



Updating vegetation classifications: an example with New Zealand's woody vegetation

Susan K. Wiser & Miquel De Cáceres

Keywords

Alliance; Association; Clustering; Community ecology; Fuzzy classification; Noise clustering; National Vegetation Survey (NVS) databank; Vegetation databases

Nomenclature

Ngā Tipu o Aotearoa – New Zealand Plants database as of 24 January 2011 <http://nzflora.landcareresearch.co.nz/>

Received 28 September 2011

Accepted 29 May 2012

Co-ordinating Editor: Sandor Bartha

Wiser, S.K. (corresponding author, wisers@landcareresearch.co.nz): Landcare Research, PO Box 40, Lincoln, 7640, New Zealand

De Cáceres, M. (miquelcaceres@gmail.com): Biodiversity and Landscape Ecology Laboratory, Forest Science Center of Catalonia, Ctra antiga St Llorenç km 2, 25280, Solsona, Catalonia, Spain

Abstract

Questions: How can existing vegetation classifications be updated when new plot data are obtained? Can we use the properties of plots classed as outliers to identify gaps in our understanding of vegetation patterns and so direct future enquiry?

Location: New Zealand.

Methods: We updated a pre-existing classification of New Zealand's forests and shrublands based on a nationally representative data set (1177 plots) by using 12 374 additional plot records from New Zealand's National Vegetation Survey Databank (NVS). We resampled the NVS plot records to remove uneven representation along floristic and geographic gradients. To update the classification at the alliance level, we first cast the original classification into the fuzzy classification framework of Noise Clustering and then discarded original types with low plot numbers and high compositional variation. We then used the plot records that could not be assigned to any original alliance to define new alliances, while retaining the original alliances as fixed elements. We also defined vegetation associations to create a classification at a lower level of abstraction and related it to the classification at the alliance level. Finally, we determined whether known rare types were represented among the new vegetation types and characterized plot records classed as outliers.

Results: After casting the 24 original alliances in the NC framework, we discarded seven. We extended the 17 remaining alliances with 12 new ones and defined 79 associations. All 12 new alliances had extents <120 986 ha, which is smaller than the original alliances, and included rare types that were known to exist but could not be defined using the objectively sampled data set underpinning the original classification. Plot records classed as outliers tended to occur at lower altitudes or in successional shrublands. Further sampling is required to adequately define vegetation types in such situations, although composition may be inherently erratic in successional shrublands.

Conclusions: Our analysis illustrates the application of a fuzzy classification framework at a national scale and provides a model for others wishing to extend and update vegetation classifications. Our approach allows rare community types to be defined and identifies portions of compositional and geographic gradients that are poorly documented.

Introduction

To be most useful, classifications of vegetation should allow for extensions, modifications or refinements as new data become available. In traditional phytosociology, new vegetation types (normally associations or sub-associations) were progressively defined and added to the

syntaxonomical system as new areas were surveyed (Braun-Blanquet 1929, 1964; Ewald 2003). Typically, such updates were made without destroying the vegetation type concepts previously defined. With the advent of numerical classification techniques, vegetation classification has become more formalized and repeatable (Mucina 1997). Most numerical classification methods, however, do not

easily accommodate the provisional nature of vegetation classifications. That is, those methods cannot adequately fulfill the need to preserve some existing types while discarding others and defining new ones. Rebuilding classifications is impractical when previously defined types are already in established use.

De Cáceres et al. (2010) recently provided a framework, based on fuzzy clustering, to deal with the 'dynamic' character of vegetation classifications. These authors stressed the need to consistently use the same methodological framework for different tasks, such as assigning new plot records to existing vegetation types, updating the definition of existing types or creating new ones. Although De Cáceres et al. (2010) illustrated the tasks of their framework using example data sets, they did not fully anticipate the range of challenges that arise when attempting to extend and refine a pre-existing classification.

The first challenge arises because the pre-existing classification may have not been derived using the same analytical framework that is needed for the extension. Given the myriad of quantitative classification methods available and that different methods may be appropriate in different situations (Mucina 1997; McCune & Grace 2002), this scenario is likely to apply to many pre-existing classifications. For example, using a different distance metric and/or clustering algorithm can result in some plots being assigned to a different type. Approaches are needed to evaluate the consequences of such changes and determine whether the original classification is valid under the new analytical framework (e.g. Feoli et al. 2006). A second set of challenges arises when one intends to update the existing classification using a large number of plot records, such as those available in large phytosociological databases (Dengler et al. 2011). Because phytosociological databases comprise many heterogeneous data sets, analytical results may largely reflect the vegetation types and geographic regions that are disproportionally represented by the plot data (Roleček et al. 2007; Lengyel et al. 2011). Accordingly, these new data may need to be resampled to retain an appropriate representation of the range of vegetation composition present without introducing substantial geographic bias. Although the resampling approach of Lengyel et al. (2011) enables the range of compositional variation in the vegetation plot data to be retained, it does not provide a solution to the problem of geographic bias. Finally, incorporating many more plot records than those of the data set that underpinned the existing classification can allow vegetation types to be defined at a higher level of resolution. In phytosociological studies following the Braun-Blanquet method syntaxa of different levels are related in fully nested relationships (e.g. Dengler et al. 2005). While classifications obtained using fuzzy clustering methods

should be related in a fuzzy way, methods to do this are still underdeveloped.

Recently, a numerically-derived classification at the alliance level was produced for New Zealand's woody vegetation based on a nationally representative plot network of 1177 vegetation plots (Wiser et al. 2011). The objective, area-proportional sampling design had the advantage of not being affected by the preconceptions of the nature of the drivers of vegetation composition that may be inherent in environmentally stratified sampling (Cooper et al. 2006) or the biases against sampling anthropogenically-disturbed vegetation often found in preferential sampling (Chytrý & Rafajová 2003; Botta-Dukát et al. 2007). One of the well-known drawbacks of area-proportional sampling, however, is that rare vegetation types may not be sampled with enough plots to allow them to be defined (Pignatti 1980; Økland 2007; Roleček et al. 2007). Indeed, the New Zealand classification provided greater partitioning within dominant forests than previous national-scale classifications and defined successional types that had not previously been recognized, but it failed to distinguish some known rare forest types. Moreover, because the New Zealand classification was based on only 1177 plots, alliances could not be partitioned as finely as some existing regional classifications, i.e. to the association level. Existing plot records from New Zealand's National Vegetation Survey Databank (NVS; Wiser et al. 2001) could be used to fill these two gaps.

In this paper we take plot records from the NVS databank and adopt the classification framework of De Cáceres et al. (2010) to illustrate how an existing vegetation classification can be extended with new plot data. In particular, we address the following two questions: (1) can we define new alliances or associations to represent known New Zealand rare forest types; and (2) can we use the properties (e.g. geographical location, environmental conditions and composition) of plots classed as outliers to identify gaps in our understanding of New Zealand woody vegetation patterns and so direct future enquiry? Additionally, in our classification exercise we demonstrate how the challenges mentioned above can be overcome. In so doing we provide solutions to the problems associated with geographic and compositional bias in the data obtained from large phytosociological databases, casting an existing classification into a new classification framework, and relating a classification at a lower level of abstraction (i.e. association level) to a higher level (i.e. alliance) in a fuzzy way.

