

1 *Phytocoenologia* 2017, in press.

2

3 *Type: Research paper*

4

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

7

8 *Author names: Robert K. Peet, Kyle A. Palmquist, Thomas R. Wentworth, Michael P. Schafale,*
9 **Alan S. Weakley, Michael T. Lee**

10

11 *Author addresses: Peet, Robert K. (Corresponding author, peet@unc.edu)¹, Palmquist, Kyle A.*
12 *(kpalmqu1@uwyo.edu)², Wentworth, Thomas R. (twentwo@ncsu.edu)³, Schafale, Michael P.*
13 *(michael.schafale@ncdcr.gov)⁴, Weakley, Alan S. (weakley@unc.edu)⁵, Lee, Michael T.*
14 *(michael.lee@unc.edu)¹*

15

16 ¹ Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-
17 3280, United States

18 ² Department of Botany, University of Wyoming, Laramie, WY 82071, United States

19 ³ Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC
20 27695-7612, United States

21 ⁴ North Carolina Natural Heritage Program, 1651 Mail Service Center, Raleigh, NC 27699-
22 1651, United States

23 ⁵ Herbarium (NCU), North Carolina Botanical Garden, University of North Carolina at Chapel
24 Hill, Chapel Hill, NC 27599-3280, United States

25

26

27 *Running head: Carolina Vegetation Survey*

28

29 **Abstract**

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

54

55 **Keywords:** Association; biodiversity; classification protocol; natural vegetation; North
56 Carolina; South Carolina; United States National Vegetation Classification; vegetation
57 classification

58

59 **Nomenclature:** Botanical nomenclature follows Weakley (2015).

60

61 **Abbreviations:** CVS = Carolina Vegetation Survey; FGDC = Federal Geographic Data

62 Committee; FL = Florida; GA = Georgia; NC = North Carolina; SC = South Carolina; TN =

63 Tennessee; USNVC = United States National Vegetation Classification; VA = Virginia; WV = West

64 Virginia

65 **The Carolina Vegetation Survey approach**

66 In 1988, a group of North Carolina ecologists, representing several universities, government
67 agencies, and non-profit organizations, formed the Carolina Vegetation Survey (CVS). These
68 founding members were inspired by the remarkable diversity of natural communities in North
69 Carolina (NC), South Carolina (SC), and surrounding southeastern US states, and they were also
70 concerned that the region's natural heritage was rapidly eroding under the combined pressures
71 of population growth and economic development. The initial intent of the organizers of CVS
72 was to develop and share a deep understanding of the pattern of vegetation across the region's
73 diverse landscapes, which range from isolated barrier islands along the Atlantic Coast to peaks
74 in excess of 2,000 m in the southern Appalachian Mountains. Although CVS was established
75 through the individual initiatives of its founding members, it has at times been supported by
76 government agencies with an interest in particular applications, including the U.S. Forest
77 Service, the U.S. Geological Survey, the NC Department of Mitigation Services (formerly NC
78 Ecosystem Enhancement Program), and the NC Natural Heritage Program.

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

90 As now envisioned, the CVS vegetation classification is intended to be used as a
91 framework for characterizing vegetation, organizing further ecological research, identifying
92 research needs, and guiding conservation of biodiversity at the community and ecosystem
93 levels. In addition, CVS and associated researchers are exploring environmental correlates of
94 vegetation and flora, ecological behavior of individual species, and spatial and temporal

95 patterns of species richness and composition in plant communities (e.g. Peet et al. 2014;
96 Palmquist et al. 2015). The NC Natural Heritage Program, one of the collaborating institutions,
97 uses a state-wide natural community classification (Schafale 2012) to guide biodiversity
98 conservation planning, and CVS contributes to advancement of that effort. Since the inception
99 of CVS, the USNVC has been developed and advanced as the national standard for vegetation
100 classification in the US. Currently, a major goal of CVS is to contribute to and refine the USNVC
101 vegetation types for the southeastern United States using quantitative analysis of vegetation
102 plot data and in the process provide a model for how ecologists in other regions could similarly
103 inform and improve the USNVC. The specific products of CVS have been intended from the
104 beginning to include a comprehensive book or books on the natural vegetation of the Carolinas
105 and a series of journal articles on specific subsets of vegetation and specific ecological topics of
106 interest.

