Diversity and Distributions, (Diversity Distrib.) (2006) **12**, 633–644

BIODIVERSITY RESEARCH

Comparing agglomerative clustering and three weed classification frameworks to assess the invasiveness of alien species across spatial scales

Roger A. Lawes[1,2], Helen T. Murphy[1]* and Anthony C. Grice[1,2]

[1]*CRC for Australian Weed Management,* [2]*CSIRO Sustainable Ecosystems, Davies Laboratory, Townsville, Queensland 4814, Australia,* [3]*CSIRO Sustainable Ecosystems, Atherton, Queensland 4883, Australia*

*Correspondence: Helen T. Murphy, CSIRO Sustainable Ecosystems, Davies Laboratory, Townsville, Queensland 4814, Australia. Tel.: +61 747538537; Fax: +61 747538600; E-mail: roger.lawes@csiro.au

ABSTRACT

To prioritize weed management at the catchment scale, information is required on the species present, their relatively frequency, abundance, and likely spread and impact. The objective of this study was to classify the invasiveness of alien species that have invaded the Upper Burdekin Catchment in Queensland, Australia, at three spatial scales. A combination of three published weed classification frameworks and multivariate techniques were employed to classify species based on their frequency and cover at a range of spatial scales. We surveyed the Upper Burdekin Catchment for alien species, and for each species determined the following distribution indices — site frequency, total cover, transect frequency per site frequency and quadrat frequency per site frequency, cover per quadrat when present, cover per transect when present, and cover per site when present. These indices capture the effect of species abundance and frequency between sites (site frequency and total cover), within sites (transect frequency per site and cover per transect when present), and within transects (quadrat frequency per site frequency and cover per site). They were used to classify the species into seven groups using a hierarchical cluster analysis. The relationship between the indices was explored to determine how effective the small scale, site-specific indices were at predicting the broader, landscape-scale patterns. Strong correlations were observed between transect frequency per site and frequency ($r^2 = 0.89$) and cover per transect when present and total cover ($r^2 = 0.62$). This suggests that if a weed is abundant at the site level, it has the potential to occupy large areas of the catchment. The species groupings derived from the application of the three published weed classification frameworks were compared graphically to the groupings derived from the cluster analysis. One of the frameworks classified species into three groups. The other two frameworks classified species into four groups. There was a high degree of subjectivity in applying the frameworks to the survey data. Some of the data were of no relevance to the classification frameworks and were therefore ignored. We suggest that the weed classification frameworks should be used in conjunction with existing multivariate techniques to ensure that classifications capture important natural variations in observed data that may reflect invasion processes. The combined use of the frameworks and multivariate techniques enabled us to aggregate species into categories appropriate for management.

Keywords

Biological invasions, classification frameworks, clustering, invasive species, landscape scale, ordination.

INTRODUCTION

New species entering a landscape are subject to community-level processes and may or may not become invasive. This fate is governed by the characteristics of the species and habitat it has entered (Williamson & Fitter, 1996; Richardson & Pyšek, 2006). The ecological definition of invasiveness has been discussed at length by Williamson (1999), Lonsdale (1999), Davis & Thompson (2000), Rejmánek *et al.* (2002), and Colautti & MacIsaac (2004). Despite these discussions, ecologists have struggled to classify

DOI: 10.1111/j.1366-9516.2006.00291.x
www.blackwellpublishing.com/ddi

Table 1 Data requirements to implement the weed classification framework devised by Richardson *et al.* (2000); Colautti & MacIsaac (2004); Davis & Thompson (2000)

Weed invasion framework of Richardson *et al.* (2000)

Classification	Additional classification	General criterion	Our criterion
Introduction	Barrier A	Overcome major geographical barriers	n/a
Introduction	Barrier B	Overcome environmental barriers at introduction	n/a
Naturalized	Barrier C	Overcome reproduction barriers	< 15 sites
Naturalized	Barrier D	Overcome local/regional dispersal barriers	< 15 sites
Invasive	Barrier E	Present in disturbed habitat	> 15 sites
Transformer	Barrier F	Present in undisturbed habitat	> 15 sites/mod-high quadrat cover