Methods

Study area and original classification

Indigenous forests and shrublands cover ca. 23% and 10% of New Zealand's 27 million hectares of land

surface, respectively (Thompson et al. 2004), whereas before humans arrived forests dominated the landscape. The forests are evergreen and dominated by different combinations of *Nothofagus* spp., broad-leaved angiosperms, *Agathis australis* and other conifers (predominantly podocarps) (Cockayne 1928; McKelvey & Nicholls 1957; Wardle 1984). Shrublands occur in sub-alpine areas and in lowland and montane regions that were presumably forested in pre-human times (Wardle 1991).

From 2002 to 2007, 1177 20-m × 20-m permanent vegetation plots were established at intersections of an 8-km × 8-km grid superimposed on the areas mapped as shrubland or indigenous forest by the New Zealand Land Cover Database (LCDB version 1; Thompson et al. 2004). Data were collected under the auspices of the New Zealand Land Use and Carbon Analysis System (LUCAS; Coomes et al. 2002; Allen et al. 2003). On each 20-m × 20-m plot, the abundance of each species present was recorded in seven fixed height tiers using a modified Braun-Blanquet cover–abundance scale (Hurst & Allen 2007). These data were the basis for a quantitative classification exercise that produced 24 alliances, each containing 19–105 plot records, and estimated extents of 144 000–794 000 ha (Wiser et al. 2011). Wiser et al. (2011) progressed earlier classification efforts in New Zealand by defining contemporary vegetation types, especially alliances in disturbed landscapes and those invaded by exotic species, in addition to recognizing already described common forest types.

Data assembly, manipulation and transformation

New Zealand's NVS contains records from approximately 77 000 vegetation survey plots (Wiser et al. 2001; <http://www.givd.info/ID/AU-NZ-001>). We selected a 12 374 plot subset of these data that were appropriate to extend the original classification. Criteria for inclusion were (1) for permanent plots only the most recent measurement was retained; (2) the plot location was recorded to within 100 m; (3) species abundance was recorded with comparable tier heights and cover class scales to the LUCAS data; (4) plots were not related to exclosures or other experimental treatments and; (5) woody species had total cover across the plot of >20% or were present in height tiers >5 m (corresponding to the LCDB v1 mapping criteria).

Taxonomic changes in species names represent a major challenge when combining data sets collected over different time periods and by different people (Chytrý & Rafajová 2003; Franz & Peet 2009; Jansen & Dengler 2010). To address this issue: (1) taxa that had been identified to the subspecies or variety level were aggregated up to the spe-

cies level; (2) homotypic synonyms were identified and the current name was applied based upon Ngā Tipu o Aotearoa, the New Zealand Plants database <http://nzflora.landcareresearch.co.nz/> and; (3) where we could identify names that had been applied to both narrow and broad taxonomic concepts, we associated the broadest concept with the name.

Cover scores within each height tier were converted to the midpoint of the percentage cover range for that cover class, and summed across tiers (e.g. Wiser et al. 2002). This generated an importance value reflecting the volume occupied by each species rather than its projected cover, and corresponded with that used in the original classification of the LUCAS plots. Resemblance between plots was defined using the Chord distance (Orlói 1967).

Noise Clustering fuzzy classification model

Noise Clustering (NC; Dave 1991) is a modification of the well-known fuzzy clustering algorithm Fuzzy C-means (FCM; Bezdek 1981). For the definition of vegetation types, NC represents a conceptual improvement over FCM because it allows plot records that should be considered outliers to be recognized (De Cáceres et al. 2010). NC identifies outliers by considering an additional class, called the 'Noise' class. The effect of considering the Noise class is that it captures objects that lie further than a pre-specified distance δ from all the centroids of the 'true' clusters. The δ parameter should be set to different values depending on the degree of resolution of the classification analysis (i.e. the level of abstraction of vegetation units) and how the distance between plots is defined.

To extend the applicability of the framework presented in De Cáceres et al. (2010), we suggest identifying three different types of plot records using the rules given in Table 1. The proportion of plots in each category will change depending on the values of the two NC parameters, the distance to the Noise class (δ) and the fuzziness coefficient (m). Larger values of m will lead to more plot records falling in the category 'transitional', whereas larger values of δ will decrease the amount of plot records falling in category 'unassigned'. Clearly, one would like to avoid too few plots 'clearly assigned to a type'. However, if m is set too low (i.e. too close to 1) and δ is too large, the advantages of NC over K-means or Fuzzy C-means are lost (De Cáceres et al. 2010). In this study we set the fuzziness coefficient to a rather low value, $m = 1.1$, because we expected many transitional plot records. Following preliminary results, the distance to the Noise class was set to $\delta = 0.83$ for analyses at the alliance level and $\delta = 0.75$ for analyses at the association level.

Table 1. Typology of plot records determined on the basis of their fuzzy membership in the NC model.

Plot record type	Definition
Clearly assigned in a type	Plot records having fuzzy membership ≥ 0.5 in a vegetation type form the 'core definition' of that type; these plots are those used to characterize the type (e.g. to calculate constancy and fidelity values)
Unassigned	Plot records having fuzzy membership ≥ 0.5 in the Noise class indicate vegetation patterns 'not represented' in the current classification scheme (i.e. outliers) and, where sufficient data are available, may be used to define new vegetation types
Transitional	Plot records whose assignments are ambiguous (i.e. with fuzzy membership < 0.5 for all types, including the Noise class) are considered 'transitional' and are excluded from the characterization of vegetation types. These plot records may be useful, however, to map transitions between vegetation types

Casting the original classification into the NC framework

The original classification of the LUCAS plot records had been obtained using the relativized Manhattan distance (Faith et al. 1987; also referred to as Sørensen's dissimilarity index) and β -flexible hierarchical clustering (Wiser et al. 2011). To cast the original alliance classification obtained into the new NC framework we first calculated the cluster centroids in the space of the Chord distance and then used the NC membership rule for alliances ($\delta = 0.83$) to determine how the assignment of the plot records was affected by the clustering model change. After casting in the NC model, some plot records could be considered transitional, others as outliers or even classified into a different type. Such changes could alter some types to the extent of having too few plot records for an adequate definition. A cluster may exhibit a large variance, which, in combination with the low number of plot records assigned to it, indicates that the cluster is defined based on a very sparse region of the multivariate space. We used the following definition of variance for fuzzy clusters:

$$FVar_i = \frac{\sum_{j=1}^n e_{ij}^2 u_{ij}^m}{\sum_{j=1}^n u_{ij}^m} \quad (1)$$

where e_{ij} is the distance from the object j to the centroid of cluster i , u_{ij} is the membership of object j to cluster i and m is the fuzziness coefficient. For those original alliances that, after casting, included fewer than or equal to 20 plot records and whose cluster variance was above 0.6, we concluded that they lacked sufficient plot records to properly represent their part of the vegetation gradient and excluded them from further analyses.

Stratification of the NVS data set

Because data in the NVS databank have been compiled from many independently conceived projects, plots are distributed unevenly across geographic, environmental and compositional gradients and sample different parts of these gradients with different levels of intensity and

completeness. We stratified the NVS plot data set with the aim of producing a data set capturing as wide a range of compositional variation as possible while avoiding geographic biases. To achieve this we extended the approach of Lengyel et al. (2011). We first constructed a set of geographical grids with points located in arrays at 2-, 4-, 6-, 8- and 10-km intervals across New Zealand. For each grid, we then determined the cell within which each plot was located. We defined the number of plot records required from a given cell as a truncated linear function of the β -diversity of that cell (Fig. 1), which we defined as the overall compositional variation of the plot records within the cell (Legendre et al. 2005). Truncation was used to ensure a minimum number of required plot records per cell. For those grid cells where the number of plot records available exceeded the number required, we conducted a heterogeneity-constrained resampling of the plot set (Lengyel et al. 2011). Otherwise, all plot records were retained. The procedure was repeated using all combinations of the different grid cell sizes and different coefficients for the relationship between the β -diversity of the cell and the required number of plots (Fig. 1). To determine an appropriate combination of grid cell size and β -diversity coefficient, we required the total number of plot records selected to be at least 75% of the required plot records. This ensured that the compositional variation within and among grid cells was adequately represented. Moreover, we required the stratification to retain between 25% and 45% of all plot records. Among the combinations that met both criteria, we chose the one with the smallest grid cell size and largest β -diversity coefficient. Finally, we created a combined data table, including both the 1177 LUCAS plots and the set of stratified NVS plots, to be used in subsequent analyses.

