cover is estimated for all plant species exceeding a specified lower level (typically 5% cover, although sometimes with a minimum number of species). Species present but with cover lower than the cut-off may be ignored, thereby making these plots quick to record. The information collected meets the Ecological Society of America guidelines (Jennings et al. 2009) and the U.S. Federal Geographic Data Committee (FGDC) standards (2008) for occurrence plots, which are those used to classify vegetation to an association within the U.S. National Vegetation Classification Standard (NVCS). The information can also be used to assess vegetation successional status as well as the presence and abundance of undesirable taxa such as invasive exotics. Optionally, woody stem data required for Level-2 plots (tallies of planted and/or natural woody stems) may be collected for Level-3 plots to allow more accurate assessment of the rate and direction of vegetation development. Level-3 plots are usually only 100 m² in size.

Level-4: Community classification plots.

Level-4 plots include all components of Level-3 plots, but in addition cover values are determined not just for dominant species, but for all vascular plant species occurring on the plot. Additional environmental data are collected, and the plots are often larger in overall area. As is the case for Level-3 plots, it is optional whether to tally woody stems. These plots conform to the requirements for "classification plots" as defined by the NVCS (FGDC 2008, Jennings et al. 2009), which are plots of sufficient detail and quality to be used in development and refinement of types recognized within the NVCS. The primary reasons for collecting Level-4 plots are to facilitate rigorous documentation of vegetation composition and to improve the classification system. An experienced field botanist is required to ensure collection of a complete list of species occurring in the plot. Level-4 plots are 100 m² to 1,000 m² in size, and may or may not be divided into modules recorded separately.

Types of plots supported by the CVS database system

The original CVS database system focused on plots adhering to the particular methodology described above (see Peet et al. 1998). Subsequently it became desirable to be able to store and manage vegetation plots with different attributes, such as those used to document local restoration efforts or less detailed inventory initiatives. CVS provides database support for five types of plots that represent levels of increasing detail, thereby making the CVS database system relevant to a much larger community of potential users. These are arranged in five levels because of the generally increasing complexity across the series.

Level-1: Planted stem inventory plots.

Level-1 plots are applicable only for restoration areas wherein planted woody stems are monitored. The primary purpose is to determine the success of planted material with respect to species, spacing, density, and individual plant dimensions (diameter and height), and to monitor the survival and growth of those installed plants. Level-1 plots are one module (100 m²) in size, although there is no conceptual impediment to allowing larger plots.

Level-2: Total woody stem inventory plots.

Level-2 plots also are designed specifically for restoration areas and represent a superset of information collected for Level-1 plots. In these single-module plots, planted woody stems are recorded exactly as for Level 1, but in addition all woody stems resulting from natural regeneration are recorded by size class. These plots allow an accurate and rapid assessment of the overall trajectory of woody-plant restoration and regeneration.
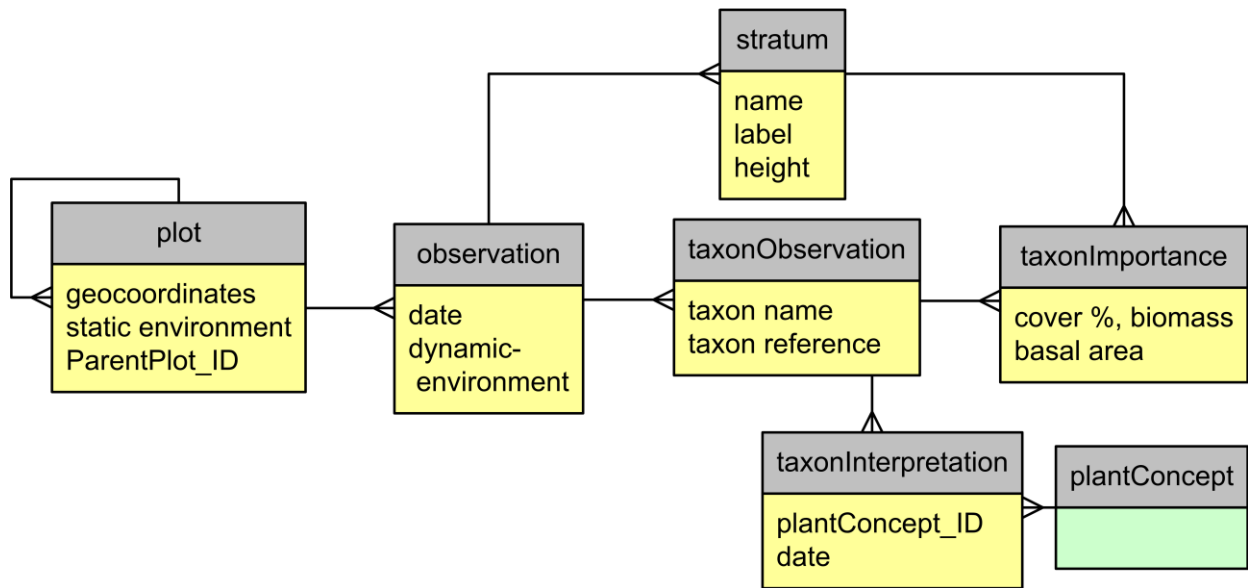
Fig. 4: Data model of the CVS database key tables; the structure follows the VegBank data model. The plot table contains attributes that never change. There can be one to many observations of a plot, and there can be one to many observations of taxa (taxonObservation) in a plot during an observation event, each associated with zero to many measures of taxon importance. Importance values can also be associated with strata (or subplots). Each taxon occurrence can be interpreted as representing a specific plant taxonomic concept zero to many times, one of which can be flagged as the current interpretation. Plant concepts are colored differently to indicate that they function as subdatabase within the full database system.