Weed invasion framework of Davis & Thompson (2000)

Classification	Criterion 1 Dispersal	Criterion 2 Colonizer Origin	Criterion 3 Species Impact	Our criterion
Type 1 (SCS)	Short	Common	Small	< 35 sites/ < 5% cover
Type 2 (SCG)	Short	Common	Great	< 35 sites/ > 5% cover
Type 3 (SNS)	Short	Novel	Small	n/a
Type 4 (SNG)	Short	Novel	Great	n/a
Type 5 (LCS)	Long	Common	Small	> 35 sites/ < 5% cover
Type 6 (LCG)	Long	Common	Great	> 35 sites/ > 5% cover
Type 7 (LNS)	Long	Novel	Small	n/a
Type 8 (LNG)	Long	Novel	Great	n/a

Weed invasion and classification framework of Colautti & MacIsaac (2004)

Classification	General criterion	Our criterion
Stage 0	Propagule in a new region.	n/a
Stage 1	Propagule survives and disperses in the new region.	n/a
Stage 2	Species reproduces in new environment.	n/a
Stage 3	Species localized and numerically rare.	< 15 sites/low cover
Stage 4a	Species widespread but rare.	> 35 sites/low cover
Stage 4b	Species localized, but dominant.	< 35 sites/mod–high cover
Stage 5	Species widespread and dominant.	> 35 sites/mod–high cover

successful invasions, partly because successful invasions are quite rare. According to the 'tens rules' (Williamson & Fitter, 1996), a species has approximately a 10^{-6} probability of becoming a pest once introduced to a novel ecosystem.

Initially, ecologists focused on classifying species based on important plant characteristics (e.g. Mack, 1996) culminating in models of plant invasiveness with a view to predicting which species are likely to become invasive (e.g. Perrins *et al.*, 1992; Rejmánek, 1995). Biosecurity agencies adopted key concepts from these studies to develop weed risk assessment protocols (Pheloung *et al.*, 1999), although recent reviews of these protocols found intentional human dispersal of propagules to be more important than plant species' ecological traits such as dispersal mechanism or seed size (Caley & Kuhnert, 2006).

Richardson *et al.* (2000), Davis & Thompson (2000), and Colautti & MacIsaac (2004) all developed frameworks to enable researchers to either describe a species as a 'type' of invader (Davis & Thompson, 2000; Richardson *et al.*, 2000) or classify the invasion itself as a series of stages (Colautti & MacIsaac, 2004) (Table 1). The ability to classify an invasion is necessary to prioritize weed management.

The framework of Richardson *et al.* (2000) comprised three components: introduction, naturalization, and invasion. Introduction implies that the species has been transported by humans across major geographical boundaries. Naturalization occurs when the species has overcome abiotic and biotic barriers to survival and can cope with environmental stochasticity. Invasion occurs when the species produces reproductive offspring distant from the parent and culminates in new self-perpetuating populations. Six key barriers that limit the spread of introduced plants were defined to enable researchers to classify species into one of the three categories (Table 1). Pyšek *et al.* (2004) recently discussed further details on the use and application of Richardson *et al.* (2000)'s framework.

Davis & Thompson (2000) proposed a scheme modeled after Rabinowitz's (1981) classification of rarity, to distinguish 'invaders' as distinct from 'successional colonizers' and 'non-invasive colonizers'. The authors defined eight different types of 'colonizers' based on three distinct characteristics: dispersal distance (short/long), uniqueness (novel/common to the region colonized), and impact on the new environment (small/great) (Table 1). The interpretation of dispersal distance and the origin of the colonizer are scale dependent, while impact may be assessed from either a productivity or an ecosystem function.