Clustering analyses

At the alliance level, our aim was to extend the original classification by defining new alliances. Initially, we had to determine which plot records could be used to define new alliances. To achieve this, we used the NC membership rules for alliances ($\delta = 0.83$) to calculate the membership

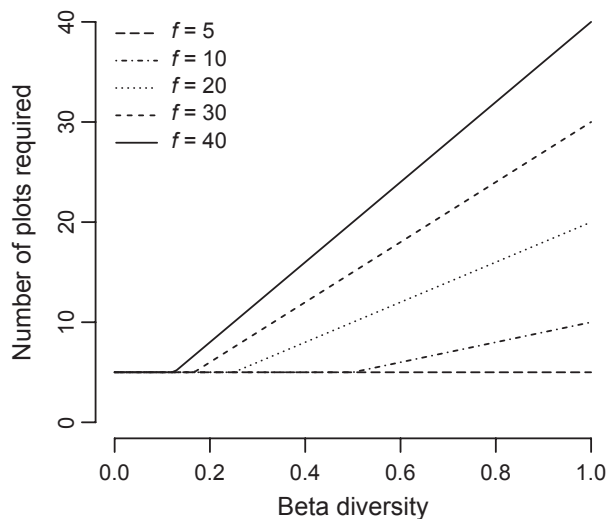


Fig. 1. Truncated linear relationship between the variation in species composition (i.e. β -diversity) of grid cells and the number of plots needed to be resampled. Different slope coefficients (f) are displayed.

of all the plot records in the combined data set (LUCAS and stratified NVS) to the original alliances, and we designated those plot records whose membership to the Noise class was higher than 0.5 as ‘unassigned’ (Table 1). We applied the Noise clustering algorithm to this set of unassigned plot records to define from one to 25 new clusters. In all these executions of the NC algorithm, the centroids of the accepted original alliances were used as fixed elements so that the newly defined alliances would be as distinct as possible from the originals (De Cáceres et al. 2010). After defining the new clusters, we re-calculated the membership of all plot records in the combined data set. All these analytical steps are summarized in a workflow diagram provided in Appendix S2.

We also aimed to define vegetation types at a lower level of abstraction (i.e. a finer level of resolution) than the previous alliances. We defined associations using all the plot records in the combined data set. In this case, we set the distance to the Noise class to $\delta = 0.75$, following De Cáceres et al. (2010). The NC algorithm was used to obtain fuzzy classifications from two to 100 clusters.

To relate the two levels of abstraction, we defined the fuzzy membership of an association into an alliance by assessing the proportion of plot records assigned to the association that were also assigned to the alliance. Let u_{ij} be the membership value of plot record j into association i , and v_{kj} be the membership value of plot record j into alliance k . The membership of association i into alliance k , w_{ki} , was calculated as (see Appendix S1 for details):

$$w_{ki} = \frac{\sum_{j=1}^n u_{ij}v_{kj}}{\sum_{j=1}^n u_{ij}} \quad (2)$$

The fuzzy membership of an association into the Noise class was calculated as one minus the sum of memberships to the true alliances.

Deciding on the number of alliances and associations

We developed several evaluation methods to guide our choice of the number of new alliances and associations to recognize. First, for each number of alliances or associations, we calculated the proportion of plots in each of the categories of Table 1. When the number of clusters increases, the number of ‘unassigned’ plot records should decrease, but these plot records may become ‘transitional’; our goal was to maximize the number of plots ‘clearly assigned’. Second, for each number of alliances, we determined how many new alliances did not comply with the acceptance criteria used for the original alliances (i.e. had fewer than or equal to 20 plot records and a cluster variance higher than 0.6). Similarly, we required associations to be defined by ten or more plot records and to have a cluster variance ≤ 0.5 .

The NC algorithm can leave large parts of the multivariate space in the Noise class. To ensure that the multivariate space spanned by associations and alliances was similar, we calculated two additional statistics for each combination of number of associations and number of alliances. First, we counted the number of associations having ≥ 20 plot records and high membership (>0.5) to the Noise class when assigned at the alliance level. If many of the plot records of an association are assigned to the Noise class in the alliance-level analysis, the multivariate space spanned by associations is larger than that spanned by alliances. Since, however, we allowed associations to have fewer numbers of plots than alliances (i.e. 10 vs 20), associations could also have high membership to the Noise class when there were insufficient plots to define the corresponding alliance. Therefore, we allowed associations to belong to the Noise class if they comprised fewer than 20 plot records. Second, we counted the number of alliances with no component associations. We first determined, for each association, which was the closest true alliance (i.e. the one for which the association showed highest membership); then we counted the number of alliances without associations. If an alliance receives no associations, this indicates that the region of the multivariate space spanned by alliances is larger than that spanned by associations.

Final assignments and characterization of vegetation types

We used the NC membership rules for alliances ($\delta = 0.83$) and associations ($\delta = 0.75$) to obtain the membership of the NVS plot records that had been discarded by the

stratification process into the final alliances and associations. Clearly assigned plots were used to determine the characteristics of the newly defined vegetation types. For each type, we assessed the altitudinal range, geographical extent, mean species richness and the mean percentage of exotic species. The geographic extent was obtained by calculating the proportion of the 1177 LUCAS plots that were assigned to the type relative to the total area of 8.9 million ha mapped as forest or shrubland by LCDB.

Recognizing known rare forest types

We used the New Zealand Forest Class Maps (NZFCM; New Zealand Forest Service 1971–1990) to identify a selection of forest types having a spatial extent that was too small to be defined by Wiser et al. (2011). We then tested whether our approach allowed such rare types to be defined. From the 1950s, Forest Class Maps were compiled at a scale of 1:250 000 (NZ Forest Service Mapping Series 6), and provided national coverage (excluding Fiordland) of 18 forest classes. NZFCM types that did not correspond to any of the alliances described in Wiser et al. (2011) include Kauri, Kauri–Softwood–Hardwoods–Beeches, Rimu–Taraire–Tawa, and Taraire–Tawa. These types had mapped extents ranging from 2722 to 40 486 ha, whereas the most narrowly distributed alliances of Wiser et al. (2011) had extents of 144 000 ha. Hereafter these types will be described as *Agathis australis*, *Agathis australis* – *Softwood–Hardwoods–Nothofagus*, *Dacrydium cupressinum* – *Beilschmiedia tarairi* – *B. tawa*, and *Beilschmiedia tarairi* – *B. tawa* forests, respectively. To test the recognition of these NZFCM types in the new classification we examined the occurrence and dominance of nominal taxa within our newly defined vegetation types.