107 To achieve these goals, CVS has been collecting vegetation-plot records since 1988
108 following a consistent and detailed protocol (Peet et al. 1998, 2012a). Plot size is flexible,
109 typically in the range of 100 to 1000m², depending on the nature of the vegetation and the
110 stand characteristics. Each record has a full list of vascular plant species with cover-class
111 estimates plus tallies of woody stems. Species concepts and nomenclature are regularly
112 updated to a consistent standard (currently Weakley 2015). Supporting data include
113 geocoordinates, soil chemical and physical properties, and other site attributes. Data have been
114 collected at annual collaborative events collectively involving over 1,100 volunteers, and by
115 graduate students and ecological professionals as components of their specific research
116 projects. Plots are permanently marked and most are located on public conservation lands.
117 Thus, most plots are available for future resampling, though this is not part of the primary focus
118 of CVS.

119 The CVS geographic focus is NC and SC with some extension into adjacent states to
120 capture the range of variation of recognized types. The focal vegetation spans the range of
121 natural, non-ruderal terrestrial vegetation, including emergent wetlands, with infrequent
122 extensions into submerged aquatic vegetation and ruderal vegetation. The ecological scope of a
123 classification covering such a large region is necessarily broad. In the southern Appalachian
124 Mountains, for example, CVS has vegetation-plot records that span the range from fertile,

125 protected valley bottoms that support large-statured, mixed-mesophytic forests, to extremely
126 exposed, high-elevation rock outcrops that support sparse herbaceous vegetation. On barrier
127 islands of the Maritime Fringe, vegetation ranges from well-developed maritime forests to
128 sparse herbaceous vegetation of dunes and salt flats. Plots were selected to represent the most
129 natural remaining examples of vegetation across the study area. Ruderal, heavily altered, and
130 exotic-dominated stands were generally avoided, though the level of alteration of vegetation in
131 CVS plots varies depending on the remaining vegetation available.

132 CVS has endeavored to make available both its protocol and research findings to the
133 scientific community. The sampling protocol is detailed in Peet et al. (1998, 2012a) and has
134 been widely adopted by other researchers. For example, the vegetation monitoring protocol of
135 the NEON program (<http://www.neonscience.org/>) is in large part modeled after the CVS
136 protocol (Barnett 2014), as is the sampling protocol of the Cumberland Piedmont Network of
137 the US National Park Service (<https://irma.nps.gov/DataStore/Reference/Profile/2192468>;
138 accessed 17 Oct 2016). Several ecologists have published vegetation research that illustrates
139 application of the CVS approach to various natural communities and geographic regions (e.g.
140 Newell et al. 1999; Carr et al. 2010; Palmquist et al. 2015). The CVS database also provides a
141 rich resource for individuals interested in addressing broader scientific questions related to
142 plant ecology. For example, there is a growing awareness among vegetation scientists that scale
143 of observation strongly influences perception of ecological pattern and process, and for this
144 reason it is important to inventory vegetation at multiple spatial scales (Shmida & Wilson 1985;
145 Stohlgren et al. 1995; Peet et al. 1998; Dengler 2009). The CVS protocol generates species
146 occurrence data at 6 spatial scales on a logarithmic scale from 0.01 to 1000 m². Moreover,
147 ongoing efforts by CVS have generated what is by far the largest dataset currently available that
148 contains observation of species co-occurrence of all vascular plants over a broad range of
149 spatial scales.

150

151 **Application of this approach**

152 Most plot data in the CVS database have been collected since 1988 and adhere to the CVS
153 protocol (Peet et al. 1998; Peet et al. 2012a). In the interest of complete and comprehensive
154 coverage we have included additional datasets collected from the Carolinas since 1975 that

155 conform to the USNVC standards (Jennings et al. 2009). In addition, to allow examination and
156 description of communities across their entire geographic range, as mandated by the USNVC
157 standards, we have incorporated plot data collected in Virginia (VA), West Virginia (WV),
158 Tennessee (TN), Georgia (GA), and Florida (FL). In some cases we have resampled plots to
159 document successional change (e.g. Taverna et al. 2005; Israel 2012) or the impact of
160 disturbance events (e.g. Reilly et al. 2005a, 2005b) or management practices (e.g. Palmquist et
161 al. 2014, 2015). As of October 2016, our database contained 19500 plot observations from our
162 target states, including 7317 from NC, 1392 from SC, 4944 from VA, 4302 from WV, 500 from
163 GA, 574 from TN and 471 from FL.