Colautti & MacIsaac (2004) recently proposed a framework that attempted to eliminate the need for universal definitions of current terms by using operational terms with no *a priori* meaning (i.e. as one of five possible stages) in a process based model. The first Stage (0) begins when a propagule enters the habitat of interest. A species proceeds through Stage 1, dispersal and survival in the habitat, and Stage 2, survival and reproduction. If the weed satisfies these criteria, it may then be classified into separate stages that describe the frequency and abundance of a species in an environment. Stage 3 species are considered localized and numerically rare. Stage 4a species are widespread but rare, Stage 4b species are localized but dominant in those localities, and Stage 5 species are considered widespread and dominant (Table 1). Again, the environment can be any region of interest. The definitions of widespread and abundant are also open to interpretation. In essence, all three frameworks put forward a series of hypotheses that can be applied to data to classify the invasive characteristics of a species. The data requirements necessary to implement the frameworks vary, although each draws heavily on frequency and abundance information to classify the invasive characteristics of a species.

The strengths and weaknesses of these frameworks were recently evaluated by Murphy *et al.* (2006) who used existing data on invasive species in Canada, and tested whether the frameworks were operational given a relatively rich species data set, and whether the different frameworks identified the same sets of species. These authors had difficulty defining the rate of spread, required by Richardson *et al.* (2000)'s scheme, and impact as required by Davis & Thompson (2000)'s, and concluded that both schemes had limited potential in classifying regional alien floras.

In this study, we surveyed the riparian zone of the Burdekin river system for alien species. This river system, the third largest in Australia, extends for over 1700 km and connects an array of pastoral properties. Alien plant species such as *Cryptostegia grandiflora* and *Ziziphus mauritiana* are present in the catchment and affect the productive capacity of pastoral properties in this catchment (Grice *et al.*, 2000). However, the extent of their geographical distribution and possible impact along the riparian zone remains unknown. Other alien grass species, such as *Andropogon gayanus*, have dramatically altered ecosystem function in other Australian rangeland systems and may be present in this catchment (Rossiter *et al.*, 2003). It was therefore necessary to document the weed status of this ecosystem to assist managers and state agencies in formulating a regional weed management strategy.

The frequency and abundance of each species were recorded at course and fine scales. We apply a series of multivariate techniques and the frameworks devised by Richardson *et al.* (2000), Davis & Thompson (2000), and Colautti & MacIsaac (2004) to each species objectively, using the data derived from the survey. The frameworks present a series of hypotheses about invasion and apply these to data. We propose evaluating weed invasions as part of a larger, existing plant community where invasive species are classified based on their relative abundance, frequency, and cover at local and landscape scales. This approach differs from those proposed by the authors of previous frameworks as species are classified using all available data acquired from an extensive, multiscale landscape survey of species frequency and cover. The patterns derived from multivariate analyses may help formulate hypotheses about invasion processes in the landscape of interest and can be applied to any data set. Once critical invasive species have been identified, management strategies can be formulated to target important species in the ecosystem.

METHODS

Survey design

Surveys were conducted on the Upper Burdekin Catchment, situated in North Queensland, Australia (Fig. 1). The Burdekin River and associated tributaries extend for 1700 km. Eighty sites were surveyed at intervals of approximately 17 km in the riparian zone of this catchment.

At each site, a transect consisting of 10 contiguous 10 m × 5 m quadrats was surveyed on the riverbank, on the mid-slope of the riparian zone and on the upper bank of the riparian zone. Within each site, this design was replicated three times, with each replicate separated by 100 m. In all, 90 quadrats were surveyed throughout a site.

At each quadrat, trained observers visually estimated and recorded the percentage canopy cover of each alien species. From these observations, the presence or absence of each species was determined at the quadrat, transect, and site levels.

Indices and statistical analysis

As part of the survey, canopy cover, c, in the i^{th} site, j^{th} transect, and k^{th} quadrat was recorded as a percentage of the quadrat. These were transformed into a portion by dividing by 100 to allow the portions in individual quadrats to be summed. Seven indices were derived from the surveys to determine what portion of area the weed occupies, when present, at the quadrat level, transect level, and site level. If cover in the quadrat was greater than zero, then the presence, p, of a weed, in the i^{th} site, j^{th} transect, and k^{th} quadrat was recorded as one, otherwise it was zero. For each species, four measures of cover were used: total cover, C; cover per site, C_s; cover per transect, C_t; and cover per quadrat, C_q (Table 2). Total cover was the cumulative sum of the proportion of cover measured in each quadrat. Cover per quadrat, cover per transect, and cover per site provided a measure of cover occupied by the weed at each scale.