Characterizing plots classed as outliers

The NVS plots are primarily located on public land; as such they are concentrated in montane areas and under sampled lowland areas. Within forests, geographic coverage of plots is widespread, but uneven, with some areas better represented than others. Successional shrublands are also poorly represented (Wiser et al. 2001). Our expectation is that plot data in such areas and ecosystems will be insufficient to define vegetation types, resulting in more plot records being assigned to the Noise class. To determine if this is the case, we used the Wilcoxon rank sum test to compare the altitude and mean top height of plots assigned to the Noise class to those that were not (Table 1). To determine whether successional shrublands were overrepresented in the Noise class compared to other kinds of shrublands (e.g. sub-alpine) we focused on the subset of plots with mean top height below 6 m (following McGlone

et al. 2010) and used a Wilcoxon rank sum test to compare the mean percentage of exotic species (early successional shrublands are frequently highly invaded) of plots assigned to the Noise class to those that were not. Finally, we had the more general expectation that where compositional gradients had been inadequately sampled, using our criteria derived from the truncated linear function of β -diversity, more plots would be assigned to the Noise class. We tested this by comparing the ratio of NVS plots available to those required for each grid cell with the proportion of plots assigned to the Noise class in that cell using Spearman rank correlation. Only the 5751 plots used to derive the classification were used in this analysis.

Results

Consequences of casting original classification into NC framework

Casting the original classification into the NC fuzzy classification framework had significant consequences. First, switching from Manhattan to Chord distance produced important changes in the dissimilarity between pairs of individual plot records, although the distances between alliance centroids were much less modified (Appendix S3). Second, applying the NC alliance membership rule resulted in some changes in plot membership of the original alliances (Appendix S4). Third, seven original alliances (mostly successional shrublands and forests) were poorly defined according to our validation criteria (Fig. 2) and were excluded from subsequent analyses. The remaining 17 original alliances were used as fixed elements in the definition of new alliances.

Reducing geographic and compositional bias from the NVS phytosociological data

The application of our criteria for stratification led to two potentially suitable combinations of grid cell size and β -diversity coefficient (Table 2). Among them, we chose a stratification using a grid point interval size of 8 km (i.e. grid cell area of 64 km²) and a β -diversity coefficient of 20, and this resulted in a selection of 4574 plots (37.0% of the original data set and 75.2% of the plots required to adequately represent β -diversity). Merging the stratified NVS data set and the LUCAS data set yielded a data set containing 5751 plot records and 1930 species (1237 species in common, 427 species in NVS that do not appear in LUCAS and 266 species in LUCAS that do not appear in NVS).

Cluster analysis results

When applying the NC membership rule for the provisional assignment of the 5751 plot records, 1771

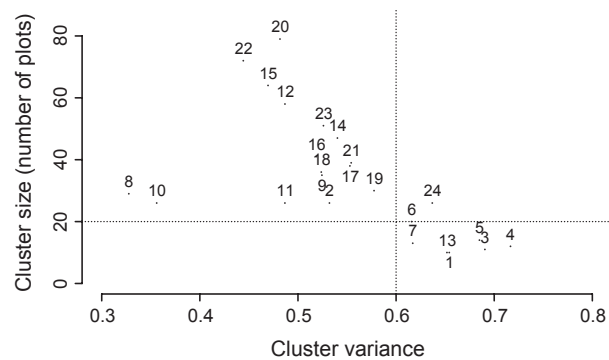


Fig. 2. Cluster variance and cluster size (number of plot records) of original alliances. Alliance numbering follows Wiser et al. (2011). Dashed lines indicate thresholds (0.6 for cluster variance and 20 plots for cluster size). These were used to determine which original alliances were discarded in the subsequent fuzzy cluster analyses.

(30.8%) were left unassigned. These plots were used to create classifications ranging from $c = 17 + 1 = 18$ to $c = 17 + 25 = 42$ alliances. We calculated evaluation statistics after reassigning the remaining 3980 plot records. Beyond 31 (=17 + 14) alliances the proportion of clearly assigned plots did not increase and the reductions observed from the Noise class went to increasing the number of transitional plots (Fig. 3a). Examining the number of alliances that had low membership and high cluster variance at the same time (Fig. 3b), we concluded that the maximum number of alliances to be defined should be 31 (=17 + 14). At the association level, the changes in the proportion of plots belonging to categories of Table 1 indicated that more than 40 associations should be defined (Fig. 3c). Associations with high cluster variance (>0.5) appeared even with low numbers of associations. In contrast, associations with low numbers of members (<10) only appeared after 58 clusters (Fig. 3d).

Table 2. Plot data sets resulting from geographical and environmental re-sampling of National Vegetation Survey (NVS) data at different grid cell sizes and different slope coefficients for the linear function of β -diversity (f).

f	Grid cell size (km ²)				
	4	16	36	64	100
5	69.6/58.4	43.0/65.3	30.8/68.8	24.3/71.6	20.5/ 75.1
10	70.7/58.8	44.2/65.9	32.2/69.6	25.7/72.4	22.0/ 76.2
20	84.0/59.4	59.7/69.4	45.5/72.8	37.0/75.2	32.4/77.8
30	90.6/54.1	72.0/66.4	56.2/69.6	46.7/72.1	41.2/73.7
40	93.3/48.3	80.0/61.2	64.2/64.7	54.2/67.4	48.2/68.5

Left values indicate the percentage of the total number of NVS plot records (i.e. 12 374); right values indicate the percentage of the number of plot records required to represent β -diversity in each case. Values in bold indicate compliance with stratification criteria.

Assignment of associations into alliances proved to be useful to guide the decision of the final number of clusters to define at both levels. First, we observed that for between 24 and 29 alliances the number of associations should be <80 in order to avoid associations with ≥ 20 plot records having high membership in the Noise class (Fig. 4a). For more than 29 alliances the number of associations should be at least 80 to avoid alliances with no component associations (Fig. 4b). Finally, we decided to retain 29 (=17 original + 12 new) alliances and 79 associations, because, although we wanted to define as many valid vegetation types as possible, 30 or 31 alliances would lead to some alliances having no component associations and a larger number of associations would lead to some of them being invalid or not being related to any recognized alliance. When the assignment of associations into alliances is categorized (following Table 1), there are 46 associations clearly assigned to one alliance, four associations assigned to the Noise class, and 29 transitional associations (i.e. between different alliances or between alliances and the Noise class).

Final assignment of all plot records

Across all plot records (i.e. stratified NVS + LUCAS and the remaining 7800 NVS plot records), 53.2% were clearly assigned to the original alliances and 9.5% were clearly assigned to the new ones (Table 3). That the portion of plot records clearly assigned to original alliances was much lower for the stratified NVS data set (39.8%) than for the LUCAS data set (60.3%) supports our premise that atypical plant communities were not well sampled with the grid-based sampling of the 1177 LUCAS plots. Around 25.5% of the plot records were considered transitional at the alliance level, and 11.9% were left unassigned (i.e. assigned to the Noise class). In comparison, at the association level fewer plot records were considered transitional, while more of them fell into the Noise class (Table 3).