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

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

185 these regions (Kalusová et al. 2014, 2015).

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

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

202

203 **Main features of the classification approach**

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

213 Although the CVS units fit within the formal USNVC hierarchy, CVS is also in the process
214 of developing an alternative structure that is more intuitive to the regional user community and

215 can be used to organize publications and websites. The current draft of this structure has four
216 levels above the Association corresponding first to geographic region, and then, variously, to
217 physiognomy, environmental setting, and sometimes dominant taxa (e.g. 1. Mountains, 2.
218 Montane upland forests, 3. Montane acid mesic forests, and finally 4. Acidic cove forests). An
219 alternative system employed by the NC Natural Heritage Program has two tiers above the
220 Association (e.g. Uplands, Montane cove forests). Although a few Associations could potentially
221 be placed in more than one of these alternative organizational categories, this is rare in that
222 regional boundaries are typically consistent with significant changes in environment and the
223 available species pool.

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

238

239 **Classification protocols**

240 **Ecological scope & typological resolution**

241 The primary classification focus of CVS is revision and documentation of the USNVC
242 Associations and Alliances that occur in the Carolinas and surrounding states. Because of the
243 large size and heterogeneity of the data set, current analysis techniques do not perform well
244 when applied simultaneously to the entire data set. Our detailed analyses typically focus on one

245 USNVC Group or a small number of Groups at one time.

246

247 **Spatial grain**

248 CVS plots consist of from 1 to 10 modules, each 100 m² in area, with cover class values reported

249 from 1 to 4 of these modules and for the entire plot, as well as species lists for a range of

250 smaller subplot sizes. Plot data derived from non-CVS sources generally range from 100 to 1000

251 m² and have cover data that apply only to the entire plot. For a particular project we select a

252 range of spatial grains that maximizes the plots available, yet assures some consistency.

253 Typically, numerical classification is performed on data collected from plots ranging in size from

254 100 to 1000 m². Where possible, we summarize composition and diversity at a standard size,

255 such as 100 m².

256

257 **Primary vegetation attributes**

258 We complete entitation based on both abundance and presence-absence data. Most commonly

259 we use CVS cover class codes (1-10; see Peet et al. 1998) as our preferred metric of abundance

260 because this provides a balanced representation of sparse and common species. We then

261 assess the differences between the abundance and presence-absence clustering solutions. We

262 have higher confidence in solutions where there is agreement between these results.

263

264 **Constraining attributes**

265 USNVC Groups are not defined based on floristic composition, but rather follow the EcoVeg

266 approach and reflect variation in vegetation with respect to geography, physiognomy and

267 environment (Faber-Langendoen 2014). To circumscribe plots to be analyzed for a particular

268 project, we initially select all plots that were assigned to Associations within a USNVC Group or

269 Groups. These initial Association assignments are based on expert interpretation of the

270 vegetation, environmental setting, and geographic location of each plot (not numerical analysis)

271 and represent temporary, initial assignments. We then add marginal plots (plots that have

272 floristic affinities to the Group or Groups in question, but may have initially been assigned to an

273 Association in another Group) to ensure inclusion of all plots potentially relevant to the scope

274 of the project. Because of the nature of USNVC Groups, we typically use physiographic region

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

280

281 **Properties of class definition procedures**

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

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

297

298 **Summary of plot-based definition procedures**

299 **Acquisition of plot data.** CVS has systematically and on an almost annual basis since 1988
300 collected high-quality plot data using the CVS protocol (Peet et al. 1998, 2012a). Because these
301 plots are subjectively located to represent the floristic, geographic and environmental range of
302 remaining high-quality natural vegetation, there is inevitably some selection bias in plot
303 location, but this approach assures that we capture far more of the unusual and rare types than
304 would be the case with either random or stratified random sampling. In addition, we

305 supplement our plot data with plot data from projects conducted by other research groups. We
306 pay close attention to these externally collected data to identify possible differences in
307 taxonomic resolution or floristic completeness and exclude plots of questionable consistency
308 with CVS-collected plots.

