



**Irish Vegetation Classification
Technical Progress Report No. 1**

Prepared by

Philip M. Perrin

BEC Consultants Ltd

11th December 2015



**An Roinn
Ealaíon, Oidhreachta agus Gaeltachta
Department of
Arts, Heritage and the Gaeltacht**

1 Scope

This document reports on progress made during the first phase of the development of the Irish Vegetation Classification (IVC). Work was conducted by BEC Consultants Ltd under contract to the National Biodiversity Data Centre between July and December 2015. The contract specified five tasks which are summarised here:

1. Establish the basic objective, principles and methods for production of the IVC
2. Produce a hierarchical framework for the IVC
3. Incorporate data from the National Vegetation Database (NVD) and available data on High Nature Value (HNV) farmland into the classification produced by the Irish Semi-natural Grasslands Survey (ISGS)
4. Develop web pages based on the revised grassland classification
5. Develop a web application to allow users to statistically classify their own data

Work conducted to complete these five tasks is reported respectively in sections 2 to 6 of this report, whilst section 7 makes recommendations for the second and subsequent phases of the project.

2 Objective, principles and methods

2.1 Objective

The overall objective for the project is as follows:

“The Irish Vegetation Classification (IVC) aims to classify, describe and map in detail all aspects of natural and semi-natural vegetation in Ireland within a single, unified framework.”

2.2 Principles

In achieving the overall objective the IVC will be guided by the following six principles:

1. **It will be statistically-based and validated.** A classification will be produced using a range of appropriate contemporary statistical methods and quantitative vegetation data from Ireland. Validation methods will be used to ensure that proposed categories are statistically robust. It will not seek to statistically reproduce categories of existing classification systems.
2. **It will be a vegetation classification.** A classification of vegetation communities will be produced, not a habitat classification, thus the fundamental categories will be produced on the basis of floristic data analysis only. Environmental data (e.g. soil depth, altitude, inundation frequency) and management information (e.g. mowing regime, peat extraction) will be used *post hoc* where available to interpret categories and provide ecological meaning.
3. **It will recognise the vegetation continuum.** Plant species respond individually to changes along environmental gradients, thus vegetation varies continuously across

space and time. For practical purposes, a vegetation classification (which is an inherently artificial framework) seeks to identify the most frequently repeating combinations of plant species as communities. However, it should preferably also recognise that a proportion of vegetation will always be transitional between these communities.

4. **It will be user-friendly.** A classification will be produced which can be readily and consistently applied by the range of potential users. These will include field surveyors, scientific officers, environmental managers and planners, policy makers and academics.
5. **It will be hierarchical.** A classification will be produced with categories organised in a hierarchy. Categories at lower levels will be combined at high levels to produce categories of increasingly broader scope. A hierarchy will facilitate application of the classification at different scales and for a variety of purposes.
6. **It will be expandable and updatable.** A classification is only as good as the data on which it is based. To remain relevant, the classification should be updated and expanded as new data become available. This is particularly relevant when a phased approach to the development of the classification is likely. Web-basing of the classification will facilitate making and logging such changes and their immediate dissemination.

2.3 Analysis methods

In line with the guiding principles, the methodology chosen to update the ISGS classification, and indeed to expand the IVC to other habitats during future phases, is based on the approach described and exemplified by De Cáceres *et al.* (2010) and Wiser & De Cáceres (2013). Two key elements of this approach are fuzzy clustering and noise clustering. The concepts behind these methods are presented here, whilst their application to Irish grassland data is covered in section 4. During the ISGS, statistical comparisons of several classification methods were made and fuzzy clustering was the top ranked method.

Fuzzy clustering. When partitioning vegetation datasets, fuzzy set theory (Zadeh 1965) is an attractive concept to apply as it recognises that real world classes of objects may not have precise membership criteria. In 'hard' or 'crisp' partitioning algorithms based on classical set theory, objects have binary membership, that is they belong (1) to one class and do not belong (0) to all other classes. In 'fuzzy' classifications each object has a probability (or goodness of fit) from 0 to 1 of belonging to each class, the sum of these probabilities being 1. Hence, this approach facilitates the reality that vegetation samples will often be transitional in nature between perceived vegetation types (De Cáceres *et al.* 2010).

Fuzzy set theory can be employed when defining new vegetation types by excluding transitional plots (plots with a maximal membership less than a threshold α) from the calculations of synoptic floristic and environmental data; this is a means of increasing the distinctiveness and cohesiveness of the defined vegetation types (De Cáceres *et al.* 2010). Transitional plots can still be classified according to the vegetation type for which they have maximal membership if desired. The user-defined fuzziness co-efficient, m , influences the proportion of plots that will be deemed transitional.

Noise clustering. A modification of fuzzy clustering involves an additional cluster termed the noise class. This class captures any plots that lie further than a pre-specified distance δ from the centroids of the true vegetation clusters (Wiser & De Cáceres 2013). These plots are outliers within the dataset, but could be used to define further vegetation communities if more data

became available. This facilitates expansion and updating of the classification as required by the guiding principles.

Following Wiser & De Cáceres (2013), three types of plots can be defined based on the results of the fuzzy noise clustering (Table 1).

Table 1. Categorising types of plots using fuzzy membership results from noise clustering analysis (after Wiser & De Cáceres 2013).

Plot type	Definition
Assigned	The plot has membership ≥ 0.5 for one of the vegetation communities and therefore relates to the core definition of that vegetation community.
Unassigned	The plot has membership ≥ 0.5 for the noise class and is poorly represented by the current classification scheme.
Transitional	The plot has membership < 0.5 for all vegetation communities and for the noise class. It falls within the scope of the current classification scheme but does not relate to the core definition of any of the vegetation communities.

2.4 Presentation methods

The following methods will be used when describing plant communities.

Naming of plant communities. Communities are to be named in a simple style similar to that of Rodwell (1992), using the scientific names of two (or, less frequently, one or three) of the most frequent or indicative species. Each community will also have an alternative title using species common names. Groups and sub-communities (see section 3) will be named in a similar fashion. As stressed by Rodwell (1992), there is more to a community than just the species in the names and it should always be remembered that these species may be absent from any given sample assigned to that community. Each community will be given a unique code reflecting its division and group (again see section 3).

Synoptic tables. Data describing the frequency and abundance of species within each community will be presented as a synoptic table. Frequency is the percentage of plots in which each species occurs, irrespective of how much is present, while abundance refers to extent of cover of a species irrespective of its frequency (Rodwell 1992). Frequency will be indicated by Roman numerals, where I = 0.1-20.0%, II = 20.1-40.0%, III = 40.1-60.0%, IV = 60.1-80.0% and V = 80.1-100.0%. The community descriptions will follow the phytosociological convention of referring to species with frequencies of IV or V as 'constants'. Following Stevens *et al.* (2010), the minimum, median and maximum cover values for each species will be presented, using the Domin scale for simplicity. The forty most frequent species will be listed in descending order.

Distribution maps. Static maps displaying the distribution of plots at the hectad level will be presented for each community. The size of symbols will indicate the number of records within each hectad. Colour-coding of symbols will indicate the time period within which the most recent record was made (pre-1971, 1971-1985, 1986-2000 or 2001-2015). This will highlight areas where there is a lack of recent data and possibly areas from which vegetation communities have been lost. The following data will be displayed alongside each map: (i) the number of clearly assigned and transitional records; (ii) the number of mapped records in each time period, (iii) the total number of hectads in which the community has been recorded, divided by the time period in which the most recent record was made; and (iv) the number of hectads in which the community was recorded in each time period. There is a possibility in the

future for these static maps to be complemented by dynamic maps such as are used by the NBDC for mapping species (<http://maps.biodiversityireland.ie>).

Environmental data. The IVC will be integrating a large number of floristic datasets with considerable variation in the range and accuracy of the accompanying environmental data. As a result it will not be possible to present a standard set of environmental variables for each community. However, where environmental data do exist for a substantial subset of the records within a community, an average measure of these data will be included in the ecological description together with the sample size. It must be stressed that these average measures will be indicative only, as the subset of data on which they are based may not be an unbiased sample of all the records for that community.

Proxy environmental variables will, however, be calculated for each community using the forty species presented in the synoptic tables. Mean cover-weighted scores will be calculated for Light, Reaction, Wetness, Fertility and Salinity using the Ellenberg values presented in Hill *et al.* (2004) for vascular plants and Hill *et al.* (2007) for bryophytes.

Affinities. The relationship of each community with the Heritage Council's 'Guide to habitats in Ireland' (Fossitt 2000, hereafter GHI) and EU Habitats Directive Annex I habitat categories will be presented. Where reliable data of these types are available for a significant subset of the plots clearly assigned to a community, the percentage of plots within the subset matching these categories will be presented together with the size of the subset.

Correspondence with the communities or sub-communities of the British NVC (e.g. Rodwell 1992), as defined by the *Modular Analysis of Vegetation Information System* software (MAVIS; Centre for Ecology and Hydrology) will be listed. This will typically be the community or sub-community with the highest goodness-of-fit score. In cases where the top match is not felt to be the most informative, a lower scoring match will also be shown.

Subjective affinities with phytosociological alliances of the Zürich-Montpellier school (for description of which see Kent 2012) will be given following the conspectus presented by Rodwell *et al.* (2002) with guidance from Rodwell (2000). The closest EUNIS Habitat Classification 200711 category will be presented (<http://eunis.eea.europa.eu/habitats-code-browser.jsp>).