Characterization of the newly defined vegetation types

We were able to define 12 new alliances, having estimated extents ranging from <7562 ha to 120 986 ha (Table 4); all of these were too narrowly distributed to be defined by the original classification based on 1177 plots, where the most narrowly distributed alliance had an extent of 144 000 ha (Wiser et al. 2011). Six new narrowly distributed shrubland alliances were defined, whereas only one that is broadly distributed (extent of 204 163 ha) was retained from the original classification (*Discaria toumatou* – *Coprosma propinqua* – *Anthoxanthum odoratum* – *Dactylis glomerata* shrubland). Of these, five predominantly occur below 1000 m in landscapes that would have supported

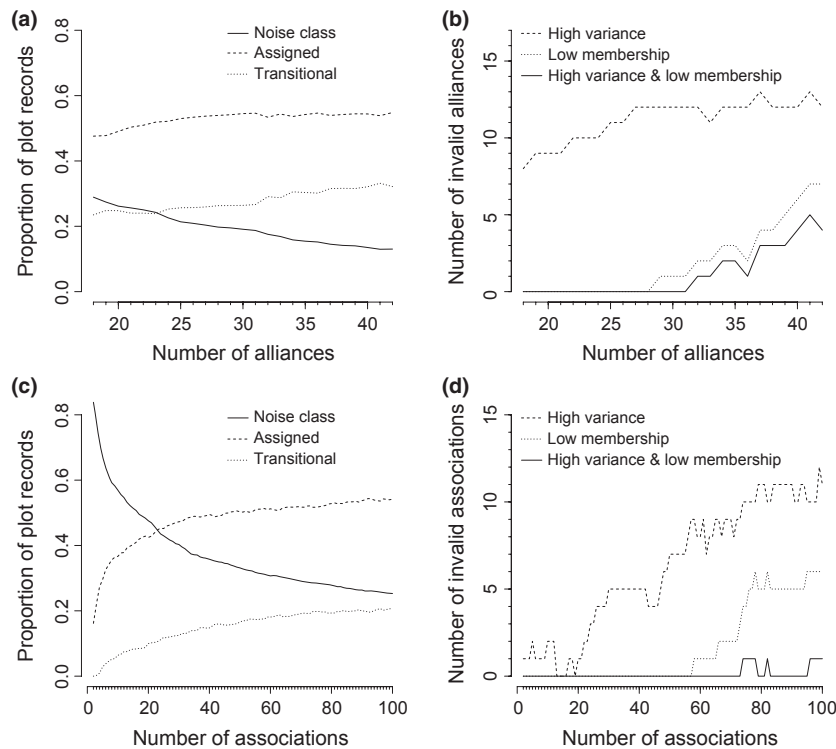


Fig. 3. Evaluation statistics of classification results at the level of alliance (a–b) and association (c–d): (a) and (c) proportion of plots belonging to the three categories; (b) number of alliances with high cluster variance (>0.6), low membership (≤ 20) or both. The x-axis starts at 18, as it accommodates the 17 alliances already defined; (d) number of associations with high cluster variance (>0.5), low membership (≤ 10) or both.

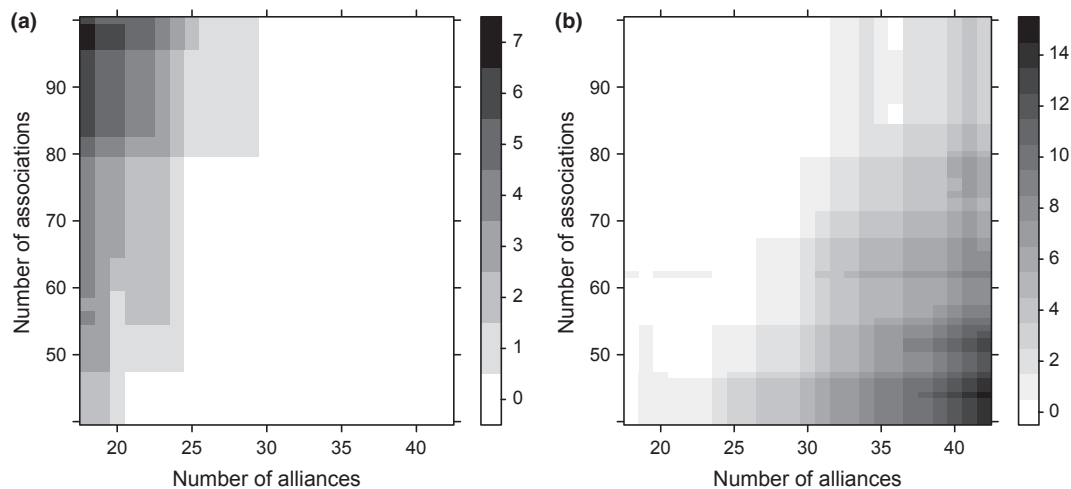


Fig. 4. Assignment of associations into alliances. For different combinations of numbers of alliances and association, we plot (a) the number of associations with membership >0.5 to the Noise class and having ≥ 20 records, and (b) the number of alliances with no component associations (after excluding the alliance-level Noise class – see text).

forests in pre-human times, and two [*Ulex europaeus* – *Cordyline australis* shrubland and (*Kunzea ericoides*)/*Coprosma rhamnoides*/*Dactylis glomerata* – *Anthoxanthum odoratum* successional shrubland] have both particularly high representation of exotic species (Table 4) and exotics included in

the nominal species (i.e. the nitrogen-fixing shrub *Ulex europaeus* and the turf-forming grasses *Dactylis glomerata* and *Anthoxanthum odoratum*). Of note is how much more narrowly distributed shrubland alliances tend to be than forest alliances (shrubland alliances range from <7562 to

Table 3. Final assignment results at the alliance and association levels.

	LUCAS	Stratified NVS	LUCAS + Stratified NVS	Remaining NVS	All plot records
Total no. plots	1177	4574	5751	7800	13 551
Alliance level (%)					
Assigned	784 (66.6)	2330 (50.9)	3114 (54.1)	5378 (68.9)	8492 (62.7)
Original alliances	710 (60.3)	1819 (39.8)	2529 (44.0)	4674 (59.9)	7203 (53.2)
New alliances	74 (6.3)	511 (11.2)	585 (10.2)	704 (9.0)	1289 (9.5)
Transitional	236 (20.1)	1279 (28.0)	1515 (26.3)	1934 (24.8)	3449 (25.5)
Unassigned	157 (13.3)	965 (21.1)	1122 (19.5)	488 (6.3)	1610 (11.9)
Association level (%)					
Assigned	569 (48.3)	2413 (52.8)	2982 (51.9)	5440 (69.7)	8422 (62.2)
Transitional	261 (22.2)	877 (19.2)	1138 (19.8)	1625 (20.8)	2763 (20.4)
Unassigned	347 (29.5)	1284 (28.1)	1631 (28.4)	735 (9.4)	2366 (17.5)

Columns indicate assignment results for different sets of plot records; values in each column are the number of plots assigned followed by the percentage of the set that this value represents.

Table 4. Characterization of new alliances.