309

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

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

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

333

334 **Grouping plot records.** An important first step in grouping plots into types is to determine the

335 set of plots to be analyzed that represent a given USNVC Group (or similar subset of the
336 USNVC). We use several approaches to select an overly inclusive starting set of plots, including
337 presence of typical dominant and indicator species and the previous subjective or numeric
338 assignment of individual plots to existing USNVC Associations within the Group(s) of interest.
339 After data preparation, initial clustering is performed on this data set, and a combination of
340 statistical indicators and expert judgment is used to remove outliers and peripheral plots.
341 Several iterations of this approach are often used to refine the data set (now a Consistent
342 Classification Section) for further analysis at the Association level. This refinement of the data
343 set represents a clarification of the conceptual boundaries of the Group(s) being analyzed.

344 The second step is to conduct entitation on the set of plots identified in the step above
345 to identify clusters that represent potential USNVC Associations. Typically, we calculate a
346 Sørensen dissimilarity matrix from the three-column vegetation dataset (plot, species, cover-
347 class code) and then use agglomerative, hierarchical clustering with flexible-group linkage ($\beta =$
348 0.25) (see Peet & Roberts 2013) on both abundance data and presence-absence data. We chose
349 this dual approach because species abundance across plots can be affected by external factors
350 other than environmental conditions (e.g. fire suppression and land-use history). As such,
351 presence-absence may give us a clearer picture of species-environmental relationships.
352 Because we are revising an existing classification hierarchy, we identify the extant number of
353 Associations for the focal USNVC Group(s) and set this as our initial cluster number. However,
354 we also run hierarchical clustering with a range of cluster numbers to quantify both finer-scale
355 and broader-scale patterns in the data. Collectively, we seek agreement between the
356 abundance and presence-absence clustering for all results from different cluster numbers. In
357 addition, we use silhouette width and the optpart function (R package optpart, Roberts 2015,
358 Roberts 2016) to assess cluster validity and reassign plots to better-fit clusters (Peet & Roberts
359 2013). While we, to some degree, are still investigating the optimal methods for analysis and
360 classification at the Association level, the protocol described here is the result of substantial
361 testing and refinement, and variants have been used in several publications (e.g. Carr et al.
362 2010; Matthews et al. 2011; Palmquist et al. 2016) and theses (e.g. Wichmann 2009; Seymour
363 2011; Faestel 2012).

364

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

376
377 **Characterization of vegetation types.** To describe the floristics of each Association and identify
378 compositional differences between types, we generate constancy tables for all clusters, which
379 include average cover % and constancy for each species in each cluster (*sensu* Matthews et al.
380 2011). In addition, we identify those species that are prevalents (i.e. those N species with the
381 highest constancy, where N is the mean number of species in a standard plot area, typically 100
382 m^2). We also use the Murdoch Preference Function (R package `optpart`, function `murdoch`;
383 Roberts 2016) to identify indicator taxa for each cluster. To summarize the topographic,
384 edaphic, and species richness gradients across Associations, we provide boxplots of the
385 environmental attributes described above and species richness values across multiple spatial
386 scales. Finally, we map our newly defined types to extant USNVC Associations to indicate
387 whether the Association concept is equal to a previous USNVC concept, is approximately equal
388 to a previous type, is greater than but includes a previous type, is less than but is included in a
389 previous type, or whether the Association does not overlap a seemingly similar established
390 USNVC concept (for examples see Carr et al. 2010; Matthews et al. 2011; Palmquist et al. 2016).

391

392 **Advantages and limitations of the approach**

393 One of the significant challenges confronted in improving and revising the USNVC is that we are
394 working with a classification system that has been in use for some years. New standards

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

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

423 Like most plot classification approaches in the literature, our approach begins with non-
424 quantitative, *a priori* assignment of individual plots to existing Associations, followed by