Level-5: Community classification and structure plots.

Level-5 plots conform to the standard CVS protocol described in the previous section and in Peet et al. (1998). These require all the information collected for Level-4 plots, plus additional information on the spatial structure of the vegetation within the plot, typically including nested quadrats and multiple modules. Woody stem data remain optional, but are strongly recommended. The primary purpose of Level-5 plots is to facilitate rigorous research on and assessment of vegetation composition and structure. Level-5 plots are one or more modules in size, though plots of 10 modules (1,000 m²) are encouraged.

Botanical nomenclature

Botanical nomenclature at present conforms to Weakley's (2010) Flora of the Southern and Mid-Atlantic States, but we expect to update in the future to subsequent versions of Weakley's flora. One advantage of the Weakly flora is that it contains explicit mapping of the taxonomic concepts (*sensu* Berendsohn 1995, Franz et al. 2008) used by authors of over 1,000 taxonomic treatments, thereby facilitating incorporation of legacy data and updating of nomenclature. Where conversion of previous nomenclature to Weakley

(2010) has resulted in unresolvable ambiguities, or where low resolution taxa were recognized, sets of taxa are presented in square brackets. For example, a record of occurrence of an undetermined *Aster* from 1990 would typically be recorded in the database as [*Ampelaster* + *Aster* + *Doellingeria* + *Eurybia* + *Ionactis* + *Oclemena* + *Sericocarpus* + *Symphyotrichum*] as a consequence of the recent division of *Aster* into multiple genera.

Database design and data management

Data model

To capture the wide range of plot complexity represented by the different levels of the CVS protocol and to ensure compatibility with emerging U.S. national standards (FGDC 2008), CVS chose the VegBank data model as the basis of its design (Fig. 4; modified from VegBank 2006, Peet aet al. 2012). VegBank supports multiple cover methods and stratum methods, which is necessary as the CVS database includes data from multiple projects not compliant with the CVS protocol, and the stratum component of the CVS protocol has evolved over the 25 years the protocol has been in use. Finally, the VegBank data model enables CVS

to mark plots as confidential and restrict access to the precise geo-coordinates of those plots.

Taxonomy for both communities and species presents a challenge when integrating data collected at different times, in different places, or by experts using different authoritative taxonomic treatments (see Franz et al. 2008, Franz and Peet 2009, Jansen and Dengler 2009, Peet and Lee et al. in press). In short, a name can refer to multiple definitions of the underlying concept, and a plant or community representing one underlying concept can have multiple names that apply (Berendsohn 1995, Franz et al. 2008). This many-to-many relationship between classification units and names makes accurate data integration impossible unless the underlying concepts for the plant and community taxa are also tracked. Use of the VegBank data model allows CVS to use the concept-based taxonomy features present in the VegBank data model for both plant and community taxa, and in addition provides CVS the ability to support successive determinations of taxon occurrences to particular concepts.

Although the VegBank data model met most needs for CVS, some modifications were necessary. This was primarily due to either CVS data fields that are nonstandard relative to VegBank or the somewhat different purpose of the CVS database. The purpose of VegBank is to

maintain an online repository, whereas the goal of the CVS database includes ongoing management of data. Where non-standard data fields are collected by CVS, we simply added these fields to store the data. Added fields include our raw measurements for McNab's Terrain Shape Index (TSI) (1989) and Landform Index (LFI) (1993), and the particular CVS method of measuring groundcover. The addition of a table was only necessary for storing raw measurements of soil depths, of which there are 4 to 16 for the standard CVS plot, whereas VegBank only allows one value for soil depth per plot.