Alliance	Distribution	Altitude range (m)	Extent (ha)	Mean species richness	Mean% exotic	No. associations defined
Shrublands						
<i>Cordyline australis</i> tall shrubland	SI	60–300	<7562	3	6	1
<i>Ulex europaeus</i> – <i>Cordyline australis</i> shrubland	NI, SI	2–570	15 123	6	39	2
<i>Dracophyllum uniflorum</i> / <i>Gaultheria crassa</i> – <i>Poa colensoi</i> – <i>Festuca novae-zelandiae</i> montane shrubland	SI	730–1360	60 493	37	5	1
<i>Kunzea ericoides</i> /(<i>Coprosma rhamnoides</i> – <i>Leptocophylla juniperina</i>) shrubland	NI, SI	20–1010	120 986	27	14	2
(<i>Kunzea ericoides</i>)/(<i>Coprosma rhamnoides</i> / <i>Dactylis glomerata</i> – <i>Anthoxanthum odoratum</i> successional shrubland	NI, SI	50–700	75 616	45	46	1
<i>Leptospermum scoparium</i> shrubland	NI, SI, Stewart	0–1020	37 808	20	3	2
Nothofagus Forests						
<i>Nothofagus menziesii</i> / <i>Hoheria glabrata</i> – <i>Myrsine divaricata</i> – <i>Coprosma ciliata</i> / <i>Polystichum vestitum</i> montane forest	SI	460–1260	75 616	28	<1	1
<i>Nothofagus truncata</i> – <i>Weinmannia racemosa</i> (<i>N. menziesii</i>)/ <i>Leucopogon fasciculatus</i> forest	NI, SI	50–990	68 054	18	<1	4
Podocarp Forest						
<i>Dracophyllum traversii</i> – <i>D. longifolium</i> – <i>Coprosma pseudocuneata</i> – <i>Archeria traversii</i> low forest and sub-alpine shrubland	SI	740–1210	22 685	40	<1	2
Broad-Leaved-Podocarp Forest (including kauri)						
<i>Pseudowintera colorata</i> – <i>Fuchsia excorticata</i> – <i>Griselinia littoralis</i> / <i>Polystichum vestitum</i> forest	NI, SI	0–900	45 370	39	2	2
<i>Prumnopitys taxifolia</i> / <i>Meliclytus ramiflorus</i> – <i>Myrsine australis</i> forest	SI	20–685	15 123	39	7	1
<i>Weinmannia silvicola</i> – <i>Beilschmiedia tawa</i> / <i>Freyinetia banksii</i> forest	NI	30–640	22 685	50	0	1

SI, South Island; NI, North Island; Stewart, Stewart Island.

204 163 ha, median = 60 493 ha, whereas forest alliances range from 15 123 to 574 681 ha, median = 241 970 ha).

We defined six new forest alliances (Table 4). Of these, three are classed as Broad-leaved-Podocarp forest, one as Podocarp forest (grading into sub-alpine shrubland at high elevations) and two as *Nothofagus* forest. The most narrowly distributed *Nothofagus* alliance is noteworthy in being dominated by *N. truncata*; none of the alliances defined by Wiser et al. (2011) were dominated by this spe-

cies, which is the most narrowly distributed of all the NZ *Nothofagus* species (Wardle 1984). We also recognized a new, very rare (estimated extent 15 123 ha) forest alliance, the *Prumnopitys taxifolia*/*Meliclytus ramiflorus* – *Myrsine australis* forest alliance, which is primarily restricted to fragmented locations on the South Island east coast, but was most likely more extensively distributed before human settlement (Moar 2008). This community was only briefly described under 'low-altitude conifer broad-leaf forest' by

Wardle (1991) and was subsumed under the more extensive type 'Rimu-Matai-Hardwoods' by the NZFCM. New alliances and their component associations are described in Appendix S5, and component associations of original alliances are listed in Appendix S6.

Representation of known rare forest types

As the 1177 plot LUCAS data set included only 11 and 26 plots containing the narrowly distributed *Agathis australis* and *Beilschmiedia tarairi* respectively, alliances where these tree species were dominant could not be defined. In contrast, the 5751 plot LUCAS + stratified NVS data set included 51 and 88 plots where these species occurred, respectively. This allowed the *Weinmannia silvicola*–*Beilschmiedia tawa*/*Freycinetia banksii* forest alliance and the *Weinmannia silvicola*–*Beilschmiedia tarairi*–*Beilschmiedia tawa* – (*Dysoxylum spectabile*–*Agathis australis*)/*Freycinetia banksii*–*Dicksonia squarrosa* forest association to be defined (Table 4, Appendices S5 and S6). This alliance and association includes stands that correspond to the concepts for two of the four rare NZFCM classes – *Dacrydium cupressinum* – *Beilschmiedia tarairi* – *B. tawa*, and *Beilschmiedia tarairi* – *B. tawa* forest and stands of the NZFCM *Agathis australis* class on the more mature end of the spectrum, but did not segregate them from each other. The *Weinmannia silvicola* – (*Agathis australis*)/*Leptospermum scoparium* – *Knightia excelsa*/*Leucopogon fasciculatus* – *Blechnum novae-zealandiae* forest association defines stands of the *Agathis australis* class on the more recently disturbed end of the spectrum, as evidenced by the abundance of post-disturbance colonizing trees such as *Leptospermum scoparium* and *Knightia excelsa*.

We did not describe any new alliances that correspond to the NZFCM class *Agathis australis* – *Softwood*–*Hardwoods*–*Nothofagus*. Our data set included only two plots where *Agathis* and *Nothofagus* co-occur; both were sufficiently distinct in composition to be assigned to the alliance-level Noise class.

Characterizing plots classed as outliers

Our expectation that in grid cells where compositional gradients had been inadequately sampled more plots would be assigned to the Noise class was met, although the relationship was weak (Spearman rank correlation between the ratio of NVS plots available to those required for each 64-km² grid cell and the proportion of plots assigned to the Noise class in that cell = -0.16 , $P < 0.00001$). As we also expected, plots assigned to the Noise class, at both the alliance and association level, tended to occur at lower elevations (mean elevation = 547 m vs 468 m, for alliance-level non-Noise and Noise class plots, respectively, $P \leq 0.0001$, and 551 m vs 480 m for the analogous

association-level plots, $P \leq 0.0001$). They were also shorter in stature (mean canopy height = 16.2 m vs 9.1 m for alliance-level non-Noise and Noise class plots, respectively, $P \leq 0.0001$, and 16.9 m vs 9.4 m for the analogous association-level plots, $P \leq 0.0001$). When we examined short-statured stands only, plots with a higher proportion of exotics also tended to be assigned to the Noise class (mean percentage exotic = 13.7 vs 14.3 for alliance-level non-Noise and Noise class plots, respectively, $P = 0.03$, and 10.4 and 15.6 for the analogous association-level plots, $P = 0.0002$).

Discussion

The on-going exploration of the vegetation continuum, both in space and time, turns any given classification of vegetation into a set of conventions that may need to be modified or extended (De Cáceres & Wiser 2012). In the traditional Braun-Blanquet method, expert phytosociologists dealt with this need by incorporating their knowledge on diagnostic species of previously defined types (either belonging to the same or different levels of abstraction) in the classification process of a particular relevé table. Following and extending the approach presented by De Cáceres et al. (2010), we have illustrated here how this need to extend existing classifications can be incorporated into numerical classification in a way that is formalized and repeatable. Overall, our approach illustrates the application of a fuzzy classification framework at a national scale, and can serve as a model for others wishing to extend and update their regional/national vegetation classifications.

Casting an existing classification into a new classification framework

When using two clustering methods on the same set of plot records, parts of the continuum will be similarly clustered while others will not. A similar situation occurs when casting the classification obtained from one method into a new classification framework. Even if one tries to preserve the original classification as much as possible, some of the original types may be poorly defined according to the way clusters are defined in the new framework, as we observed for seven of the 24 alliances originally defined by Wiser et al. (2011). Here, we assessed the robustness of the classification to changes in the clustering framework by focusing on the consequences for the validity of specific clusters (alliances). While we defined invalid vegetation types as those having a small number of members and high cluster dispersion, other methods can be employed to assess the validity of previously defined vegetation types

when changing the clustering framework (e.g. Feoli et al. 2006). In general, the validation of vegetation classifications is an area that deserves further research. For example, recent progress has been made to assess the stability of classification results to sampling variations (Tichý et al. 2011).

Sampling and the extension of vegetation classifications

An important issue reflecting the provisional character of classifications is the need to define vegetation types on parts of the continuum that so far have not been described because of the sampling design of earlier studies. Wiser et al. (2011) noted that their original classification based on objectively sampled plots failed to capture some previously documented, iconic forest types because of their rarity in the landscape.