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

439 One additional limitation inherent in most plot-based approaches, ours included, is the
440 degree to which the plots are intended to represent natural vegetation. The USNVC mandates
441 that Associations be based on existing floristics, but our goals for classification, and for most
442 intended uses of the USNVC, benefit from having units that represent natural, unaltered
443 vegetation of particular ecological settings. We have attempted to sample the most natural
444 vegetation remaining for each type and region, but plots inevitably vary in the degree to which
445 they have been altered as a direct or indirect consequence of post-European human activity
446 (e.g. lumbering, introduction of exotic species, altered herbivore populations, altered
447 disturbance regimes, and climate change). At times, despite our best efforts to sample high-
448 quality sites, effects of human-mediated alteration of the environment and vegetation create a
449 stronger signal in the quantitative analysis than the underlying ecological processes that the
450 classification is meant to reflect. As an example, in *Pinus palustris* dominated vegetation of the
451 Coastal Plain, vegetation structure, especially shrub cover relative to herb cover, varies
452 substantially with fire history. We found plots from different environments and regions
453 grouping together in our quantitative analyses, apparently solely because they had high cover
454 of common, wide-spread shrub species and had reduced cover of ecologically diagnostic species

455 owing to reduced fire frequency. In such cases, it is necessary to apply expert knowledge and
456 judgment to delete plots representing such degraded sites. In addition, vegetation types that
457 are well defined in the USNVC are, nevertheless, occasionally not well characterized by the plot
458 data, owing to alterations in all of the remaining examples. In such cases, some description is
459 better than none, but it is important to be clear to the user what the specific data represent.

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

470

471 **Author contributions**

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

477

478 **Acknowledgements**

479 We are pleased to thank the 1100+ persons who have participated in the collection of the
480 vegetation plots in the Carolina Vegetation Survey database. Collection of these data and
481 development of the database were supported by the U.S. Forest Service, the U.S.D.A. Natural
482 Resources Conservation Service, the NC Department of Mitigation Services (formerly NC
483 Ecosystem Enhancement Program), the NC Natural Heritage Program, and numerous other

484 groups, programs and persons. Analysis of the data was supported by the above organizations
 485 plus the U.S. Geological Survey and the U.S. National Park Service.

486

487 **References**

- 488 Barnett, D.T. 2014. TOS Science design for plant diversity. National Ecological Observatory
 489 Network. NEON.DOC.000912.
 490 <http://data.neonscience.org/documents/10179/1723439/NEON.DOC.000912vA/>.
- 491 Brown, R.L. & Peet, R.K. 2003. Diversity and invisibility of southern Appalachian plant
 492 communities. *Ecology* 84: 32-39.
- 493 Carr, S.C., Robertson, K.M. & Peet, R.K. 2010. A vegetation classification of fire-dependent
 494 pinelands of Florida. *Castanea* 75: 153-189.
- 495 De Cáceres, M., Chytrý, M., Agrillo, E., Attorre, F., Botta-Dukát, Z., Capelo, J., Czúcz, B., Dengler,
 496 J., Ewald, J. (...) & Wisser, S.K. 2015. A comparative framework for broad-scale plot-based
 497 vegetation classification. *Applied Vegetation Science* 18: 543-560.
- 498 Dengler, J. 2009. A flexible multi-scale approach for standardised recording of plant species
 499 richness patterns. *Ecological Indicators* 9: 1169-1178.
- 500 Faber-Langendoen, Baldwin, D.K., Peet, R.K., Meidinger, D., Muldavin, E., Keeler-Wolf, T. &
 501 Josse, C. 2017. The EcoVeg Approach in the Americas: U.S., Canadian and International
 502 Vegetation Classifications. *Phytocoenologia* (in press).
- 503 Faber-Langendoen, D., Keeler-Wolf, T., Meidinger, D., Tart, D., Hoagland, B., Josse, C., Navarro,
 504 G., Ponomarenko, S., Saucier, J.-P., Weakley, A. & Comer, P. 2014. EcoVeg: a new approach
 505 to vegetation description and classification. *Ecological Monographs* 84: 533-61.
- 506 Faestel, M.A. 2012. *Classification and description of alluvial plant communities of the North*
 507 *Carolina Coastal Plain*. M.S. Thesis, University of North Carolina, Chapel Hill, NC.
- 508 FGDC (Federal Geographic Data Committee). 2008. *National Vegetation Classification standard,*
 509 *version 2*. FGDC-STD-005-2008. Vegetation Subcommittee, U.S. Geological Survey, Reston,
 510 Virginia, USA.
- 511 Fridley, J.D., Peet, R.K., Wentworth, T.R. & White, P.S. 2005. Connecting fine- and broad-scale
 512 patterns of species diversity: species-area relationships of southeastern U.S. flora. *Ecology*
 513 86: 1172-1177.