3 Hierarchical framework

A simple hierarchical framework for the IVC has been constructed (Fig. 1). At this stage of the project the framework is of value as it will facilitate planning of future phases. Ultimately, however, the aim of this hierarchy will be to simply organise the defined plant communities in an intuitive structure and provide options for users in terms of the level of detail used in recording or mapping, as required by guiding principle 5. The purpose of the hierarchy is not to facilitate integration with any existing schemes. With this said, the higher levels are loosely based on the concepts of Faber-Langendoen *et al.* (2012) as used in the [U.S. National Vegetation Classification](#) and the [Canadian Vegetation Classification](#). In comparison to these two systems, the IVC hierarchy is relatively simple due to the limited variation in climate and environmental gradients in Ireland and absence of levels differentiating vegetation at a continental or regional

5

← →					
Physiognomy and ecology			Floristics		
Class	Formation	Division	Group	Community	Sub-community
Tree Vegetation (Mesomorphic)	Forest Vegetation	Woodland	e.g. GL3 <i>Dactylis glomerata</i> – <i>Centaurea nigra</i> group (Cock's-foot – Knapweed group)	e.g. GL3A <i>Briza media</i> – <i>Thymus polytrichus</i> grassland (Quaking-grass – Wild Thyme grassland)	e.g. GL3aai <i>Briza media</i> – <i>Thymus polytrichus</i> grassland <i>Sesleria caerulea</i> – <i>Tortella tortuosa</i> sub-community (Quaking-grass – Wild Thyme grassland Blue Moor-grass – Frizzled Crisp-moss sub-community)
		Plantation Forests			
		Scrub			
		Hedgerows			
Shrub and Herb Vegetation (Mesomorphic and Cryomorphic)	Inland Shrub and Herb Vegetation	Grasslands →			
		Heaths			
		Wetlands			
		Montane Communities			
		Herb Fringe Communities			
		Ruderal Communities			
	Coastal Shrub and Herb Vegetation	Salt-marsh			
		Sand-dunes			
Aquatic Vegetation (Hydromorphic)	Freshwater Vegetation	Lakes			
		Watercourses			
	Saltwater Vegetation	Marine Waters			
Rock Vegetation (Cryptogamic and Open Mesomorphic)	Inland Rock Vegetation	Siliceous Rocks			
		Calcareous Rocks			
	Coastal Rock Vegetation	Coastal Rocks			

Figure 1. Hierarchical structure of the IVC with examples of the lower levels from the grasslands division.

scale. Also, it is proposed that widespread types of cultural vegetation (e.g. arable weed communities, coniferous plantations, improved grassland, invasive plant communities) are incorporated into the same framework with natural and semi-natural vegetation, rather than held in a parallel framework.

The upper three levels are based on physiognomy and ecology, whereas the lower three levels are based on floristic data as set out in the guiding principles. Class, the highest level of the hierarchy, divides vegetation based on dominant growth forms. The Formation level essentially divides the vegetation into coastal and inland variations as these represent Ireland's major biogeographical zones. The Division level uses dominant growth forms and broad sets of diagnostic species. Categories at this level will be familiar as broad habitat types comparable to the habitat subgroups (or level 2) of the GHI. The Group level represents broad divisions of vegetation reflecting significant differences in the major environmental gradients such as pH and soil moisture. The Community level is the fundamental unit of the IVC and divisions here reflect relatively specific differences in overall floristic composition. The final level, Sub-community divides communities on specific indicator species or subtle differences in floristic composition; not all communities will have sub-communities. As mentioned in subsection 2.4 and demonstrated in Fig. 1, the lower levels of the hierarchy are assigned a code.

4 Grassland classification

4.1 Scope of work

The task of incorporating additional data into the ISGS classification (introduced in the next subsection) comprised a number of stages which are summarised here to give an overview of work conducted. These stages are then dealt with in more detail in the subsequent subsections, with the workflow presented in Figs. 2a-d.

1. Transfer the ISGS classification into a noise clustering framework.
2. Extract grassland data from other available datasets.
3. Assess the quality and suitability of these data for reviewing and redefining the ISGS classification and extract the subset of suitable data.
4. Review the ISGS classification using this data subset in terms of gaps and resolution.
5. Redefine the ISGS communities statistically and descriptively based on the review.
6. Assign all grassland data (including those rejected in stage 3) to the new IVC classification.
7. Produce distribution maps for communities based on all assigned plots.

4.2 ISGS classification

The analysis methodology used to create the ISGS classification is presented here in a slightly summarised format so that the starting point for the IVC is clearly defined and later decisions are placed in context. For full details and an account of the survey methodology see O'Neill *et al.* (2013).

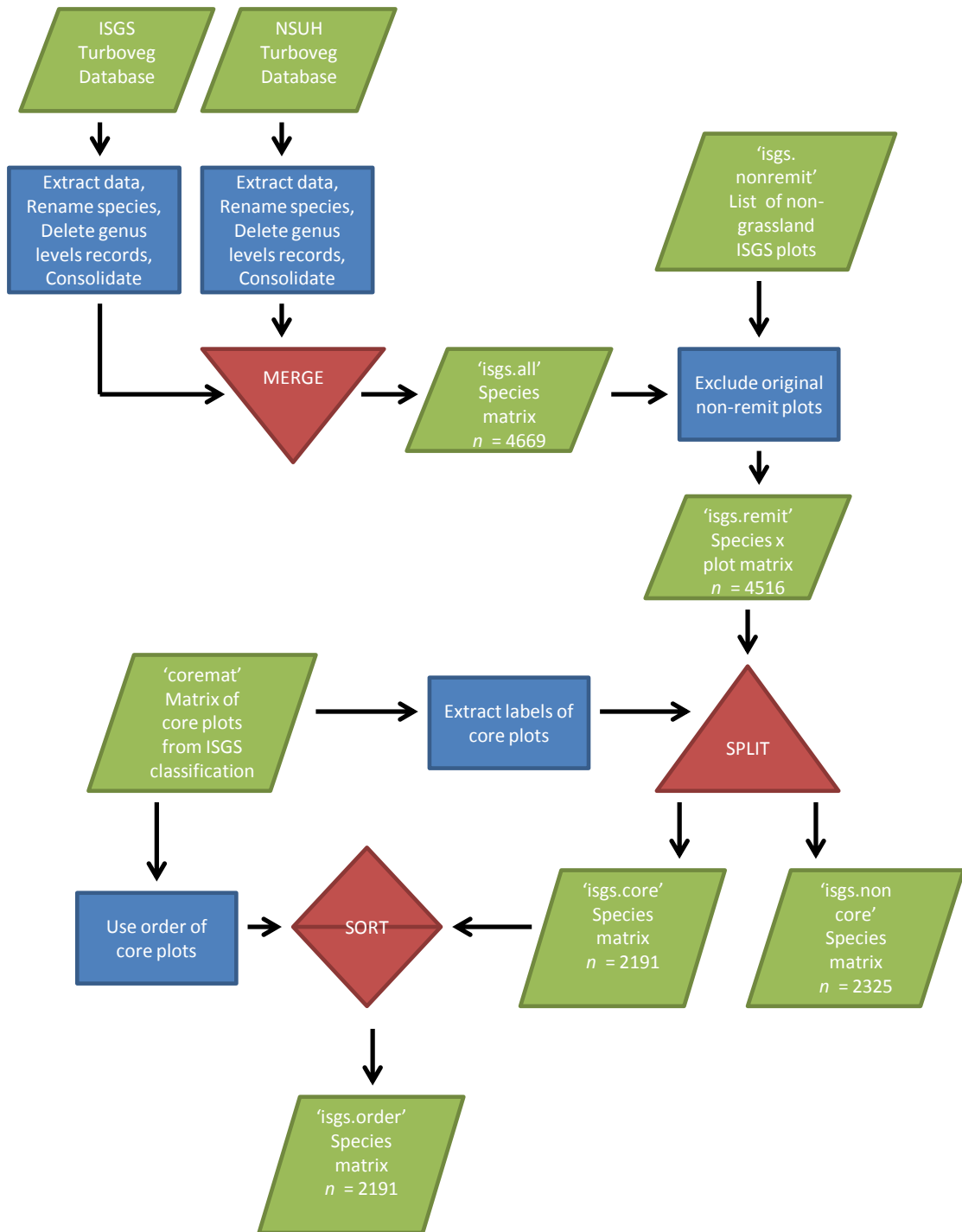


Figure 2a. Workflow showing how data from the ISGS and the NSUH were compiled and prepared. Green boxes indicate datasets, red boxes indicate data manipulation and blue boxes indicate other processes.

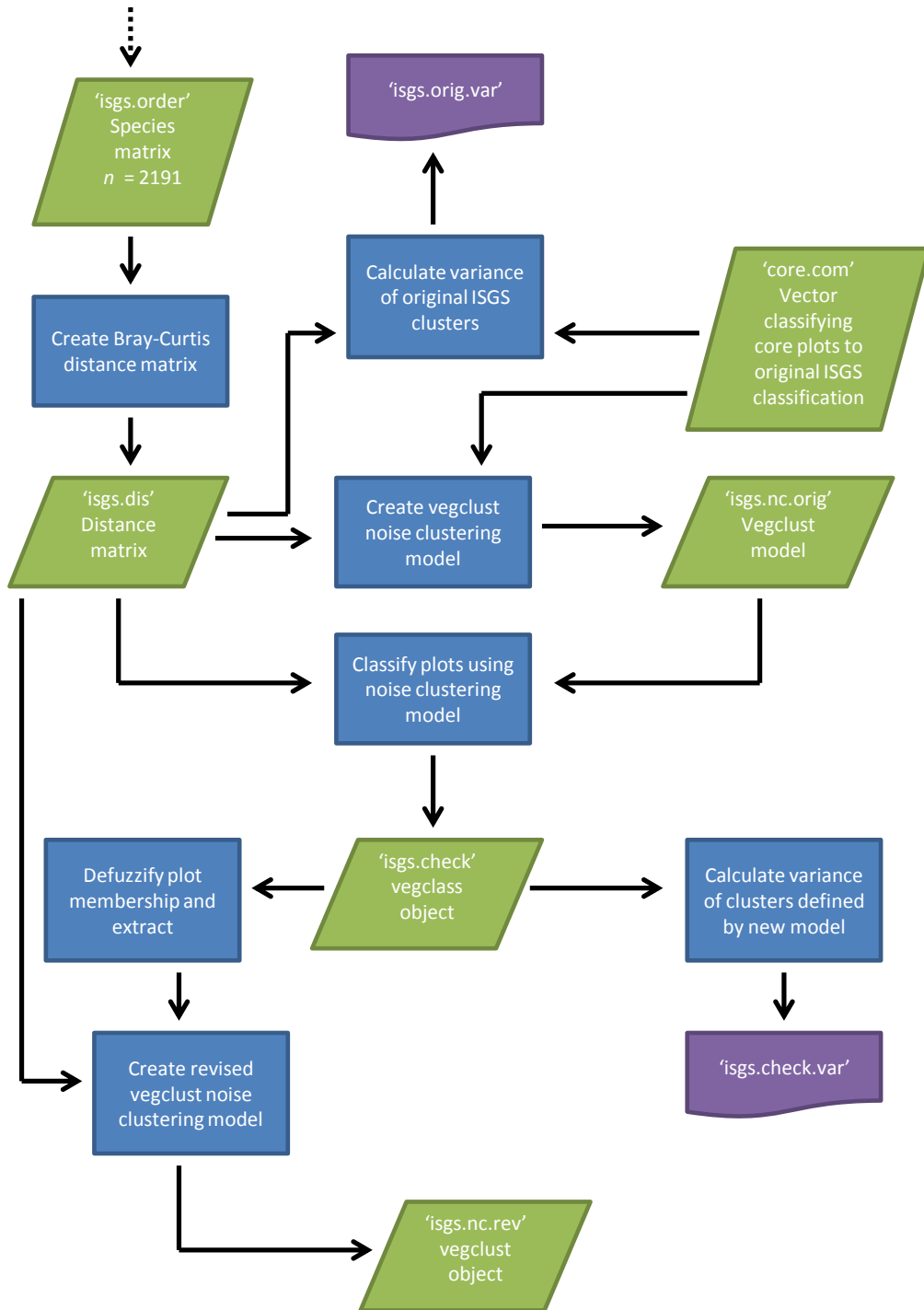


Figure 2b. Workflow showing how the ISGS classification was transferred into a noise clustering framework. Green boxes indicate datasets, blue boxes indicate processes, purple boxes indicate outputs and dotted lines where datasets were produced at an earlier stage.