As total cover was the cumulative sum of the portion of cover in all quadrats, it was necessary to scale lower level frequency measures to match this estimate of total cover. Therefore, transect frequency was multiplied by 10, as there are 10 quadrats

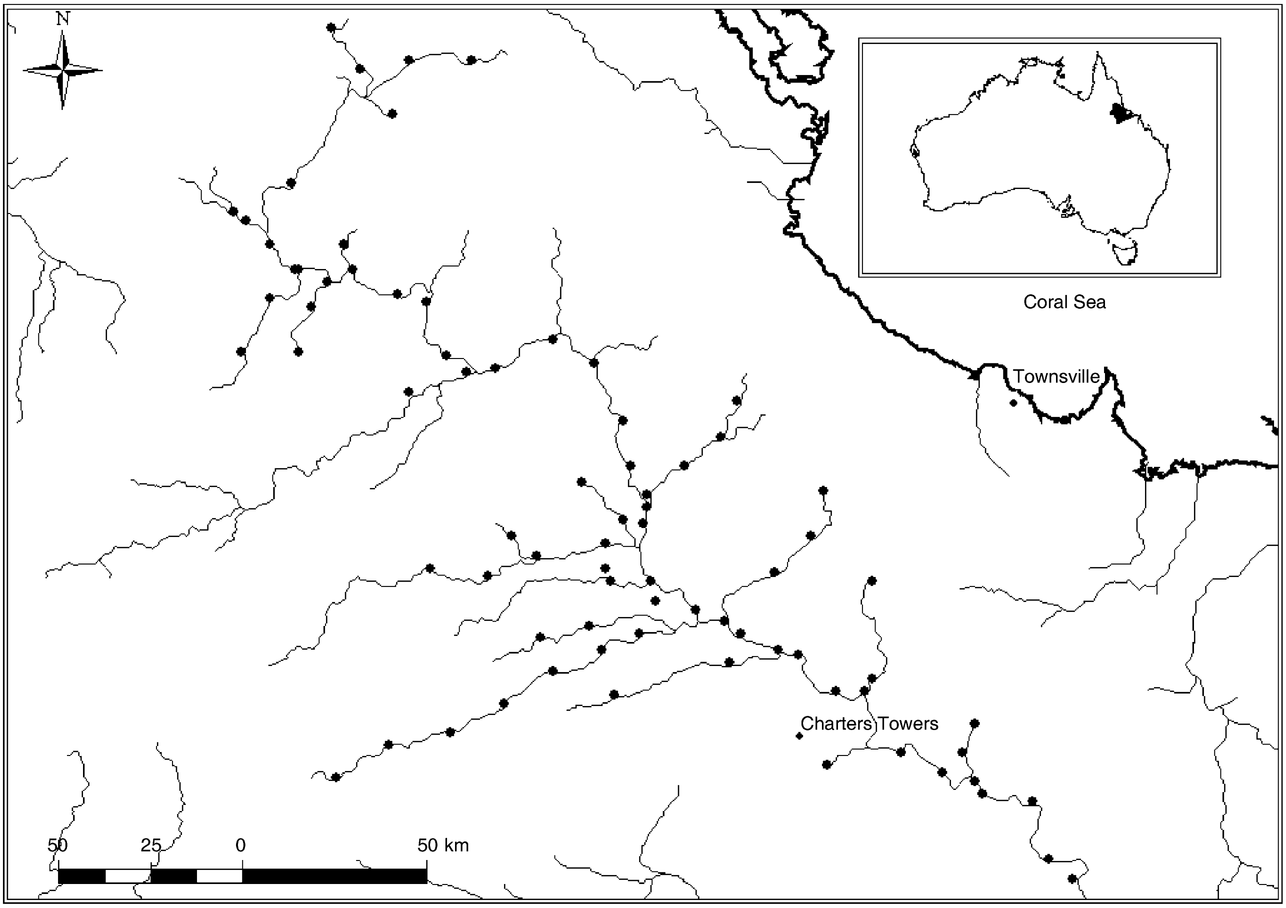

Figure 1 Location of the Burdekin River and sampling sites.