- 514 Israel, K.A. 2012. *Vegetation change in the Duke Forest, 1977-2010*. M.S. Thesis, University of
515 North Carolina, Chapel Hill, NC.
- 516 Jennings, M.D., Faber-Langendoen, D., Loucks, O.L., Peet, R.K. & Roberts, D. 2009. Standards for
517 associations and alliances of the U.S. National Vegetation Classification. *Ecological*
518 *Monographs* 79: 173-199.
- 519 Kalusová, V., Chytrý, M., Peet, R.K. & Wentworth, T.R. 2014. Alien species pool influences the
520 level of habitat invasion in intercontinental exchange of alien plants. *Global Ecology and*
521 *Biogeography* 23:1366-1375.
- 522 Kalusová, V., Chytrý, M., Peet, R.K. & Wentworth, T.R. 2015. Intercontinental comparison of
523 habitat levels of invasion between temperate North America and Europe. *Ecology* 96: 3363-
524 3373.
- 525 Manthey M., Fridley, J.D. & Peet, R.K. 2011. Niche expansion from past species extinctions? A
526 comparative assessment of habitat generalists and specialists in the tree floras of
527 Southeastern North America and Southeastern Europe. *Journal of Biogeography* 38: 840-
528 853.
- 529 Matthews, E.M., Peet, R.K. & Weakley, A.S. 2011. Classification and description of alluvial plant
530 communities of the Piedmont region, North Carolina, U.S.A. *Applied Vegetation Science* 14:
531 485-505.
- 532 Newell, C.L. & Peet, R.K. 1998. Vegetation of Linville Gorge Wilderness, North Carolina.
533 *Castanea* 63: 275-322.
- 534 Newell, C.L., Peet, R.K., Ulrey, C.J., Wentworth, T.R., Patterson, K.D. & McLeod, D.E. 1999.
535 Geographic variation in forest distribution across five landscapes in the Southern
536 Appalachian Mountains of North and South Carolina. In: Eckerlin, R.P. (ed.), *Proceedings of*
537 *the Appalachian Biogeography Symposium*. Virginia Museum of Natural History Special
538 Publication Number 7, pp. 19-33.
- 539 Palmquist, K.A., Peet, R.K. & Carr, S.C. 2016. Xeric Longleaf Pine Vegetation of the Atlantic and
540 East Gulf Coast Coastal Plain: an Evaluation and Revision of Associations within the U.S.
541 National Vegetation Classification. *Proceedings of the U.S. National Vegetation*
542 *Classification 1: 1-70*. (<http://50.87.248.75/~usnvcorg/proceedings/index.php>).

- 543 Palmquist, K.A., Peet, R.K. & Mitchell, S.R. 2015. Scale-dependent responses of longleaf pine
544 vegetation to fire frequency and environmental context across two decades. *Journal of*
545 *Ecology* 103:998-1008.
- 546 Palmquist, K.A., Peet, R.K. & Weakley, A.S. 2014. Changes in plant species richness following
547 reduced fire frequency and drought in one of the most species-rich savannas in North
548 America. *Journal of Vegetation Science* 25:1426-1437.
- 549 Peet, R.K. Ecological classification of longleaf pine woodlands. 2006. In: Jose, S., Jokela, E. &
550 Miller, D. (eds.), *Longleaf pine ecosystems: ecology, management, and restoration*. Chapter
551 3, pp. 51-93. Springer, New York, NY.
- 552 Peet, R.K., Fridley, J.D. & Gramling, J.M. 2003. Variation in species richness and species pool size
553 across pH gradients in forests of the Southern Blue Ridge Mountains. *Folia Geobotanica* 38:
554 391-401.
- 555 Peet, R.K., Lee, M.T., Boyle, M.F., Wentworth, T.R., Schafale, M.P. & Weakley, A.S. 2012a.
556 Vegetation-plot database of the Carolina Vegetation Survey. *Biodiversity and Ecology* 4:
557 243-253.
- 558 Peet, R.K., Lee, M.T., Jennings, M.D. & Faber-Langendoen, D. 2012b. VegBank – a permanent,
559 open-access archive for vegetation-plot data. *Biodiversity and Ecology* 4: 233-241.
- 560 Peet, R.K., Palmquist, K.A. & Tessel, S.M. 2014. Herbaceous layer species richness of
561 southeastern forests and woodlands: Patterns and causes. In: Gilliam, F.S. (ed.), *The*
562 *herbaceous layer in forests of eastern North America*, Second edition. Chapter 10, pp. 255-
563 276. Oxford University Press, Oxford, UK.
- 564 Peet, R.K. & Roberts, D.W. 2013. Classification of natural and semi-natural vegetation. In:
565 Franklin, J. & van der Maarel, E. (eds.), *Vegetation Ecology*, Second edition. Chapter 2, pp.
566 28-70. Oxford University Press, New York, NY.
- 567 Peet, R.K., Wentworth, T.R. & White, P.S. 1998. A flexible, multipurpose method for recording
568 vegetation composition and structure. *Castanea* 63: 262-274.
- 569 Reilly, M.J., Wimberly, M.C. & Newell, C.L. 2005a. Wildfire effects on plant species richness at
570 multiple spatial scales in forest communities of the southern Appalachians. *Journal of*
571 *Ecology* 94: 118-130.