The most significant change CVS made to the VegBank data model was in the way subplots are stored. VegBank specifies that each subplot be stored as a separate plot record to allow each subplot to have any attributes stored with it that may be stored with a full plot. For example, the subplot in VegBank could have a different latitude and longitude than the full plot. Such detailed information for subplots is not recorded for CVS plots. It seemed unwise to populate separate records for each subplot as changes for a single plot such as a correction of latitude and longitude would need to be propagated to each subplot. Moreover, updates to plant taxonomy for a plot might need to be applied multiple times, depending upon the presence of a species in the full plot and in any number of subplots. Strata in VegBank are modeled as a subset of plot, delineated by vertical height or occasionally by life form (e.g., liana/vine). A species present in a plot may have attributes attached to it that apply to the entire scope of the plot and may also be recorded as present in zero, one, or multiple strata. Subplots, or in the CVS vocabulary "modules," have the same properties as strata in that we wish to record the presence of particular species in the modules along with some importance measures, such as cover. CVS simply uses the VegBank stratum table to document both strata and modules, with strata being differentiated vertically and modules horizontally. Updates to taxonomy are made on a different table and, thus, CVS avoids duplication and makes management of the data simpler.

Data management in a system of databases

Several key databases together comprise the CVS database system. All are implemented in Microsoft Access and all are available from the senior author upon request. The **CVS Central Archive Database**, implemented following the data model described above, contains the majority of the vegetation plots. The archive database is optimized for efficient, normalized storage of vegetation plots. As the archive database is not optimized for data entry, searching, or viewing, CVS created supporting database products for those purposes.

A separate **CVS Data Entry Tool** is needed for several reasons. Most critical, however, is that data entry is a widely distributed task accomplished at various places by various parties, not all of whom can have access to the CVS central database. Moreover, persons engaged in data entry often wish to have very close control over their data until it reaches a certain level of completeness and quality assurance. For these reasons one or more separate copies of the CVS Data Entry Tool can be used by anyone wishing to enter data into the CVS data system. The CVS Data Entry Tool is optimized to make entry very efficient by mimicking the standard CVS field datasheets. Entry of scientific names is facilitated by pick lists that enable selection of any scientific name with 4 to 6 keystrokes. The tool also includes functions for downloading data from previously sampled plots from the archive database, printing datasheets with previous data as a template for monitoring, error-checking, and simple reporting. Entry tool users may download a new version of the species list from our website through the entry tool interface. We do periodically upgrade the species list of the database to the newest version of the Weakley flora. This process is simplified by our ability to use a digitally parsed version of the flora and the presence of taxon concept mapping information within the flora, a feature that makes it almost unique among floristic treatments.

The Data Entry Tool is designed to facilitate quality control. The tool automatically checks for a range of logical errors in the data. Optionally, double-entry of species names may be performed using an interface for this purpose. Comparison of field datasheets with digitally entered data may also be performed line-by-line. We can even automatically embed errors to estimate the accuracy of the data entry and the quality assurance steps. Experienced local botanists scan the projects for species names that seem out of place, and unique or new taxa are double-checked. Names of persons and places are checked against established lists to minimize divergence in form or spelling. Once the extensive quality control steps and error-checking procedures have been completed for an entry tool, its contents are migrated to the archive database.

Some projects move through this process more slowly than others. Many users want to have access to all of the data in the central archive and to simultaneously be able to view and search data in any of the distributed entry tool databases. It would not be efficient to migrate data from the entry tools into the archive database before the people primarily responsible for the data have finished the entry and quality control process, as the archive is not optimized for the quantity of corrections that are likely to be necessary. To solve this problem a third component database was created, the **CVS Analysis Database**. The analysis database contains a combined snapshot of all the databases that contain plots in the CVS data system. The data model for the analysis database is far simpler than that of the central archive database or the data entry tool.

The analysis database is intended to provide a point of access that allows efficient discovery and export of data for specific analytic tasks that the investigator might devise. It is optimized for searching and analysis. An automated script is run to create a new version of the analysis database whenever a significant amount of data entry or updating has been completed, typically from once a week to once a month, depending on how quickly the data are changing. Because we follow the VegBank model, data are never deleted, but only added. A new determination of a taxon or community can be added and marked as current, but the old determination will be retained. Older versions of the database are archived and available if needed.