Table 2 Cover indices and frequency calculations used in principal component analysis and cluster analysis

Index	Equation	Comment
Total cover (portion)	$C = \sum_{i}^{n=80} \sum_{j}^{n=9} \sum_{k}^{n=10} c_{ijk}$	Cumulative sum of portion of cover in each quadrat
Cover per quadrat, C_q	$C_q = \frac{C}{\left(\sum_i \sum_j \sum_k p_{ijk}\right)}$	C_q = total cover/frequency of quadrats where weed was present
Cover per transect, C_t	$C_t = \frac{C}{\left(\sum_i \sum_j p_{ij}\right) \times 10}$	C_t = total cover/frequency of transects where weed was present
Cover per site, C_s	$C_s = \frac{C}{\left(\sum_i p_i\right) \times 10 \times 9}$	C_s = total cover/frequency of sites where weed was present
Quadrat frequency	$if\ c_{ijk} > 0\ then\ p_{ijk} = 1$ $\therefore Frequency = \sum_i \sum_j \sum_k p_{ijk}$	Cumulative sum of all quadrats where species were present
Transect frequency	$if\ c_{ij} > 0\ then\ p_{ij} = 1$ $\therefore Frequency = \sum_i \sum_j p_{ij}$	Cumulative sum of all transects where species were present
Site frequency	$if\ c_i > 0\ then\ p_i = 1$ $\therefore Frequency = \sum_i p_i$	Cumulative sum of all sites where species were present

in each transect. Site frequency was multiplied by 9 and 10 as there were nine transects in each site and 10 quadrats in each transect. These indices differ from estimates of mean cover because absences, at different scales, are deliberately ignored in order to give an indication of the proportion of land a weed occupies once it has invaded.

Three measures of frequency were also defined: site frequency, transect frequency per site frequency, and quadrat frequency per site frequency. These were simply derived from the three frequency indices presented in Table 2. Site frequency was the number of sites where the species was present. Transect frequency per site frequency provides a measure of the mean number of transects occupied by the species when present at the site level. Nine transects were surveyed at each site and this was the maximum possible value. Quadrat frequency per site frequency provides an indication of the mean number of quadrats occupied by the weed when present. The maximum possible value was 90. The two relative frequencies indicate the extent to which a species occupies a given site. For each species, these indices collectively capture the variation in frequency and cover at three spatial scales. Site frequency and total cover provide information on these effects at the landscape scale. Transect frequency per site and cover per transect explain the effects of a species along a 100 m transect within a site. Quadrat frequency per site and quadrat cover per site capture details of species effects at a quadrat level within the site.

Prior to analysis, data were centred by subtracting the attribute mean, and standardized by dividing by the centred attribute's root mean square, enabling all attributes to be considered with equal weight. Data were analyzed using non-metric multidimensional scaling to summarize the relationship between the seven indices for the 66 species identified in the survey. The dissimilarities between species cover values were calculated using a Euclidean distance measure.

Data were furthered summarized using an agglomerative hierarchical cluster analysis using the same distance matrix as the non-metric multidimensional scaling (Gordon, 1999). Species were agglomerated using Ward's minimum variance method that seeks to minimize the within group variation (see Gordon, 1999 for discussion). Small groups of species were agglomerated into larger groups using the same method and this iterative process continued until all species formed a single group. In this way, group membership was optimized at each stage of the agglomeration, and was essentially a local, or single stage optimization (Gordon, 1999). Results of this process were displayed in a dendrogram, arbitrarily cut at a height of seven. It is realized that trade-offs exist between a parsimonious interpretation of data and the need to preserve and identify the inherent variation that exists between species groups.