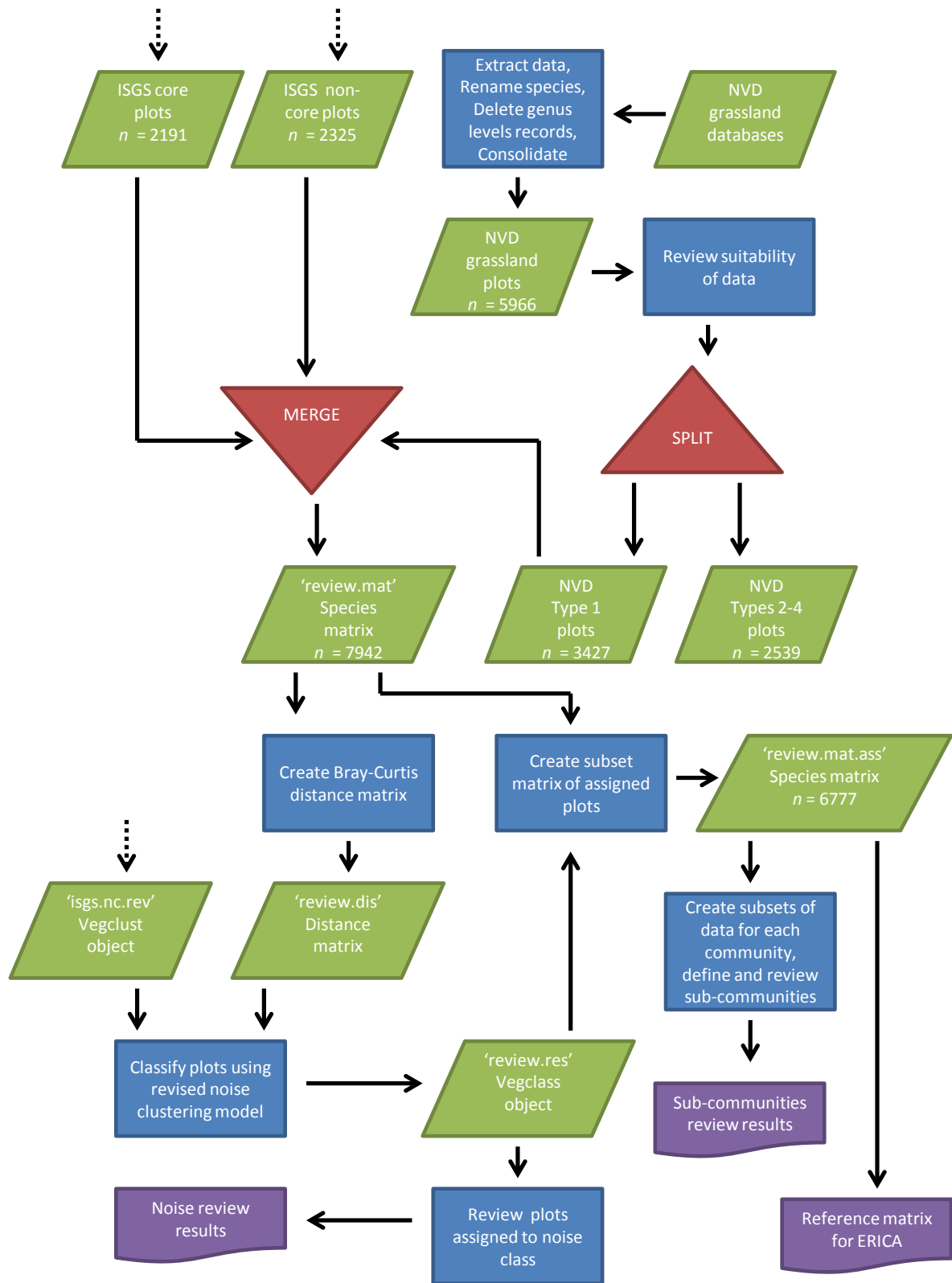


Figure 2c. Workflow showing the data review processes. Green boxes indicate datasets, red boxes indicate data manipulation, blue boxes indicate other processes, purple boxes indicate outputs and dotted lines indicate where datasets were produced at an earlier stage.

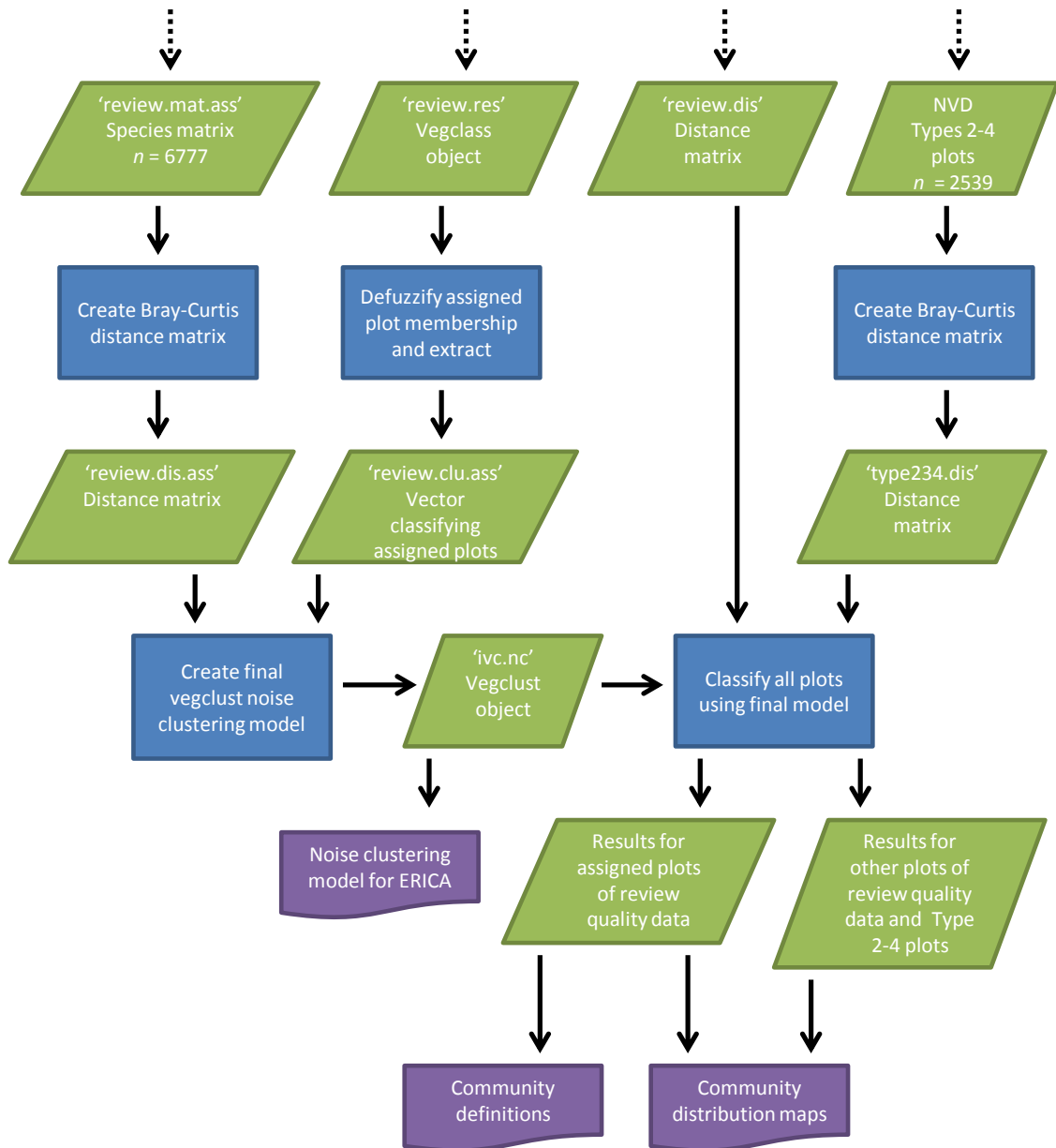


Figure 2d. Workflow showing how the final assignments of plots were made. Green boxes indicate datasets, blue boxes indicate processes, purple boxes indicate outputs and dotted lines indicate where datasets were produced at an earlier stage.

Data preparation. The combined ISGS 2007-2012 dataset contains 4,633 relevés. Only relevés classified by the GHI as GS (semi-natural grassland), GA (GSi, semi-improved grassland) or GM (freshwater marsh) were included in the analysis; 153 relevés classified (or reclassified following review) as HH (heath), PF (fen/flush), FS (swamp), CM (salt marsh) or CD (sand dunes) were excluded. An additional 36 relevés classified as GS from the National Survey of Upland Habitats (NSUH) were included to bolster coverage of the dataset in upland areas (it is implicit henceforth that when referring to the ISGS dataset it means the combination of ISGS and NSUH data).