The searching features and ability to view information in the analysis database, or the central archive or entry tool, is separately housed in the **CVS Viewer Tool**. The viewer's major functions include (1) providing a detailed view of all data associated with a particular plot of interest, (2) creation and execution of searches for plots based on user-specified criteria, (3) creation of a dataset of plots selected by hand or through the results of various searches, (4) export of vegetation plots of interest in any of a number of formats, and (5) automated creation of summary constancy tables.

Plate: Vegetation types featured by the vegetation-plot database GIVD NA-US-006.

A: A CVS plot dominated by *Taxodium disticum* located in the Francis Marion National Forest, South Carolina. Carolina Vegetation Survey Plot 62-07-2926 (Photo: R.K. Peet).

B: CVS sampling along a coastal marsh transect near Wilmington, North Carolina from dominance by *Spartina alternifolia* to dominance by *Spartina patens* and *Juncus roemerianus*. Carolina Vegetation Survey Plots 64-03-0960 and 64-03-0961 (Photo: R.K. Peet).

C: A CVS plot containing pools dominated by the local endemic *Gratiola amphiantha* located in Forty Acre Rock Heritage Preserve, South Carolina. Carolina Vegetation Survey Plot 111-04-1345 (Photo: R.K. Peet).

Database content and availability

As of December 2011 the CVS database contained records from 8,466 plots, most of which are from the Carolinas (Table 1; Figs. 1, 5, Plates A, B). Some 5,339 of these plots conform to the CVS protocol and have nested subsamples (i.e., conform to level-5 standards). Of the remaining plots, 1,161 were collected with a protocol developed for the U.S. National Park Service that is similar and based on the CVS level-4 protocol in that the CVS cover scale is used and plots are usually 20 m × 50 m. A total of 680 vegetation associations in the NVCS are represented by plots in the CVS database. Of the 5,339 CVS plots, 49% are 1,000 m² (n=2,618), 8% are >=500 and <1000 m², 13% are >100 m² and <500 m², 26% are 100 m², and 2% are odd plots that cover less area than 100 m².

Table 1. The numbers of plots in the CVS database by method and with respect to whether cover was recorded by strata and whether tree stems were tallied by DBH.

Method	Plots	With Strata	With Stems
CVS Level 5	5,339	3,378	4,357
NPS protocol	1,161	1,130	0
Cover only	1,133	0	137
Other	833	193	131
TOTAL	**8,466**	**4,701**	**4625**

Summarized content of the CVS database by NVC association type--the lowest scale within the NVCS classification hierarchy--can be viewed on the Carolina Vegetation Survey website (http://cvs.bio.unc.edu). Most raw data in database format can be obtained by application to the senior author of this paper. For a small percentage of the data, access is partially or completely restricted, owing to such issues as occurrence of endangered species or requests from private land owners. In a few cases government agencies have shared data with us, but have requested that they remain confidential. Our long-term plan is to place all public data in both VegBank (http://vegbank.org) and the digital repository of the University of North Carolina Library (https://cdr.lib.unc.edu/).

Applications of the Carolina Vegetation Survey database

Plot records in the CVS database have been used, and are continuing to be used, for many purposes by many investigators. Below we detail some of the published projects that have been based on these data. Others are in progress, including a detailed synthesis of the vegetation of the Carolinas.

Vegetation classification and description

The majority of the datasets in the CVS database system were collected to contribute to our understanding of compositional variation in the vegetation of the southeastern U.S. In most cases the expectation was that the plots would contribute to improving and expanding the NVC, as well as our understanding of variation in vegetation across major environmental and geographic gradients. Because of the diversity of vegetation types found in the Carolinas, and more broadly in the Southeast, descriptive vegetation work is best sorted by physiographic region. Below we recognize three regions: the southern Appalachian Mountains, the ancient rolling hills of the Piedmont to the south and east, and the flat lands of the Southeastern Coastal Plain.