- 572 Reilly, M.J., Wimberly, M.C. & Newell, C.L. 2005b. Wildfire effects on β -diversity and species
573 turnover in a forested landscape. *Journal of Vegetation Science* 17: 447-454.
- 574 Roberts, D.W. 2015. Vegetation classification by two new iterative reallocation optimization
575 algorithms. *Plant Ecology* 216: 741-758.
- 576 Roberts, D.W. 2016. Optpart: Optimal partitioning of similarity relations. R package version 2.3-
577 0. <https://CRAN.R-project.org/package=optpart>.
- 578 Schafale, M.P. 2012. *Guide to the natural communities of North Carolina, fourth approximation*.
579 North Carolina Natural Heritage Program, Department of Cultural Resources, Raleigh, NC.
580 208 pp.
- 581 Schafale, M.P. & Weakley, A.S. 1990. *Classification of the natural communities of North*
582 *Carolina, Third Approximation*. North Carolina Natural Heritage Program, Department of
583 Environment and Natural Resources, Raleigh, NC. 321 pp.
- 584 Seymour, S.D. 2011. *Vegetation of non-alluvial wetlands of the southeastern Piedmont*. M.S.
585 Thesis, University of North Carolina, Chapel Hill, NC.
- 586 Shmida, A. & Wilson, M.V. 1985. Biological determinants of species diversity. *Journal of*
587 *Biogeography* 12: 1–20.
- 588 Stohlgren, T. J., Falkner, M.B. & Schell, L.D. 1995. A modified-Whittaker nested vegetation
589 sampling method. *Vegetatio* 117: 113-121.
- 590 Taverna, K., Peet, R.K. & Phillips, L. 2005. Long-term change in ground-layer vegetation of
591 deciduous forests of the North Carolina Piedmont, USA. *Journal of Ecology* 93: 202-213.
- 592 Weakley, A.S. 2015. *Flora of the Southern and Mid-Atlantic States*. University of North Carolina
593 Herbarium, North Carolina Botanical Garden, University of North Carolina, Chapel Hill, NC.
594 1320 pp.
- 595 Wentworth, T.R., Schafale, M.P., Weakley, A.S., Peet, R.K., White, P.S. & Frost, C.C. 1992. A
596 preliminary classification of North Carolina barrier island forests. In: C.A. Cole and K. Turner
597 (eds.), *Barrier Island Ecology of the Mid-Atlantic Coast: A Symposium*, pp. 31–46. U.S. Department
598 of Interior, National Park Service, Southeastern Regional Office, Atlanta, GA.
- 599 Wichmann, B.L. 2009. *Vegetation of geographically isolated montane non-alluvial wetlands of*
600 *the southern Blue Ridge of North Carolina*. M.S. Thesis, North Carolina State University,
601 Raleigh, NC.
- 602 Wisser, S.K., Peet, R.K. & White, P.S. 1996. High-elevation rock outcrop vegetation of the

603 southern Appalachian Mountains. *Journal of Vegetation Science* 7: 703-722.

604