Nomenclature for the analysis followed the Ireland2008 species checklist with some amendments. Records at the sub-species or variant level were combined at the species level. Species records that had only been identified to the genus level were excluded in most cases as they may be amalgams of species with markedly different ecological preferences and therefore misleading to utilise (e.g. *Carex* sp.). An exception was made for *Hieracium* spp. due to recognised identification issues with microspecies. All records for *Euphrasia* spp. were combined as *Euphrasia officinalis* agg. with the exception of *Euphrasia salisburgensis*. Due to taxonomic or identification issues, the following pairs of species were combined: *Agrostis canina* / *A. vinealis*, *Juncus acutiflorus* / *J. articulatus*, *Poa humilis* / *P. pratensis*, *Thuidium tamariscinum* / *T. delicatulum*, *Chiloscyphus pallescens* / *C. polyanthos*, *Fissidens adianthoides* / *F. dubius*. All records of *Viola* species were interpreted as *Viola riviniana*, as it is highly unlikely that *Viola reichenbachiana* would be encountered in grassland and other *Viola* species are distinctive. *Rubus fruticosus* and *Taraxacum officinale* were recorded in the field to the aggregate by the ISGS.

As the ISGS relevé data were recorded using the Domin scale, they were converted to percentages using mid-range values (Table 2a); mean values needed for the analysis cannot be calculated directly from a non-linear scale. Data from the NSUH were recorded as percentages in the field. Where species records were excluded due to identification only to the genus level, the relevé was excluded if these records totalled 5% or more cover, due to the lack of data on a significant portion of the vegetation. This resulted in the removal of 39 relevés.

Species recorded in fewer than 20 relevés in the combined dataset were excluded to reduce noise. Multivariate outlier analysis was used to examine the dataset in PC-Ord 6.09 (MjM Software, Gleneden Beach, Oregon). The mean distance of each sample from each other sample was calculated using Quantitative Sørensen (Bray-Curtis) as the distance measure. A threshold of three standard deviations of the grand mean for all distances between samples was used.

Eight samples were flagged as outliers but all were deemed to be within the remit of the analysis; several were samples from rank grasslands. These stages of data preparation yielded a final data matrix of 4,477 plots and 299 species/taxa, which were square-root transformed to down-weight the influence of abundant species.

Data analysis. The aim of the analysis was to sort the relevés to produce a two-tier classification, with broad groups divided into a number of communities at a level of resolution akin to the vegetation communities of the British NVC and the association level of the Zürich-Montpellier school.

Table 2. Conversion of cover data on ordinal scales to percentage cover using mid-range values: (a) the Domin scale, (b) modified version 3 of the Domin scale, (c) versions of the Braun-Blanquet scale.

(a)			(b)	
Domin (original)	Cover range (%)	Converted cover (%)	Domin (modified 3)	Converted cover (%)*
10	91-100	96	10	96
9	76-90	83	9	83
8	51-75	63	8	63
7	34-50	42	7	38
6	26-33	30	6	19
5	11-25	18	5	9
4	5-10	8	4	5
3	1-4	3	3	3
2	<1	0.5	2	1
1	<1	0.3	1	0.5
+	<1	0.1		

* Information of the cover ranges of this scale were not available, therefore conversion values proposed by the NBDC were used.

(c)				
Braun-Blanquet (original)	Braun-Blanquet (modified 1)	Braun-Blanquet (modified 2)	Cover range (%)	Converted cover (%)
5	6	9	76-100	88
4	5	8	51-75	63
3	4	7	26-50	38
.	.	6	12.6-25	19
.	.	5	5-12.5	9
2	3	.	5-25	15
.	2	1-4	<5	3
1	.	.	1-5	3
.	1	.	Single, isolated small plant	0.5
+	.	.	<1	0.5
r	.	.	<1	0.1

The data matrix of n samples \times p species was used to calculate an $n \times n$ dissimilarity matrix defining the dissimilarity between each pair of samples. Quantitative Sørensen (Bray-Curtis) dissimilarity was selected as the distance measure, as it has been shown to be one of the most effective measures for ecological community analysis, being less prone to exaggerating the influence of outliers and retaining greater sensitivity with heterogeneous datasets (McCune & Grace 2002).

Six classification methods, hierarchical agglomerative cluster analysis (HAC), hierarchical divisive cluster analysis (DIANA), two-way indicator species analysis (TWINSPAN), partitioning around medoids (PAM), fuzzy analysis (FANNY) and fuzzy c-medoids (FCMdd) were extensively tested with the dissimilarity matrix by comparing a variety of cluster validation measures over a range of cluster levels to see which performed best. Analyses were all conducted in the R statistical environment, except the TWINSPAN analysis which was conducted in PC-Ord 6.09. FANNY was selected as the best performing method. This non-hierarchical method is fully described in Kaufman & Rousseeuw (1990) and was implemented using function `fanny` in package `cluster`. For all analyses, `fanny` was run with a fuzziness co-efficient (m) of 1.1 (as trials indicated that higher values produced too much fuzziness) and with the number of maximum iterations set to 10,000.

FANNY was conducted over a range of cluster levels (two to six). At each level, transitional samples ($\alpha = 0.5$) were excluded and the remaining subset of the data was reanalysed to produce a crisp classification. Following examination of constancy (species frequency) tables, the four-group level was selected (subset size = 3,507, number of intermediates = 970). At the three-group level, wet and dry acidic grassland was grouped together with marsh, while at the five-group level, dry semi-improved grassland was separated out; neither scenario was deemed desirable. While cluster validation measures can be used to select the optimal level of clustering (number of groups), expert judgement was preferred to ensure that clusters were ecologically meaningful and would be intuitive for fieldworkers.

Separate analyses were then conducted on the data assigned to each of the four groups across a range of cluster levels to define the communities. Again, the procedure was analysis, exclusion of intermediates and reanalysis to produce a crisp classification. This ultimately resulted in a classification composed of four groups divided into nineteen communities and based on 2,191 'core' relevés, with the other 2,286 relevés deemed transitional.

Cluster validation. The validity of the four-group structure was assessed by silhouette analysis (also termed average silhouette width) in R using the function `silhouette` in package `cluster`¹. It was initially proposed by Rousseeuw (1987). The silhouette width of a sample is calculated from the average dissimilarity of that sample to all samples in the same cluster and from the average dissimilarity of that sample to all samples in the next most similar cluster (Maechler *et al.* 2013). Positive values indicate a good fit and negative values indicate that a sample would fit better elsewhere. The mean silhouette width for a cluster indicates the quality of that cluster, and the global mean silhouette width (or silhouette coefficient) of all samples indicates the quality of the classification (Peet & Roberts 2012). This value is ideally maximised. The silhouette plot for the four-cluster solution (Fig. 3) indicated a very small number of misclassified plots.

Assignment of excluded plots. Relevés not used in defining the classification were assigned statistically to communities in R on the basis of best fit. As no assignment function for FANNY had been implemented, the fuzzy c-medoids (FCMdd) algorithm described by Krishnapuram *et al.* (1999) was used through the `vegclass` function in package `vegclust` (De Cáceres *et al.* 2010) with a fuzziness co-efficient (m) of 1.1. This method is similar in principle to FANNY and came second in the initial tests with different classification methods². The 2,286 transitional relevés were assigned to the community for which they scored maximum membership probability. Of the 39 relevés originally excluded from the analysis due to genus level records, a more cautious approach was taken and only the 16 relevés with maximum membership probability >0.5 were assigned.

¹ This corrects the erroneous statement in O'Neill *et al.* (2013) that function `clustvar` in package `vegclust` was used for this analysis.

² Note, however, that this assignment procedure was not optimally applied as a Euclidean distance was used rather than the more suitable Quantitative Sørensen (Bray-Curtis) dissimilarity measure.

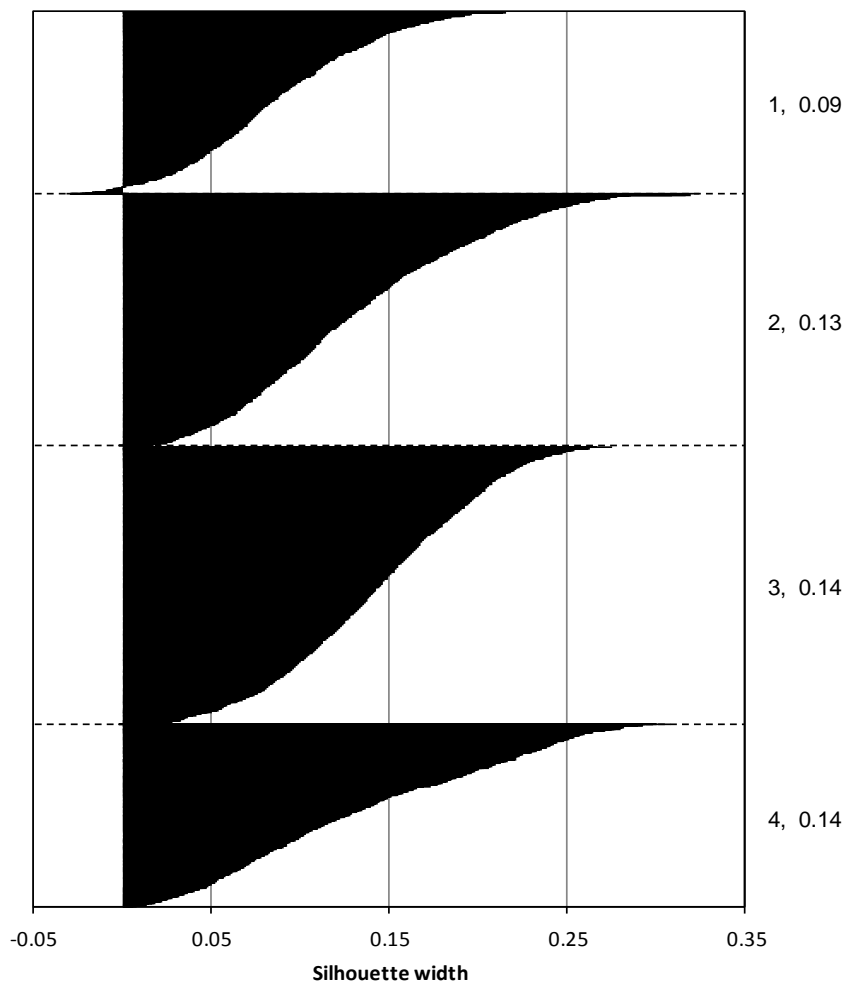


Figure 3. Silhouette plot of grasslands dataset (ISGS and NSUH relevés) with four clusters following fuzzy analysis with exclusion of intermediates ($n = 3507$). Numbers on the right indicate cluster number and mean silhouette width of cluster. Overall mean silhouette width = 0.13.