An appropriate sampling design is critical to adequately represent compositional combinations that are geographically rare. By applying a geographic and compositional stratification approach to the NVS database, the proportion of plot records that represent compositionally rare combinations was decidedly increased in our resampled data set. This allowed us to define 12 new alliances for species assemblages having extents smaller than 125 000 ha. Because of the relatively small size of the LUCAS data set, the original classification was unable to partition alliances into associations, as is particularly desirable at the regional scale. In our extension analysis, we were able to recognize more than one association (and as many as seven) within 14 of the 17 original alliances and seven of the 12 new alliances.

Our approach allowed us to define alliances or associations that corresponded to three of the four rare forest classes identified by the NZFCM, although our community concepts were either broader or narrower than those of the NZFCM. This demonstrates that when data are available, our approach can allow rare vegetation types to be defined. Further, the presence of the narrowly distributed *Agathis australis* and *Beilschmiedia tarairi*, are characteristic of these rare NZFCM classes, in the unassigned plots (*A. australis* occurs in 46 and 73 of the alliance and association Noise classes respectively, *B. tarairi* occurs in 49 and 91), suggests that with more data further rare types may be able to be described.

Although large, our data set did not have enough plots where *Agathis* co-occurs with *Nothofagus* to allow us to describe an alliance or association corresponding to the NZFCM *Agathis australis* – *Softwood–Hardwoods–Nothofagus* class. We have identified 101 plots in the NVS databank where these two species co-occur, but unfortunately these plots use tier classes or abundance measures that were incompatible with the original classification. How to make

different abundance measures compatible for classification purposes is a challenge that may be worthwhile to explore in future research.

Dealing with plot records that are transitional or remain unclassified

A second consequence of stratifying the data set to increase the proportion of compositionally rare species assemblages was to increase the number of plot records that were either transitional or left unassigned (Table 3). Although vegetation scientists accept the vegetation continuum, there is often a need to put every plot observation into a known vegetation type (e.g. Kočí et al. 2003). Adopting a framework where plot records can be transitional or even left unclassified may be regarded initially as impractical. However, it is desirable to exclude transitions and unassigned plots from the set of plots used to define vegetation types, because this ensures a more distinct characterization of those types (e.g. number and identity of diagnostic species, distinct environmental and geographical range). Since many existing quantitative classification algorithms require that all plots in the data set be assigned to a type, the analyst has to make any decisions about exclusion/inclusion of plots having unusual composition before the analysis is begun. Noise clustering provides a way to identify such outliers and exclude them from the classification, thus making the defined types more robust and cohesive in composition. Moreover, if a single answer is desired, transitional or unassigned plot records may be *a posteriori* assigned to the closest type. Despite the advantage of allowing a more distinct definition of vegetation types, acknowledging transitional and unassigned plot records creates the additional task of determining why those plots were not clearly assigned.

Using outliers to direct future enquiry

Plot records that are left unassigned represent species assemblages that are uncommon in the study area. Allowing plots to remain unassigned is important because it retains the option to extend the current classification by defining new types as new data become available. Inspecting the characteristics of unassigned plots may indicate the parts of the vegetation continuum for which additional data are required. Our stratification approach can be used to determine which locations (here 64-km² grid cells) appear to be under-sampled relative to their apparent compositional variation. Further, our analysis confirmed our expectations that outliers tend to occur at lower altitudes or in shrublands that are often successional. Other results provide further evidence that shrubland vegetation types are challenging to define. Of the seven alliances that were

poorly defined according to the new analytical framework, five were shrublands. Luckily, with the addition of more plot data some of their plots were assigned to more compositionally cohesive new shrubland alliances. The inherent difficulty of cohesively defining shrubland types is further reiterated in that three of the four associations assigned to the alliance-level Noise class were shrublands.

Why was it so difficult for us to define shrubland community types? The most proximate explanation is that such ecosystems are under-sampled relative to their spatial extent. Indeed although indigenous shrublands are estimated to cover 30% of the area of woody indigenous vegetation in New Zealand (Thompson et al. 2004) only 13% of the plots in our stratified data set sampled shrublands. Certainly there is an urgent need for more plot data to be collected or gathered to allow shrubland types to be more comprehensively defined. Nevertheless, NZ shrubland communities may be inherently more challenging to define than forest communities. That the shrubland alliances that we did define are estimated to be more narrowly distributed than forest alliances implies that the former exhibit more compositional turnover between locations. We suggest two potential explanations for this. Island areas, such as New Zealand, tend to have a disproportionate number of shorter (i.e. <15 m) arborescent species trees that have narrow ranges (McGlone et al. 2010). The absence of wide-ranging, tall species from sub-alpine shrublands ensures higher compositional turnover among sub-alpine shrublands than among nearby forests. In successional shrublands, where the native seral woody species tend to have wide ranges (McGlone et al. 2010), composition may be inherently unpredictable from environmental factors, and less repeated across geographic gradients because site history, seed availability and on-going disturbance are the primary determinants of composition (Christensen & Peet 1984; Wardle 1991; Ozinga et al. 2005). In regions, such as New Zealand, exotic invaders may be an important component of woody successions, and these invaders will differ in their residence times and the degree to which they have established in suitable habitats. This is further compounded when exotic invaders have novel traits that alter successional trajectories (e.g. D'Antonio & Vitousek 1992; Sullivan et al. 2007). Defining distinct robust vegetation types in these cases may be a challenging task.

Acknowledgements

SKW was supported by New Zealand Department of Conservation and the New Zealand Ministry of Science and Innovation contract (C09X0916, <http://www.frst.govt.nz/>). MDC was supported by a Beatriu de Pinós postdoctoral

grant (2009 BP-B 00342) from the Catalan Agency for Management of University and Research Grants, and a Consolider Montes CSD2008-00040 project granted by the Spanish Ministry of Education and Science (MEC) and the International Mobility Fund administered by the Royal Society of New Zealand (contract SPN 10-13). We acknowledge the use of data drawn from the National Vegetation Survey Database (NVS) and all of those people who wore out their boots collecting the original data. We thank Hamish Maule for working tirelessly on data corrections and Shirley Vickers for helping us link growth form, taxonomic, geographic and plot data. We thank Norm Mason, Rob Allen and Christine Bezar, Enrico Feoli, and Zoltán Botta-Dukát for helpful comments on earlier versions, and Sarah Richardson, Peter Bellingham, John Leathwick, Elaine Wright, Fiona Thomson and Matt McGlone for useful discussions.