Numerous studies have been conducted that focus on the southern Appalachian Mountain region. Newell prepared a series of descriptive treatments of three National Forest Wilderness Areas, which were subsequently summarized in her dissertation (Newell 1997) and in a pair of publications (Newell and Peet 1998, Newell et al. 1999). One major finding was that, contrary to the widely-known results of Whittaker (1956), the single-most important gradient for understanding variation in southern Appalachian vegetation is soil chemistry, largely driven by parent rock type. Ulrey (2002) built on Newell's work and prepared a region-wide summary of forest vegetation. Montane riparian vegetation was a particular focus of Brown (2002, Brown and Peet 2003) who examined compositional variation and diversity. She found Appalachian riparian vegetation to be extremely species-rich owing to a combination of high propagule pressure, stream deposition of fertile soil, and chronic disturbance from flooding events. At the other end of the topographic spectrum, Wiser et al. (1996) examined compositional variation in high-elevation rock outcrop vegetation and found substantial geographic divergence among occurrences owing to limited dispersal capability and the somewhat stochastic survival of alpine species that had been more widespread during the last full glacial. A study with similar results but in a very different setting was conducted by Wichmann (2009) on isolated montane wetlands, locally referred to as mountain bogs. She, like Wiser, found substantial variation attributable to isolation of these patchy habitats, and stochastic survival of disjunct occurrences of what are today largely northern bog species. Long-term dynamics have also been of interest. Reilly et al. (2006a, b) resampled a series of Newell's plots following wild fire, and Gerschutz (2009) resampled a widely dispersed series of plots following mortality of the dominant canopy of *Tsuga* following attack by the hemlock woolly adelgid (*Adelges tsugae*).

Our initial work on Piedmont forests was conducted by Peet and Christensen (1980) in the Duke Forest (near the Duke University campus) and was based on a variant of the standard Whittaker plot (1960). A subset of these plots were resampled in 2000–01 (Taverna et al. 2005) and again in 2009–10 (Israel 2011) and their work revealed significant changes resulting from increased populations of white-tailed deer (*Odocoileus virginianus*) and the arrival of exotic species. In addition, there has been considerable work on riparian systems to provide a template for restoration activities. Matthews et al. (2011) systematically sampled the riparian vegetation of all major drainage basins in the North Carolina Piedmont and used the data to provide a revised classification of these poorly understood systems. These data were also used to develop a software tool to aid contractors in selection of material to plant during restoration efforts (Matthews 2011). Seymour (2011) systematically sampled the range of Piedmont isolated wetlands including both seepage slopes and isolated depressions.

Research on the Coastal Plain region based on CVS plots has been extensive. Perhaps most comprehensive has been the treatment of the fire-maintained longleaf pine (*Pinus palustris*) woodlands and savannas, well known for their extraordinary small-scale species richness and associated frequency of local endemics (Walker and Peet 1983, Sorrie and Weakley 2001, Peet, Palmquist and Tessel in press). Plots have been accumulated across the eastern two-thirds of the range

of this vegetation type from Virginia south to south-central Florida and west to Alabama with the composition summarized in Peet (2006) and Carr et al. (2010). Carr et al. (2009) quantified and compared different metrics of regional turnover and beta diversity across a broad range of longleaf pine habitats. Almost as heavily studied are the wetlands of the Coastal Plain. Vegetation of riparian systems has been documented by Townsend (2001) and Faestel (2012). Wentworth et al. (1993) examined maritime forest vegetation, and Nifong (1998) examined the great diversity of communities found in isolated depression wetlands.

Broader applications in vegetation science

The CVS database, in addition to being applied to advance description of composition and geographic pattern of regional vegetation, has been used extensively for work on a number of broader conceptual issues. Of particular interest have been questions related to patterns in species diversity. Peet et al. (2003) used data from throughout the Blue Ridge Mountain region to test Pärtel's (2002) thesis that variation in species pool size along a pH gradient should reflect variation in the extent of area with soils of various pH levels. They failed to find support for Pärtel's hypothesis, and instead showed that the species pool was largest for the rather restricted high-pH sites. Subsequently, both Wheeler (2011) and Peet, Palmquist and Tessel (in press) have examined more broadly species richness patterns in the full CVS dataset. They describe very different patterns and processes operating at different scales and in different geographic regions. Wheeler found that richness can be related to soil fertility in a predictive way if latitude, the cover of herbaceous species and the filtering effect of woody shrubs are incorporated. At a broader scale, Costanza et al. (2011) related landscape heterogeneity as derived from satellite images with the diversity of CVS plots. Both Gramling (2006) and Wheeler (2011) incorporated estimates of species pool size into their analyses of species richness of CVS plots and found species pool to be an important predictor of local richness.

The availability of data from multiple spatial scales allows novel uses of the CVS plots. Fridley et al. (2005) were able to use these data to produce the most comprehensive evaluation of species-area curves to date and found that, contrary to prevailing theory, the Arrhenius (power law) model gave consistently a better fit than the Gleason (exponential) model. In another study Fridley et al. (2006) used CVS and other data to show that if one corrects for the space limitation to the number of individuals that fit in a small plot by using time-series data, then the species-area relationship is much more consistent in shape with that found at larger scales.