Application of weed classification frameworks

Weed classification frameworks were applied to every species. The criteria used for each of the three frameworks are shown in Table 1. Davis & Thompson (2000) required information on dispersal distance, uniqueness to the region, and impact on the new environment. Dispersal distance (short or long) was derived from site frequency, not from an understanding of the species dispersal mechanism. Plants with high site frequency (> 35 sites out of 80 surveyed, or 43%) were considered long-distance dispersers, whereas those occupying fewer sites were considered short-distance dispersers. Uniqueness to the region, whether common or novel, was difficult to classify from the data alone. Davis & Thompson (2000) argue that species introduced into North America during colonization should now be considered common as they are unlikely to ever be eradicated. Nearly all the species surveyed here satisfy that criterion, with most present in Australia for over 100 years. Therefore, all cluster groups were considered common. Impact, either little or great, was determined as cover per quadrat. Species occupying more than 5% of a quadrat when present were considered to have a great impact, and those that occupied less than this were considered to have a small impact. Impact was assessed at a very local scale. Site and transect level impacts were excluded from the assessment for parsimony, although the three measures were strongly correlated (Fig. 2b).

Colautti & MacIsaac (2004)'s classification system is largely devised around concepts of frequency and abundance and it was a relatively simple process to classify species based on the survey data, where cover provided a surrogate measure of abundance. Species were all able to reproduce and survive, and by default, had reached Stage 3. Species with low cover per quadrat (< 5%) and very low site frequency (< 15 sites) were considered Stage 3 species (localized and rare). Species with high cover per quadrat (> 5%) and low to moderate site frequency (< 35 sites) were considered Stage 4b species (localized but dominant). Species present at more than 35 sites but with low quadrat cover were considered Stage 4a species (widespread but rare). All other species, with moderate to high quadrat cover and a high site frequency were considered Stage 5 species (widespread and dominant).

It was more difficult to apply the survey data to Richardson *et al.* (2000)'s framework. All species had overcome major geographical barriers, environmental barriers at introduction, reproduction barriers, and local/regional dispersal barriers. Therefore, all species were naturalized. We decided that a species should be present at more than 15 sites before being considered invasive and those that occupied more than 15 sites and had moderate or high quadrat cover were considered transformer species. Data were not available to assess whether high levels of quadrat cover actually transformed the ecosystem, as suggested by Richardson *et al.* (2000).

We adopted a threshold approach to determine when a species moved from one category to the next. The application of frequency and cover data to the frameworks was complicated by the subjectivity associated with determining when this transition should occur. This is an inherent weakness of all three classification systems. As such, many of the thresholds chosen were arbitrary. It was arguably easier to apply data to Davis & Thompson (2000)'s and Colautti & MacIsaac (2004)'s frameworks. Nevertheless, subjective elements of the classifications varied between the frameworks and this complicated our ability to arrive at seemingly analogous classifications across the three frameworks. This