References

- Allen, R.B., Bellingham, P.J. & Wiser, S.K. 2003. Developing a forest biodiversity monitoring approach for New Zealand. *New Zealand Journal of Ecology* 27: 207–220.
- Bezdek, J.C. 1981. *Pattern recognition with fuzzy objective functions*. Plenum Press, New York, NY, US.
- Botta-Dukát, Z., Kovács-Láng, E., Rédei, T., Kertész, M. & Garadnai, J. 2007. Statistical and biological consequences of preferential sampling in phytosociology: theoretical considerations and a case study. *Folia Geobotanica* 42: 141–152.
- Braun-Blanquet, J. 1929, 1964. *Pflanzensoziologie, grundzüge der vegetationskunde*. 1. Aufl. Springer, Berlin, DE X + 330 pp; 3. Aufl. Springer, Vienna, At–New York, US. XIV + 845 pp.
- Christensen, N.L. & Peet, R.K. 1984. Convergence during secondary forest succession. *Journal of Ecology* 72: 25–36.
- Chytrý, M. & Račajová, M. 2003. Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. *Preslia* 75: 1–15.
- Cockayne, L. 1928. *The vegetation of New Zealand*, 2nd ed. Engelmann Press, Leipzig, DE.
- Coomes, D.A., Allen, R.B., Scott, N.A., Goulding, C. & Beets, P. 2002. Designing systems to monitor carbon stocks in forests and shrublands. *Forest Ecology and Management* 164: 89–108.
- Cooper, A., McCann, T. & Bunce, R.G.H. 2006. The influence of sampling intensity on vegetation classification and the implications for environmental management. *Environmental Conservation* 33: 118–127.
- D'Antonio, C.M. & Vitousek, P.M. 1992. Biological invasions by exotic grasses, the grass/fire cycle, and global change. *Annual Review of Ecology and Systematics* 23: 63–87.
- Dave, R.N. 1991. Characterization and detection of noise in clustering. *Pattern Recognition Letters* 12: 657–664.
- De Cáceres, M. & Wiser, S. 2012. Towards consistency in vegetation classification. *Journal of Vegetation Science* 23: 387–393.

- De Cáceres, M., Font, X. & Oliva, F. 2010. The management of numerical vegetation classification with fuzzy clustering methods. *Journal of Vegetation Science* 21: 1138–1151.
- Dengler, J., Berg, C. & Jansen, F. 2005. New ideas for modern phytosociological monographs. *Annali di Botanica Nuova Serie* 5: 193–210.
- Dengler, J., Jansen, F., Glöckler, F., Peet, R.K., De Cáceres, M., Chytrý, M., Ewald, J., Oldeland, J., Lopez-Gonzalez, G., Finckh, M., Mucina, L., Rodwell, J.S., Schaminée, J.H.J. & Spencer, N. 2011. The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science. *Journal of Vegetation Science* 22: 582–597.
- Ewald, J. 2003. A critique for phytosociology. *Journal of Vegetation Science* 14: 291–296.
- Faith, D.P., Minchin, P.R. & Belbin, L. 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69: 57–68.
- Feoli, E., Ferro, G. & Ganis, P. 2006. Validation of phytosociological classifications based on a fuzzy set approach. *Community Ecology* 7: 99–108.
- Franz, N.M. & Peet, R.K. 2009. Towards a language for mapping relationships among taxonomic concepts. *Systematics and Biodiversity* 7: 5–20.
- Hurst, J.M. & Allen, R.B. 2007. *A permanent plot method for monitoring indigenous forests – field protocols*. Manaaki Whenua – Landcare Research, Lincoln, NZ.
- Jansen, F. & Dengler, J. 2010. Plant names in vegetation databases – a neglected source of bias. *Journal of Vegetation Science* 21: 1179–1186.
- Kočí, M., Chytrý, M. & Tichý, L. 2003. Formalized reproduction of an expert-based phytosociological classification: a case study of sub-alpine tall-forb vegetation. *Journal of Vegetation Science* 14: 601–610.
- Legendre, P., Borcard, D. & Peres-Neto, P.R. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs* 75: 435–450.
- Lengyel, A., Chytrý, M. & Tichý, L. 2011. Heterogeneity-constrained random resampling of phytosociological databases. *Journal of Vegetation Science* 22: 175–183.
- McCune, B. & Grace, J.B. 2002. *Analysis of ecological communities*. MjM Software Design, Gleneden Beach, OR, US.
- McGlone, M.S., Richardson, S.J. & Jordon, G.J. 2010. Comparative biogeography of New Zealand trees: species richness, height, leaf traits and range sizes. *New Zealand Journal of Ecology* 34: 137–151.
- McKelvey, P.J. & Nicholls, J.L. 1957. A provisional classification of North Island forests. *New Zealand Journal of Forestry* 7: 8–101.
- Moar, N. 2008. Late Quaternary vegetation. In: Winterbourn, M., Knox, G., Burrows, C. & Marsden, I. (eds.) *The natural history of canterbury*. pp. 169–192. Canterbury University Press, Christchurch, NZ.
- Mucina, L. 1997. Classification of vegetation: past, present and future. *Journal of Vegetation Science* 8: 751–760.
- New Zealand Forest Service 1971–1990. *Forest service mapping series 6*. New Zealand Forest Service, Wellington, NZ.
- Økland, R.H. 2007. Wise use of statistical tools in ecological field studies. *Folia Geobotanica* 42: 123–140.
- Orlóci, L. 1967. An agglomerative method for classification of plant communities. *Journal of Ecology* 55: 193–206.
- Ozinga, W.A., Schaminée, H.J., Bekker, R.M., Bonn, S., Poschod, P., Tackx, O., Bakker, J. & van Groenendael, J.M. 2005. Predictability of plant species composition from environmental conditions is constrained by dispersal limitation. *Oikos* 108: 555–561.
- Pignatti, S. 1980. Reflections on the phytosociological approach and the epistemological basis of vegetation science. *Vegetatio* 42: 181–185.
- Roleček, J., Chytrý, M., Hájek, M., Lvončík, S. & Tichý, L. 2007. Sampling design in large-scale vegetation studies: do not sacrifice ecological thinking to statistical purism! *Folia Geobotanica* 42: 199–208.
- Sullivan, J.J., Williams, P.A. & Timmins, S.M. 2007. Secondary forest succession differs through naturalized gorse and native kānuka near Wellington and Nelson. *New Zealand Journal of Ecology* 31: 22–38.
- Thompson, S., Gruner, I. & Gapare, N. 2004. *New Zealand Land Cover Database Version 2. Illustrated guide to target classes*. Report for the Ministry for the Environment, Wellington, NZ. Available on CD-ROM as The illustrated guide to target land cover classes, at <http://www.terralink.co.nz/docs/lcdborderform.pdf>. Accessed 24 January 2011.
- Tichý, L., Chytrý, M. & Šmarda, P. 2011. Evaluating the stability of the classification of community data. *Ecography* 34: 807–813.
- Wardle, J.A. 1984. *The New Zealand beeches: ecology, utilisation and management*. New Zealand Forest Service, Christchurch, NZ.
- Wardle, P. 1991. *Vegetation of New Zealand*. Cambridge University Press, Cambridge, UK.
- Wiser, S.K., Bellingham, P.J. & Burrows, L. 2001. Managing biodiversity information: development of the National Vegetation Survey Databank. *New Zealand Journal of Ecology* 25: 1–17.
- Wiser, S.K., Drake, D.R., Burrows, L.E. & Sykes, W.R. 2002. The potential for long-term persistence of forest fragments on Tongatapu, a large island in western Polynesia. *Journal of Biogeography* 29: 767–787.
- Wiser, S.K., Hurst, J.M., Allen, R.B. & Wright, E.F. 2011. New Zealand's forest and shrubland communities: a classification based on a nationally representative plot network. *Applied Vegetation Science* 14: 506–523.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Relating fuzzy vegetation classifications.

Appendix S2. Workflow diagram for the analyses at the alliance level.

Appendix S3. Changes in distance between the Manhattan distance and Chord distance.

Appendix S4. Membership effects of changing the cluster model.

Appendix S5. Characterization of new alliances and associations.

Appendix S6. Characterization of associations related to alliances defined in Wiser et al. (2011).

[Correction added after online publication 12 July 2012 – on line 516, in Appendix S5, '(Waipoua Forest, Puketi, Waihaha)' was changed to '(Waipoua Forest, Puketi, Waikere)']

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.