Numerous other studies have employed the CVS database. Several investigators (Fridley et al. 2007, McChesney 2009, Manthey et al. 2011) have used the database to develop a new approach to quantification of species generalism versus specialization and then have used the approach to compare across species, habitats and geographic regions. Reilly (2010) used the database to develop a new quantitative framework for comparing different forms of species rarity. Ott (2010) compared the degree of divergence in composition of various types of vegetation characterized by isolated occurrences to see if composition was more consistent when corrections were made for species phylogenetic relatedness.

Future directions

The CVS database offers many exciting opportunities for future research. The importance of spatial scale is certain to be a major focus as the CVS database is currently the largest database available that includes multiple observations across a consistent and broad range of spatial scales (each plot spanning at least 4 orders of magnitude of area with at least 5 scales of observation). The careful documentation of plot locations and their long-term field marking allows exact relocation of the plots and all subplots, a benefit being used in a number of on-going studies of vegetation change (e.g. Palmquist et al. 2010). Assessment of long-term change may ultimately prove to be the greatest value of the dataset when many decades from now investigators wish documentation of past community composition. Finally, the dataset is being used to develop a regional synthesis and to provide summary information on all significant vegetation types found in and near the Carolinas.

Acknowledgements

References

Berendsohn, W.G. (1995): The concept of "potential taxa" in databases. – Taxon **44**: 207–212 CrossRef

Brown, R.L. (2002): Biodiversity and exotic species invasion in southern Appalachian riparian plant communities. – PhD thesis, University of North Carolina, Chapel Hill, NC.

Brown, R.L., Peet, R.K. (2003): Diversity and invisibility of southern Appalachian plant communities. – Ecology **84**: 32–39. CrossRef

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

Carr, S.C., Robertson, K.M., Platt, W.J., Peet, R.K. (2009): A model of geographic, environmental and regional variation in vegetation composition of pyrogenic pinelands of Florida. – Journal of Biogeography **36**: 1600–1612. CrossRef

Costanza, J.K., Moody, A., Peet, R.K. (2011): Multi-scale environmental heterogeneity as a predictor of plant species richness. – Landscape Ecology **26**: 851–864. CrossRef

Faestel, M. (2012): Classification and description of alluvial plant communities of the North Carolina Coastal Plain. – MSc thesis, Un at Chapel Hill, Chapel Hill, NC.

Franz, N.M., Peet, R.K. (2009): Towards a language for mapping relationships among taxonomic concepts. – Systematics and Biodiversity **7**: 5–20. CrossRef

Franz, N.M., Peet, R.K., Weakley, A.S. (2008): On the use of taxonomic concepts in support of biodiversity research and taxonomy. – In: Wheeler, Q. D. [Ed.]: The New Taxonomy. Systematics Association Special Volume **74**: 63–86. Boca Raton, FL: Taylor & Francis.

Fridley, J.D., Peet, R.K., van der Maarel, E., Willems, J.H. (2006): Integration of local and regional species-area relationships from space-time species accumulation. – The American Naturalist **168**: 133–143. CrossRef

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. CrossRef

Fridley, J.D., Vandermast, D.B., Kuppinger, D.M., Manthey, M., Peet, R.K. (2007): Co-occurrence-based assessment of habitat generalists and specialists: a new approach for the measurement of niche width. – Journal of Ecology **95**: 707–722. CrossRef

Gerschutz, A.D. (2009): Changes in old-growth and second-growth eastern hemlock (Tsuga canadensis) communities in the southern Appalachians after the introduction of the hemlock woolly adelgid, Adelges tsugae . – MSc thesis, University of North Carolina, Chapel Hill, NC.

Gramling, J.M. (2006): Understanding local and regional plant diversity: Species pools, species saturation, and the multi-scalar effects of plant productivity. – PhD thesis, University of North Carolina, Chapel Hill, NC.

Israel, K. (2011): Vegetation change in the Duke Forest, 1977–2010. – MSc thesis, University of North Carolina, Chapel Hill, NC.

Jansen, F., Dengler, J. (2010): Plant names in vegetation databases - a neglected source of bias. – Journal of Vegetation Science **21**: 1179–1186. CrossRef

Jennings, M.D., Faber-Langendoen, D., Loucks, O.L., Peet, R.K., Roberts, D. (2009): Characterizing Associations and Alliances of the U.S. National Vegetation Classification. – Ecological Monographs **79**: 173–199. CrossRef

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. CrossRef

Matthews, E.R. (2011): Piedmont alluvial vegetation: classification, geographic variation, and restoration. – PhD thesis, University of North Carolina, Chapel Hill, NC.