4.3 Transferring the ISGS classification

Whilst the fuzzy clustering method used to create the ISGS classification had proven itself in the statistical trials that were conducted at that time and in the cluster validation stage, it was decided to transfer the classification to a (conceptually better) noise clustering framework as this would allow the IVC to take advantage of the integrated identification of outliers and the procedure for updating classifications proposed by Wiser & De Cáceres (2013).

Some minor changes were made to the ISGS dataset after the initial analysis but prior to submission of deliverables for that project and hence are present in the final dataset. These were (i) changes to the GHI habitat type for three relevés, and (ii) amendments to a small number of bryophyte identifications following feedback from experts on critical voucher specimens. The changes to the GHI habitat types impacted on whether these relevés came within the remit of the analysis. However, as these were borderline cases, it was decided for the sake of simplicity to proceed with the IVC analysis using the same set of relevés used previously. The updated bryophyte data were used without issue.

Some changes were also made at this stage to the structuring of the ISGS species data. *Euphrasia salisburgensis* was absorbed within *Euphrasia officinalis* agg. *Hieracium* spp. was renamed as *Hieracium* agg. The combination of *Poa pratensis* and *Poa humilis* was formalised as *Poa pratensis* agg. following the advice of Rich & Jermy (1998). The couplings of *Juncus acutiflorus* / *J. articulatus*, *Thuidium tamariscinum* / *T. delicatulum*, *Chiloscyphus pallescens* / *C. polyanthus* and *Fissidens adianthoides* / *F. dubius*, which were concerned with identification issues rather than recognised taxonomic issues, were dissolved as it was deemed that this would set an unwanted data standard for the IVC moving forward.

Analysis at this stage was concerned with the 2,191 ‘core’ relevés from the ISGS which defined the classification. A dissimilarity matrix constructed using the Quantitative Sørensen (Bray-Curtis) measure was combined with the community identities determined by the ISGS analysis to produce a noise clustering framework for the data using functions `as.vegclust` and `vegclass` from package `vegclust` (Fig. 2). A fuzziness co-efficient (m) of 1.1 was again used. Following preliminary analysis, δ for the model was set as 0.75; higher values of δ resulted in poor detection of non-grassland samples as outliers.

The size of clusters representing the nineteen ISGS communities and the variance in community composition (function `clustvar`, package `vegclust`) within the plots of this set of clusters were examined (i) in the original fuzzy clustering framework and (ii) in the revised noise clustering framework (Table 3).

Converting the classification to noise clustering framework had widespread but relatively minor changes to the size of clusters and cluster variance. In their example, Wiser & De Cáceres (2013) disbanded clusters which, following ‘casting’ of an existing classification into the noise clustering framework, had low numbers of plot records (< 20 plots) and high variance (> 0.6). According to these guidelines, there were grounds for retaining all of the ISGS communities.

Table 3. Size of clusters and cluster variance for the communities of the ISGS classification in the original fuzzy clustering framework and revised noise clustering framework.

ISGS community	Number of plots (original)	Number of plots (revised)	Cluster variance (original)	Cluster variance (revised)
1a <i>Juncus acutiflorus/articulatus</i> – <i>Holcus lanatus</i>	84	93	0.17	0.17
1b <i>Agrostis stolonifera</i> – <i>Filipendula ulmaria</i>	91	96	0.23	0.22
1c <i>Molinia caerulea</i> – <i>Succisa pratensis</i>	50	52	0.17	0.17
1d <i>Molinia caerulea</i> – <i>Potentilla erecta</i>	61	70	0.19	0.19
1e <i>Juncus acutiflorus/articulatus</i> – <i>R. squarrosus</i>	62	66	0.14	0.14
2a <i>Agrostis stolonifera</i> – <i>Ranunculus repens</i>	142	135	0.25	0.25
2b <i>Juncus effusus</i> – <i>Holcus lanatus</i>	206	197	0.15	0.15
2c <i>Holcus lanatus</i> – <i>Lolium perenne</i>	212	207	0.19	0.18
2d <i>Juncus effusus</i> – <i>Rumex acetosa</i>	230	231	0.16	0.16
3a <i>Briza media</i> – <i>Thymus polytrichus</i>	149	150	0.17	0.17
3b <i>Cynosurus cristatus</i> – <i>Trifolium repens</i>	60	59	0.13	0.13
3c <i>Festuca rubra</i> – <i>Plantago lanceolata</i>	70	70	0.16	0.16
3d <i>Cynosurus cristatus</i> – <i>Trifolium pratense</i>	69	79	0.12	0.13
3e <i>Festuca rubra</i> – <i>Rhinanthus minor</i>	64	65	0.12	0.12
3f <i>Festuca rubra</i> – <i>Lotus corniculatus</i>	69	71	0.19	0.19
4a <i>Agrostis capillaris</i> – <i>Trifolium repens</i>	163	155	0.15	0.15
4b <i>Nardus stricta</i> – <i>Festuca ovina</i>	160	161	0.17	0.17
4c <i>Agrostis capillaris</i> – <i>Festuca rubra</i>	114	106	0.18	0.17
4d <i>Agrostis canina/vinealis</i> – <i>Carex echinata</i>	135	128	0.16	0.16

4.4 Extracting data from NVD datasets

The NVD was screened for other relevant data with the aid of the habitat type notes in the NVD bibliography dated 2013. Within most of the constituent databases of the NVD, plots³ have been classified using the GHI, either by NBDC staff upon entry of the data into the NVD or by the original authors. The classification of plots categorised as GS (semi-natural grassland), GA (improved grassland) or GM (freshwater marsh) was reviewed to check that they did actually fall within the remit of this phase of the project. Some of these plots were deemed to be incorrectly classified (e.g. plots with high cover of *Calluna vulgaris*, *Dryas octopetala* or bare rock), but the majority were selected. Plots categorised as FL6 (turloughs) were also reviewed as this habitat can contain vegetation communities that may also be classified as GS, GA or GM and plots of this nature were selected. It was beyond the resources of this phase to review the classification of plots in other GHI categories. Within databases where plots had not been assigned to GHI categories, notably the dataset of Austin O’Sullivan (database code NPWS057), all plots were reviewed to select those related to these three GHI categories. Some plots were rejected at this stage as they had been recorded on a presence/absence basis rather than a quantitative scale. The full list of databases from which data were selected is presented in Table 4.

Enquiries were made into the availability of grassland data from HNV farmland collected by Dr Caroline Sullivan. However, it transpires that these data were actually recorded at a field level rather than a plot level and therefore were not suitable for inclusion in the classification revision process.

Table 4. NVD datasets from which data were extracted, with cover scales and number of remit plots classified by Type.

Dataset reference	NVD code	Cover scale	Number of remit plots	Number of plots of each Type			
				1	2	3	4
N. McGough	NPWS001	Braun-Blanquet	60	44	16	0	0
T. Curtis 1	NPWS002	Braun-Blanquet	7	0	0	7	0
Protected flora	NPWS003	Braun-Blanquet	29	0	5	24	0
Threatened flora	NPWS004	Domin/Braun-Blanquet	6	0	0	6	0
Scarce flora	NPWS005	Braun-Blanquet	30	0	4	26	0
A. Bleasdale 3	NPWS008	Domin/Braun-Blanquet	6	5	1	0	0
N. Lockhart 2	NPWS011	Braun-Blanquet	34	0	2	32	0
R. Ivimey-Cook	INDEP001	Domin	67	65	1	0	1
J. Ryan	NPWS014	Braun-Blanquet	18	0	0	18	0
E. Regan	NUIG001	Percentage	63	0	14	49	0
G. Tobin	NPWS015	Percentage	25	0	5	20	0
R. Dwyer	NPWS016	Percentage	5	5	0	0	0
C. Farrell	NPWS017	Braun-Blanquet	14	14	0	0	0
Natura	NPWS019	Braun-Blanquet	10	0	0	10	0
M. Loftus	NPWS022	Domin	2	2	0	0	0
L. Weekes	NPWS024	Braun-Blanquet	19	19	0	0	0

³ The general term ‘plot’ is used from this point forward as not all records within the NVD are relevés *sensu stricto*.

Table 4. continued.

Dataset reference	NVD code	Cover scale	Number of remit plots	Number of plots of each Type			
				1	2	3	4
A. Bleasdale 4	NPWS027	Domin	4	4	0	0	0
A. Bleasdale 5	NPWS031	Braun-Blanquet	3	3	0	0	0
A. Bleasdale 2	NUIG004	Braun-Blanquet	132	132	0	0	0
S. Parr	INDEP006	Domin	132	28	104	0	0
J. Moran 1	INDEP007	Domin	42	0	25	17	0
J. Moran 2	NUIG006	Percentage	135	135	0	0	0
M. Telford	TCD002	Domin	12	12	0	0	0
A. Bleasdale 1	INDEP014	Domin	3	3	0	0	0
B. Macgowran	NUIG009	Braun-Blanquet	84	84	0	0	0
B. Dunford	UCD014	Modified Domin 3	1010	149	861	0	0
T. Constable	NPWS034	Braun-Blanquet	25	14	11	0	0
S. Keane	NUIG013	Braun-Blanquet	28	28	0	0	0
D. Lynn	INDEP021	Percentage	10	0	6	4	0
G. O'Donovan 1	INDEP023	Domin	176	124	52	0	0
N. Lockhart 1	NPWS036	Braun-Blanquet	78	24	18	36	0
G. O'Donovan 2	NPWS037	Domin	20	0	0	20	0
G. Doyle	INDEP027	Braun-Blanquet	5	0	0	5	0
C. Brennan	NUIG026	Braun-Blanquet	65	0	0	65	0
B. Ní Bhriain 1	NUIG033	Braun-Blanquet	15	0	0	15	0
S. Heery	NPWS042	Modified BB 1	122	0	4	118	0
C. Borggreve	NPWS043	Modified BB 2	18	0	0	18	0
T. Curtis 2	NPWS045	Braun-Blanquet	8	0	4	4	0
S. Murphy	NUIG036	Braun-Blanquet	22	22	0	0	0
G. Smith	TCD004	Percentage	114	111	3	0	0
J. Braun-Blanquet	NPWS058	Braun-Blanquet	101	25	6	0	70
C. Roden	NPWS048	Braun-Blanquet	3	0	0	3	0
S. Reynolds	NPWS050	Braun-Blanquet	11	2	1	8	0
J. Conaghan 2	NPWS051	Braun-Blanquet	2	0	0	1	1
E. MhicDaeid	TCD007	Braun-Blanquet	40	36	0	0	4
J. Conaghan 3	NPWS049	Braun-Blanquet	7	2	0	5	0
J. Conaghan 1	NPWS053	Braun-Blanquet	1	0	0	1	0
M. Long	TCD008	Percentage	39	0	39	0	0
A. Browne	UCD021	Braun-Blanquet	123	0	2	121	0
C. Byrne	TCD009	Domin	91	91	0	0	0
A. O'Sullivan	NPWS057	Braun-Blanquet	2788	2228	59	0	501
S. Wilson	NPWS061	Domin	101	15	86	0	0
<i>Total</i>			<i>5965</i>	<i>3426</i>	<i>1329</i>	<i>633</i>	<i>577</i>

4.5 Preparation and assessment of NVD data

The NVD data were prepared in the same manner as the ISGS data (see subsection 4.2 but also amendments in subsection 4.3 and Tables 2b and 2c). A few further minor issues became apparent when inspecting the NVD dataset. Firstly, the Ireland2008 species list lists both *Carex flava* agg. and *Carex viridula* although these are synonymous in an Irish context; records from this group were therefore combined as *Carex viridula sensu lato*. Secondly, it was also required to combine records of *Trichophorum* × *foersteri* with those of *T. germanicum* and *T. cespitosum*, as these were already combined. Thirdly, the NVD included records for *Rubus caesius*, a taxon regarded as part of *Rubus fruticosus* agg. by Stace (2010) but which is fairly easily distinguished. It was therefore decided to keep data for this species separate from the data for the rest of the aggregate in the IVC.

A more significant issue surrounded the recording of *Festuca rubra* and *Festuca ovina*. Several of the datasets in the NVD from the Burren (e.g. INDEP006, INDEP007) combined these species but the ISGS had not. A decision was required on whether data for these two species should be combined within the IVC. It was decided to keep the species separate because (i) combining the data is a workaround due to identification issues rather than a strict taxonomic issue (see section 4.3) and (ii) the distinction is ecologically important as *F. ovina* occurs in more stressful and lower nutrient conditions than *F. rubra*. Unfortunately, as a result, the aforementioned records of this combination had to be excluded as they were *de facto* *Festuca* sp. It was also decided not to aggregate sub-species of *F. ovina* and *F. rubra* into *F. ovina* agg and *F. rubra* agg. respectively at this stage, although this may need to be reviewed in the future⁴.

The suitability of the selected NVD grassland/marsh data for inclusion in the analysis was then assessed. Plots were classified according to the standard of the data as follows (see Table 3 for breakdown by dataset):

- Type 1. Plots with none of the issues described for Types 2-4.
- Type 2. Plots where excluded taxa data (e.g. *Carex* sp., *Festuca rubra/ovina*) comprised ≥ 5% cover in total.
- Type 3. Plots with an apparent significant deficiency in the recording of bryophytes, with the exception of plots categorised as Type 2; whilst recording of lower plants often varies between datasets (and habitats) in terms of both effort and accuracy, it is desirable that at least the main bryophytic components of vegetation should be recorded.
- Type 4. Plots with a recorded size < 1 m² or >25 m² and plots with no recorded size, with the exception of plots categorised as Types 2 or 3; this was to ensure that reasonably valid comparisons were made with the ISGS data, which had all been recorded from 4 m² plots.

4.6 Review of the ISGS classification

Three datasets were used to review the ISGS classification at this stage. These were (i) the ISGS core dataset ($n = 2191$), (ii) the ISGS non-core dataset ($n = 2325$), and (iii) Type 1 plots from the NVD dataset ($n = 3426$). The review was in two parts, firstly examining the need to define new grassland communities from outlier data (i.e. plots assigned to the noise cluster) and secondly

⁴ The use within the IVC of aggregates other than those mentioned hitherto is unresolved at this time as they are not characteristic taxa of grasslands.

examining the need to define sub-communities. The integrity of the existing communities themselves had already been assessed (sub-section 4.2).

Definition of new communities: The three datasets were classified using the noise clustering framework (Table 5). As would be expected, the ISGS core dataset which was used to define the classification in the first place had a low proportion of Transitional plots (1.7%) and no plots in the noise cluster. The ISGS non-core dataset had proportionately the highest number of Transitional plots (24.5%); again this would be expected given the definition of the dataset set. Of the NVD Type 1 dataset, 16.0% of plots were Transitional.

Table 5. Assignment of plots from three datasets to the ISGS classification noise clustering framework. A = number of Assigned plots, T = number of Transitional plots, U = number of Unassigned plots.

	ISGS community	ISGS core			ISGS non-core			NVD Type 1		
		A	T	U	A	T	U	A	T	U
1a	<i>Juncus acutiflorus/articulatus</i> – <i>Holcus lanatus</i>	95	0	-	97	21	-	25	13	-
1b	<i>Agrostis stolonifera</i> – <i>Filipendula ulmaria</i>	97	1	-	99	45	-	72	32	-
1c	<i>Molinia caerulea</i> – <i>Succisa pratensis</i>	49	4	-	173	36	-	69	29	-
1d	<i>Molinia caerulea</i> – <i>Potentilla erecta</i>	70	2	-	59	7	-	104	11	-
1e	<i>Juncus acutiflorus/articulatus</i> – <i>R. squarrosus</i>	66	1	-	93	41	-	79	26	-
2a	<i>Agrostis stolonifera</i> – <i>Ranunculus repens</i>	128	5	-	28	14	-	94	19	-
2b	<i>Juncus effusus</i> – <i>Holcus lanatus</i>	191	5	-	129	45	-	69	31	-
2c	<i>Holcus lanatus</i> – <i>Lolium perenne</i>	201	4	-	66	38	-	191	24	-
2d	<i>Juncus effusus</i> – <i>Rumex acetosa</i>	230	1	-	42	20	-	13	2	-
3a	<i>Briza media</i> – <i>Thymus polytrichus</i>	150	0	-	217	27	-	274	13	-
3b	<i>Cynosurus cristatus</i> – <i>Trifolium repens</i>	57	1	-	75	29	-	880	79	-
3c	<i>Festuca rubra</i> – <i>Plantago lanceolata</i>	69	2	-	157	56	-	87	47	-
3d	<i>Cynosurus cristatus</i> – <i>Trifolium pratense</i>	80	2	-	121	55	-	498	83	-
3e	<i>Festuca rubra</i> – <i>Rhinanthus minor</i>	64	1	-	54	18	-	51	29	-
3f	<i>Festuca rubra</i> – <i>Lotus corniculatus</i>	71	0	-	111	23	-	21	8	-
4a	<i>Agrostis capillaris</i> – <i>Trifolium repens</i>	151	2	-	85	34	-	147	34	-
4b	<i>Nardus stricta</i> – <i>Festuca ovina</i>	160	1	-	27	9	-	134	15	-
4c	<i>Agrostis capillaris</i> – <i>Festuca rubra</i>	104	0	-	82	30	-	32	17	-
4d	<i>Agrostis canina/vinealis</i> – <i>Carex echinata</i>	122	4	-	35	18	-	32	8	-
	Noise cluster	-	0	0		4	5	-	29	5
	Total	2155	36	0	1750	570	5	2872	549	5

More importantly, however, the overall number of Unassigned plots was very low at just 10 (0.01% of all plots). These plots included a *Poa annua* – *Plantago major* community of trampled ground, *Arrhenatherum elatius* swards and *Elytrigia repens* swards. The first of these three elements may be better dealt with later in the project with other communities of disturbed ground. The latter two elements of the noise cluster appear to represent genuine grassland communities but there are insufficient data at present to warrant statistically defining these as new communities within the noise clustering framework. Instead, they have been designated as ‘placeholder communities’. This mechanism will be used within the IVC to flag vegetation assemblages which probably warrant recognition at the community level but for which insufficient data are currently available. The first two placeholder communities are:

PH1: *Arrhenatherum elatius* grassland

PH2: *Elytrigia repens* grassland

Definition of sub-communities: The ISGS defined sub-communities for only two grassland communities (3a and 3f). These were subjective divisions of the communities based on specific indicator species and were not statistically defined. The appropriateness of defining sub-communities was statistically reviewed for all communities by running noise clustering analysis

(function `vegclustdist`, package `vegclust`, $m = 1.1$, $\delta = 0.75$, number of random starts = 10, maximum number of iterations = 100) on datasets comprising the Assigned plots (as enumerated in Table 5). For each community, the analysis was run to investigate 2 and 3 potential sub-communities.

Partitions were assessed against the criteria: i) clusters should be relatively homogeneous and distinct from one another; and ii) clusters should have ecologically interpretable patterns of species distribution (Peet & Roberts 2013). The same criteria were used by Devaney & Perrin (2015). The first criterion was assessed using silhouette analysis (see section 4.3) and partition analysis (function `partana`, package `optpart`). Partition analysis is a progression of the silhouette analysis concept and defines a global statistic, the PARTANA ratio, which is the mean similarity of samples within a cluster to the mean similarity of samples among all clusters (Peet & Roberts 2013). The second criterion was assessed by expert judgement examination of constancy (species frequency) tables for ecologically meaningful clusters. Ancillary guidance was derived from the variation in community composition.

Fifteen sub-communities were described based on this analysis (Table 6). Community 3b posed some additional issues. This is a community of well-drained neutral soils where there has been agricultural improvement. The two sub-communities described by the available data represent variations in the degree of improvement. Neither sub-community, however, adequately describes the highly modified, species-poor *Lolium* grasslands of intensive farmland or amenity areas (habitats GA1 and GA2 of GHI) as there is a lack of data of this nature in the datasets. Artificially constructed plots of this type are, however, assigned to this community by noise cluster analysis. A 'place-holder sub-community' was therefore created for community 3b in addition to the two sub-communities suggested by the analysis, to accommodate this very abundant and widespread cultural community until such time as sufficient data become available to define it statistically.

Table 6. Assessment of potential sub-communities defined by noise clustering analysis on all Assigned plots. Silhouette indicates the overall average silhouette width, Partition indicates the PARTANA ratio, No. indicates the number of sub-communities decided upon.