Matthews, E.R., 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. CrossRef

McChesney, J.L. (2009): Realizing the niche's breadth: inferring ecological process with species generalism. – MSc thesis, University of North Carolina, Chapel Hill, NC.

McNab, W.H. (1989): Terrain Shape Index: quantifying effect of minor landforms on tree height. – Forest Science **35**: 91–104.

McNab, W.H. (1993): A topographic index to quantify the effect of mesoscale landform on site productivity. – Canadian Journal of Forest Research **23**: 1100–1107. CrossRef

Newell, C.L. (1997): Local and regional variation in the vegetation of the South-ern Appalachian Mountains. – PhD thesis, University of North Carolina, Chapel Hill, NC.

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.]: Biogeography of the Southern Appalachian Mountains. Proceedings of the Appalachian Biogeography Symposium, Special Publicatons of the Virginia Museum of Natural History **7**: 19–34.

Nifong, T.D. (1998): An ecosystematic analysis of Carolina bays in the coastal plain of the Carolinas. – PhD thesis, University of North Carolina, Chapel Hill, NC.

Ott, J. (2010): Individualistic and phylogenetic perspectives on plant community patterns. – PhD thesis, University of North Carolina, Chapel Hill, NC.

Palmquist, K., Peet, R.K., Weakley, A.S., Christensen, N.L. (2010): Quantifying species turnover and compositional shifts in the longleaf pine ecosystem after 18 years. – Abstract. 95th Annual meeting, Ecological Society of America. URL: http://www.esa.org/pittsburgh/program.php.

Pärtel, M. (2002): Local plant diversity patterns and evolutionary history at the regional scale. – Ecology **83**: 2361–2366.

Peet, R.K. (1981): Forest vegetation of the Colorado Front Range: composition and dynamics. – Vegetatio **45**: 3–75.

Peet, R.K. (2006): Ecological classification of longleaf pine woodlands. – In: Jose, S., Jokela, E., Miller, D. [Eds.]: Longleaf pine ecosystems: ecology, management, and restoration: 51–94. New York: Springer.

Peet, R.K., Christensen, N.L. (1980): Hardwood forest vegetation of the North Carolina piedmont. – Veröffentlichungen des Geobotanischen Instituts der Eidgenössischen Technischen Hochschule, Stiftung Rübel, in Zürich **69**: 14–39.

Peet, R.K., Fridley, J.D., Gramling, J.M. (2003): Variation in species richness and species pool size across a pH gradient in forests of the southern Blue Ridge Mountains. – Folia Geobotanica **38**: 391–401. CrossRef

Peet, R.K., Lee, M.T., Jennings, M.D., Faber-Langendoen, D. (2012): VegBank – VegBank – a permanent, open-access archive for vegetation-plot data. – In: Dengler, J., Oldeland, J., Jansen, F., Chytrý, M., Ewald, J., Finckh, M., Glöckler, F., Lopez-Gonzalez, G., Peet, R.K., Schaminée, J.H.J. (2012) [Eds.]: Vegetation databases for the 21st century. – Biodiversity & Ecology **4**: 233–241.

Hamburg: Biocentre Klein Flottbek and Botanical Garden. CrossRef

Peet, R.K., Palmquist, K.A, Tessel, S.M. (in press): Herbaceous layer species richness of southeastern forests and woodlands: Patterns and causes. – In: Gilliam, F.S., Roberts, M.R. [Eds.]: The herbaceous layer in forests of eastern North America. 2nd ed. Oxford, UK: Oxford University Press.

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

Reed, R.A., Peet, R.K., Palmer, M.W., White, P.S. (1993): Scale dependence of vegetation-environment correlations: a case study of a North Carolina piedmont woodland. – Journal of Vegetation Science **4**: 329–340. CrossRef

Reilly, L.A.J. (2010): A quantitative approach for defining rarity. – PhD thesis, University of North Carolina, Chapel Hill, NC.

Reilly, M.J., Wimberly, M.C., Newell, C.L. (2006a): Wildfire effects on plant species richness at multiple spatial scales in forest communities of the southern Appalachians. – Journal of Ecology **94**: 118–130. CrossRef

Reilly, M.J., Wimberly, M.C., Newell, C.L. (2006b): Wildfire effects on *β*-diversity and species turnover in a forested landscape. – Journal of Vegetation Science **17**: 447–454. CrossRef

Seymour, S.D. (2011): Vegetation of non-alluvial wetlands of the southeastern Piedmont. – MSc thesis, University of North Carolina, Chapel Hill, NC.

Shmida, A. (1984): Whittaker Plant Diversity Sampling Method. – Israel Journal of Botany **33**: 41–46.

Sorrie, B.A., Weakley, A.S. (2001): Coastal Plain vascular plant endemics: phytogeographic patterns. – Castanea **66**: 50–82.

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. CrossRef

Townsend, P.A. (2001): Relationships between vegetation pattern and hydroperiod on the Roanoke River floodplain, North Carolina. Plant Ecology **156**: 43–58. CrossRef

Ulrey, C.J. (2002): The relationship between soil fertility and the forests of the southern Appalachian region. – PhD thesis, Department of Botany, North Carolina State University, Raleigh, NC. USA.

U.S. F.G.D.C. (2008): National Vegetation Classification Standard, Version 2. – U.S.Federal Geographic Data Committee. FGDC-STD-005-2008 (Version 2). http://www.fgdc.gov/standards/projects/FGDC-standards-projects/vegetation/NVCS_V2_FINAL_2008-02.pdf.

VegBank. (2006): The VegBank Data Dictionary. – Santa Barbara, California: The National Center for Ecological Analysis and Synthesis. URL: http://vegbank.org/ [accessed 2006-05-16].

Walker, J., Peet, R.K. (1983): Composition and species diversity of pine-wiregrass savannas of the Green Swamp, North Carolina. – Vegetatio **55**: 163–179. CrossRef

Weakley, A.S. (2010): Flora of the Southern and Mid-Atlantic States. – Chapel Hill, NC: UNC Herbarium, North Carolina Botanical Garden, University of North Carolina. URL: http://herbarium.unc.edu/ flora.htm [accessed 2010-03-08].

Wentworth, T.R. (1981): Vegetation on limestone and granite in the Mule Mountains, Arizona. – Ecology **62**: 469–482. CrossRef

Wentworth, T.R., Schafale, M.P., Weakley, A.S., Peet, R.K., White, P.S., Frost, C.C. (1993): A Preliminary classification of North Carolina barrier island forests. – In: Cole, C.A., Turner, F.K. [Eds.]: Barrier island ecology of the mid-Atlantic coast: A symposium: 31–46. U.S. National Park Service. Technical Report NPS/SERCAHA/NRTR-93/04. Atlanta, GA.

Wheeler, B.E. (2011): Species diversity of vegetation in the Carolinas: the influence and interaction of scale of observation, soil nutrients, and disturbance events. – PhD thesis, University of North Carolina, Chapel Hill, NC.

Whittaker, R.H. (1956): Vegetation of the Great Smoky Mountains. – Ecological Monographs **26**: 1–80. CrossRef

Whittaker, R.H. (1960): Vegetation of the Siskiyou Mountains, Oregon and California. – Ecological Monographs **30**: 279–338. CrossRef

Whittaker, R.H., Niering, W.A., Crisp, M.D. (1979): Structure, pattern, and diversity of a mallee community in New-South-Wales. – Vegetatio **39**: 65–76. CrossRef

Wichmann, B.L. (2009): Vegetation of geographically isolated montane non-alluvial wetlands of the southern Blue Ridge of North Carolina. – MSc thesis, Department of Plant Biology, 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. CrossRef

Robert K. Peet* (peet@unc.edu) &
Michael T. Lee (michael.lee@unc.edu)
Department of Biology, University of North Carolina
Chapel Hill, NC 27599-3280, UNITED STATES

M. Forbes Boyle
(maxwell_boyle@fws.gov)
U.S. Fish and Wildlife Service,
Okefenokee National Wildlife Refuge,
2700 Suwannee Canal Road
Folkston, GA 31537-7906, UNITED STATES

Thomas R. Wentworth
(tom_wentworth@ncsu.edu)
Department of Plant Biology, North Carolina State University
Box 7612
Raleigh, NC 27695-7612, UNITED STATES

Michael P. Schafale
(michael.schafale@ncdenr.gov)
North Carolina Natural Heritage Program
Green Square DENR Building
1601 Mail Service Center
Raleigh, NC 27699-1601, UNITED STATES

Alan S. Weakley (weakley@unc.edu)
University of North Carolina Herbarium
North Carolina Botanical Garden,
University of North Carolina
CB#3280
Chapel Hill, NC 27599-3280, UNITED STATES

*Corresponding author