ISGS community	Cluster variance	2 sub-communities		3 sub-communities		No.
		Silhouette	Partition	Silhouette	Partition	
1a <i>Juncus acutiflorus/articulatus</i> – <i>Holcus lanatus</i>	0.18	0.08	1.135	0.07	1.151	0
1b <i>Agrostis stolonifera</i> – <i>Filipendula ulmaria</i>	0.26	0.08	1.211	0.09	1.317	2
1c <i>Molinia caerulea</i> – <i>Succisa pratensis</i>	0.20	0.07	1.148	0.08	1.173	0
1d <i>Molinia caerulea</i> – <i>Potentilla erecta</i>	0.21	0.14	1.207	0.15	1.329	3
1e <i>Juncus acutiflorus/articulatus</i> – <i>R. squarrosus</i>	0.17	0.07	1.112	0.06	1.134	0
2a <i>Agrostis stolonifera</i> – <i>Ranunculus repens</i>	0.25	0.09	1.266	0.10	1.328	2
2b <i>Juncus effusus</i> – <i>Holcus lanatus</i>	0.17	0.07	1.106	0.07	1.134	0
2c <i>Holcus lanatus</i> – <i>Lolium perenne</i>	0.21	0.14	1.299	0.10	1.293	2
2d <i>Juncus effusus</i> – <i>Rumex acetosa</i>	0.17	0.07	1.117	0.07	1.176	0
3a <i>Briza media</i> – <i>Thymus polytrichus</i>	0.22	0.11	1.280	0.06	1.263	2
3b <i>Cynosurus cristatus</i> – <i>Trifolium repens</i>	0.15	0.14	1.202	0.08	1.186	2*
3c <i>Festuca rubra</i> – <i>Plantago lanceolata</i>	0.20	0.07	1.154	0.06	1.162	0
3d <i>Cynosurus cristatus</i> – <i>Trifolium pratense</i>	0.15	0.08	1.113	0.06	1.112	0
3e <i>Festuca rubra</i> – <i>Rhinanthus minor</i>	0.15	0.07	1.092	0.06	1.108	0
3f <i>Festuca rubra</i> – <i>Lotus corniculatus</i>	0.24	0.10	1.240	0.09	1.285	2
4a <i>Agrostis capillaris</i> – <i>Trifolium repens</i>	0.17	0.08	1.128	0.08	1.149	0
4b <i>Nardus stricta</i> – <i>Festuca ovina</i>	0.21	0.09	1.174	0.08	1.216	2
4c <i>Agrostis capillaris</i> – <i>Festuca rubra</i>	0.19	0.08	1.162	0.07	1.170	0
4d <i>Agrostis canina/vinealis</i> – <i>Carex echinata</i>	0.18	0.09	1.143	0.08	1.165	0

* does not include place-holder sub-community.

At this stage of the IVC project, it seems prudent to be somewhat conservative with the creation of sub-communities. It is not desirable to create divisions in the hierarchy which are too subtle in terms of floristics to be differentiated in the field and further sub-communities can be fairly easily defined at a later date once feedback from practitioners has been received.

4.7 Definition of IVC communities

The plots categorised as Assigned from the review process ($n = 6777$) were used to define a statistical model (an R object of class `vegclust`) for the IVC grassland classification based on the cluster assignments made during the review process. All ISGS plots ($n = 4516$) and NVD plots ($n = 5966$) were then reclassified using this new model (Table 7). ISGS plots and Type 1 NVD plots categorised as Assigned by this definition process were then used to describe the nineteen communities following the guidelines of presentation set down in sub-section 2.4. The names of the communities were reviewed and some minor alterations made.

Table 7. Final assignment of plots to the IVC grassland classification. A = number of Assigned plots, T = number of Transitional plots, U = number of Unassigned plots. Name changes are underlined

IVC community	ISGS and Type 1 plots				Type 2-4 plots				Total
	A	T	U	Sub-total	A	T	U	Sub-total	
GL1A <i>Juncus acutiflorus</i> – <i>Holcus lanatus</i>	246	32	-	278	20	10	-	30	308
GL1B <i>Agrostis stolonifera</i> – <i>Filipendula ulmaria</i>	266	73	-	339	60	24	-	84	423
GL1C <i>Molinia caerulea</i> – <i>Succisa pratensis</i>	284	58	-	342	17	11	-	28	370
GL1D <i>Molinia caerulea</i> – <i>Potentilla erecta</i>	234	23	-	257	29	8	-	37	294
GL1E <i>Juncus acutiflorus</i> – <i>R. squarrosus</i>	244	69	-	313	40	7	-	47	360
GL2A <i>Agrostis stolonifera</i> – <i>Ranunculus repens</i>	242	25	-	267	187	20	-	207	474
GL2B <i>Juncus effusus</i> – <i>Holcus lanatus</i>	399	69	-	468	47	17	-	64	532
GL2C <i>Holcus lanatus</i> – <i>Lolium perenne</i>	385	59	-	444	164	30	-	194	638
GL2D <i>Juncus effusus</i> – <i>Rumex acetosa</i>	295	28	-	323	25	16	-	41	364
GL3A <i>Briza media</i> – <i>Thymus polytrichus</i>	657	44	-	701	931	31	-	962	1663
GL3B <i>Lolium perenne</i> – <i>Trifolium repens</i>	932	52	-	984	255	23	-	278	1262
GL3C <i>Festuca rubra</i> – <i>Plantago lanceolata</i>	374	112	-	486	129	44	-	173	659
GL3D <i>Cynosurus cristatus</i> – <i>Trifolium pratense</i>	810	89	-	899	85	37	-	122	1021
GL3E <i>Festuca rubra</i> – <i>Rhinanthus minor</i>	205	58	-	263	23	18	-	41	304
GL3F <i>Festuca rubra</i> – <i>Lotus corniculatus</i>	200	32	-	232	13	9	-	22	254
GL4A <i>Agrostis capillaris</i> – <i>Trifolium repens</i>	403	58	-	461	65	21	-	86	547
GL4B <i>Nardus stricta</i> – <i>Potentilla erecta</i>	335	19	-	354	71	7	-	78	432
GL4C <i>Agrostis capillaris</i> – <i>Potentilla erecta</i>	223	49	-	272	10	4	-	14	286
GL4D <i>Agrostis canina/vinealis</i> – <i>R. squarrosus</i>	200	36	-	236	9	3	-	12	248
Noise cluster	-	18	5	23	-	16	3	19	42
Total	6934	1003	5	7942	2180	356	3	2539	10481

To produce the hectad maps, a grid reference identifying the 10 km × 10 km square was needed for each of the plots. These were available for all of the ISGS plots and the majority of the NVD plots. For some plots, however, there was either no grid reference available or the grid reference data did not identify the specific hectad (Table 8). For example, the grid reference information for all 1010 plots in UCD014 is simply “M10, 11, 20” referring to three hectads which cover much of the Burren. Unfortunately, this meant that there were 861 plots from this dataset which could be used neither to define the communities nor to map them.

Grid references were also not available for a substantial number of plots in NPWS057 which was by far the largest of the NVD datasets. Furthermore, a brief examination of the grid references that are available for this dataset revealed a number of errors. There are location descriptions for most of the plots in NPWS057 and Bourke *et al.* (2007) mention that the exact location of the fields in which plots were recorded is marked on a set of paper maps at Johnstown Castle, so it should be possible to check or generate grid references at the hectad

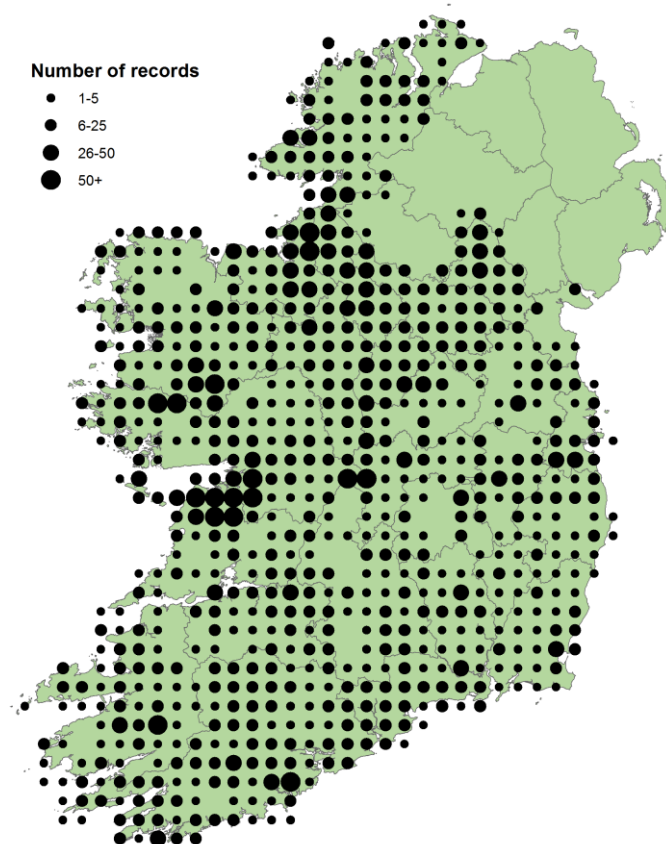
resolution. It was beyond the resources of this phase of the project to do this (all grid references were used as provided), but it would greatly improve the quality of the location data.

Table 8. Grassland plots from the NVD for which hectads could not be determined.

Dataset reference	NVD code	Plots without adequate grid references		
		NVD Type 1	NVD Types 2-4	NVD total
L. Weekes	NPWS024	19	0	19
A. Bleasdale	NUIG004	18	0	18
M. Telford	TCD002	12	0	12
B. Dunford	UCD014	149	861	1010
S. Keane	NUIG013	28	0	28
G. O'Donovan 1	INDEP023	49	20	69
J. Braun-Blanquet	NPWS058	0	3	3
A. O'Sullivan	NPWS057	1003	265	1268
<i>Total</i>		<i>1278</i>	<i>1149</i>	<i>2427</i>

The distribution of all mapped plots is shown in Fig.4. Coverage is very widespread but there are still some hectads with no records.

Figure 4. Distribution and frequency of grassland plots mapped at the hectads level.



5 Web pages

In collaboration with the NBDC, a set of web pages were developed and hosted on the NBDC website. These consist of:

1. A home page introducing the project and its aims.
2. A background page detailing the rationale for the project and previous projects in this area.
3. An explore page with a collapsing/expanding tree to navigate the classification.
4. A community page for each of the nineteen grassland communities.
5. A progress page which will log all updates and changes to IVC, with download links for the Technical Progress Reports.
6. A page from which the web application, ERICA, can be launched.
7. A links page linking to websites dealing with vegetation classification.

A “community synopsis” has been produced in PDF format for each community and includes the following sections: scientific name, common name, code, vegetation, sub-communities, similar communities, ecology, records and distribution (with map), synoptic table, affinities, proxy environmental data, conservation value, management, threats, key references and two example photographs. The version number, date and author of the synopsis are also included. The structure and tone of these synopses is based in part on the British NVC community descriptions in Averis *et al.* (2004), for example the inclusion of brief notes on conservation value and management which places each community in a more practical context. These PDFs can be downloaded individually from a link on the web page of the relevant community, or they can be downloaded in bulk, with or without photographs, from the exploration web page. The community pages themselves present a subset of the information in the synopsis.

6 Web application

A web application entitled ERICA (Engine for Relevés to Irish Communities Assignment) was developed using the R statistical environment V3.2.2 and the `shiny` package. ERICA assigns vegetation data to communities defined by the IVC. Data can be uploaded, checked for errors and analysed and the results can then be downloaded. ERICA is hosted on the website of the National Biodiversity Data Centre. The release version is V1.0. A full description of the application is given in the User’s Manual.

7 Recommendations

Some brief notes are made here to highlight issues which arose during this phase and to give guidance to futures phases of the IVC.

1. The IVC classification is based on a dataset compiled from many different surveys and projects, therefore sampling effort has not been geographically distributed in an even manner. There has undoubtedly been more intensive sampling of certain habitats in certain parts of the country (e.g. calcareous grassland in the Burren). Site-specific surveys may have contributed multiple examples of the same vegetation types.

Stratification of the dataset as conducted by Wiser & De Cáceres (2013) would have reduced this geographical bias but at the cost of excluding a large proportion of plots.

2. The range of the sizes of plots used in the review process, and subsequently to define the communities (1-25 m²), is quite broad. Peet & Roberts (2013) discuss conceptual problems with comparing vegetation samples from plots of different sizes. Decisions in this regard were again guided by the desire to retain as many plots as possible. Use of consistent plot sizes for future recording is strongly encouraged: 2 m x 2 m plots are suitable for Irish grasslands.
3. Bryophyte recording was, apparently, often absent or patchy in the NVD datasets; bryophytes can be sparse in some grassland types, therefore it was not easy to decide whether a genuine lack of recording had occurred. Exclusion of plots with poor bryophyte recording was used to improve the accuracy of the classification. However, it is likely that bryophytes are still under-represented in the synoptic tables. Recording of all bryophytes is strongly encouraged for future recording.
4. Many of the NVD datasets used a version of the Braun-Blanquet scale, which has a strong tradition in Ireland. However, in practice it yields fairly crude estimates of vegetation cover (and certainly should never be used for monitoring purposes). The Domin scale is an improvement on the Braun-Blanquet scale but the use of percentage cover in future recording is strongly encouraged.
5. Further work on other divisions may alter these grassland communities. This is an inevitable consequence of the phased IVC development. For example, work on fens and turloughs will impact on the classification of wet grassland and marsh, and work on limestone pavement may impact on that of calcareous grassland. Heaths may impact on the definition of acid grassland; indeed upland grasslands are possibly under-represented as the ISGS was primarily lowland in its focus. The core nature of each community is, however, unlikely to be radically altered.
6. Some taxonomic issues remain to be resolved but these are not of primary importance in grasslands. For example, two records of *Sphagnum recurvum* were overlooked during the NVD data assessment stage; in later stages these records should be excluded or merged with other members of the *S. recurvum* agg. (e.g. *S. fallax*). In other habitats, *S. auriculatum* agg. will need similar treatment as these records could be *S. denticulatum* or *S. inundatum*. It may also be necessary to merge records of *Conocephalum conicum* and *C. salebrosum* and those of *Dactylorhiza majalis* and *D. kerryensis*.
7. Due to taxonomic changes, there were some difficulties in directly matching several taxa in the checklist with taxa listed in Hill *et al.* (2004), Hill *et al.* (2007) and Blockeel *et al.* (2014); this was necessary for calculation of mean cover-weighted Ellenberg scores and rarity indices by ERICA (see the ERICA manual for details of this work).
8. Future recording should always record all plants in the primary dataset to at least the species level. Recording to the genus level or combining species records (e.g. *Festuca rubra/ovina*) greatly impacts on the quality and versatility of the data. Taxa can always be combined at the analysis stage if required.
9. Datasets with deficiencies in grid reference data should be reviewed, using the original source data if available, to see if these deficiencies can be remedied.
10. There are some existing classifications for Irish vegetation that could be incorporated into the IVC framework (e.g. the National Survey of Native Woodlands classification and the Saltmarsh Angiosperm Assessment Tool for Ireland classification). Some progress has also been made with upland habitats by the National Survey of Upland Habitats, but

this is not completed to the same level. Other habitats are likely to proportionately require a greater amount of work.

8 Acknowledgements

Many thanks are due to my colleagues at BEC Consultants for their input on this project, in particular Fionnuala O'Neill, and their assistance on the original ISGS classification. Thanks also go to the staff of the NBDC for their technical support, in particular Úna FitzPatrick for her advice and feedback. The IVC project makes extensive use of the `vegclust` R package written by Miquel De Cáceres; without this tool the IVC would not have been possible in its current guise. I would also like to thank Oliver Pescott, who provided unpublished data on bryophyte distributions. This phase of the IVC was funded by National Parks and Wildlife Service.

9 References

- Averis, A., Averis, B., Birks, J., Horsfield, D., Thompson, D., Yeo, M. (2004) *An illustrated guide to upland British vegetation*, Joint Nature Conservation Committee, Peterborough, UK.
- Blockeel, T.L., Bosanquet, S.D.S., Hill, M.O., Preston, C. (eds.) (2014) *Atlas of British and Irish bryophytes*. Pisces Publications, Newbury, UK.
- Bourke, D., Hochstrasser, T., Nolan, S., Schulte, R. (2007) Historical grassland Turboveg database project: 2067 relevés recorded by Dr Austin O'Sullivan 1962-1982. Unpublished report submitted to National Parks and Wildlife Service of the Department of the Environment, Heritage and Local Government, Dublin, IE.
- De Cáceres, M., Font, X., Oliva, F. (2010) The management of vegetation classifications with fuzzy clustering. *Journal of Vegetation Science*, **21**, 1138-1151.
- Devaney, F.M., Perrin, P.M. (2015) Saltmarsh Angiosperm Assessment Tool for Ireland (SMAATIE) (2013-W-DS-10). STRIVE Report. Environmental Protection Agency, Johnstown Castle, Wexford, IE.
- Faber-Langendoen, D., Keeler-Wolf, T., Meidinger, D., Josse, C., Weakley, A., Tart, D., Navarro, G., Hoagland, B., Ponomarenko, S., Saucier, J.-P., Fults, G., Helmer, E. (2012) Classification and description of world formation types. Part I (Introduction) and Part II (Description of formation types). Hierarchy Revisions Working Group, Federal Geographic Data Committee, FGDC Secretariat, U.S. Geological Survey, Reston, VA, and NatureServe, Arlington, Virginia, US.
- Hill, M. O., Preston, C.D., Roy, D.B. (2004) PLANTATT. Attributes of British and Irish plants: status, size, life history, geography and habitats, Centre for Ecology & Hydrology, Huntingdon, UK.
- Hill, M. O., Preston, C.D., Bosanquet, S.D.S., Roy, D.B. (2007) BRYOATT. Attributes of British and Irish Mosses, Liverworts and Hornworts with information on native status, size, life form, life history, geography and habitat, Centre for Ecology & Hydrology, Huntingdon, UK.

- Kaufman, L., Rousseeuw, P.J. (1990) *Finding groups in data: an introduction to cluster analysis*. Wiley, New York, US.
- Kent, M. (2012) *Vegetation description and data analysis: a practical approach*, 2nd edition. Wiley-Blackwell, Chichester, UK.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2013) Reference manual, cluster: Cluster Analysis Basics and Extensions. R package version 1.14.4.
- McCune, B., Grace, J.B. (2002) *Analysis of ecological communities*. MjM Software Design, Oregon, US.
- O'Neill, F.H., Martin, J.R., Devaney, F.M., Perrin, P.M. (2013) The Irish semi-natural grasslands survey 2007-2012. *Irish Wildlife Manuals*, No. 78. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, IE.
- Peet, R.K., Roberts, D.W. (2013) Classification of natural and semi-natural vegetation. pp. 28-70 in: Franklin, J., van der Maarel E. (eds.) *Vegetation Ecology*. 2nd edition. Oxford University Press, New York, US.
- Rich, T.C.G. Jermy, A.C. (1998) *Plant crib 1998*, Botanical Society of the British Isles, London, UK.
- Rodwell, J. S. (ed.) (1992) *British plant communities. Volume 2. Grasslands and montane communities*. Cambridge University Press, Cambridge, UK.
- Rodwell, J.S. (ed) (2000) *British plant communities Volume 5: Maritime communities and vegetation of open habitats*. Cambridge University Press, Cambridge, UK.
- Rodwell, J.S., Schaminée, J.H.J., Mucina, L., Pignatti, S., Dring, J., Moss, D. (2002) *The diversity of European vegetation: an overview of phytosociological alliances and their relationships to EUNIS habitats*. National Reference Centre for Agriculture, Nature and Fisheries, Wageningen, NL.
- Rousseeuw, P.J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**, 53-65.
- Stace, C.A. (2010) *New flora of the British Isles*, 3rd Edition. Cambridge University Press, Cambridge, UK.
- Stevens, D.P., Smith, S.L.N., Blackstock, T.H., Bosanquet, S.D.S, Stevens, J.P. (2010) *Grasslands of Wales. A survey of lowland species-rich grasslands, 1987-2004*, University of Wales Press, Cardiff, UK.
- Webb, D.A. (1980) The biological vice-counties of Ireland, *Proceedings of the Royal Irish Academy*, **80B**, 179-196.
- Wiser, S.K., De Cáceres, M. (2013) Updating vegetation classifications: an example with New Zealand's woody vegetation. *Journal of Vegetation Science*, **24**, 80-93.
- Zadeh, L.A. (1965) Fuzzy sets. *Information and Control*, **8**, 338-353.