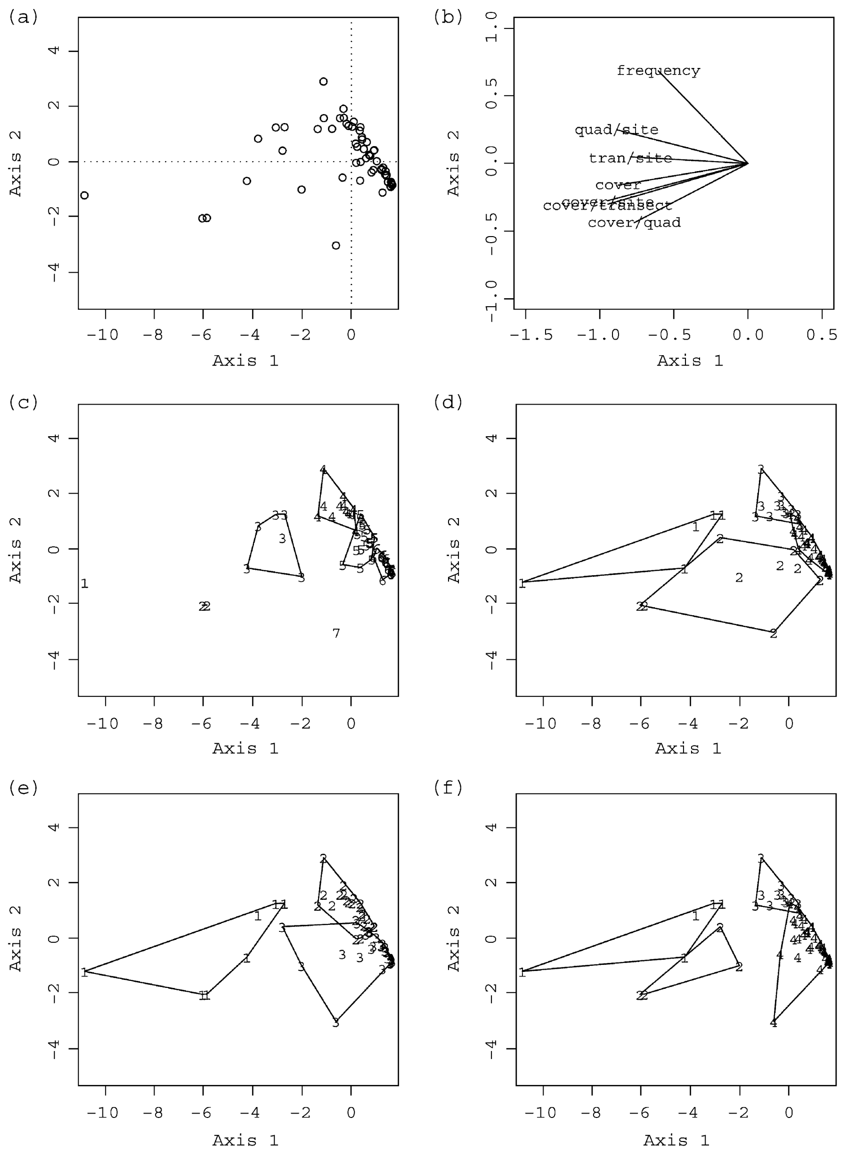

Figure 2 (a) First and second dimension from the non-metric multidimensional scaling of the seven indices. (b) Relationship between the seven attributes and the two axes derived from the non-metric multidimensional scaling. (c) Convex hulls indicate the location of the seven cluster groups in multidimensional space. (d) Convex hulls indicate the location of the four groups derived from the application of Davis & Thompson's (2000) framework in multidimensional space. (e) Convex hulls indicate the location of the three groups derived from the application of Richardson *et al.*'s (2000) framework in multidimensional space. (f) Convex hulls indicate the location of the four groups derived from the application of Colautti & MacIsaac's (2004) framework in multidimensional space.

may be because none of the frameworks was designed to be used with frequency and cover data, such as those collected here. It was therefore difficult to implement the frameworks as their authors intended. Processes such as impact and dispersal were not explicitly captured and therefore inferred from the available data. The ecological impact of invasive species in the Burdekin Catchment has not been quantified, further limiting our ability to implement the frameworks.

Groupings derived from cluster analysis and the three frameworks were evaluated graphically by superimposing all four groupings onto the ordination to assess the performance of each framework in reduced multivariate space.

RESULTS

Species were classified into three of the four possible categories using Richardson *et al.* (2000)'s framework, where 37 were 'naturalized', 22 were 'invasive', and 7 were 'transformer' species (Table 3). Species were classified into four of the six groups using Davis & Thompson (2000)'s framework (Table 3). Five were considered long-distance dispersal, common with great impact. Twelve were evaluated as long-distance dispersal, common with small impact, and nine were short-distance dispersal, common with great impact (Table 3). The remaining 39 species were short dispersers, common, with small impact. Similarly, Colautti & MacIsaac (2004)'s framework also classified the 66 species into four groups. Forty-five species were classified as Stage 3, localized, but numerically rare, 12 as Stage 4a, widespread but numerically rare, four as Stage 4b, localized and abundant. Five species were classified as Stage 5, widespread and abundant (Table 3).

Analogous classifications were obtained for the most abundant and widespread species, *Bothriochloa pertusa*, *Cenchrus ciliaris*, *C. grandiflora*, *Parthenium hysterophorous*, and *Urochloa mosambicensis*. These were classified as transformer (Richardson *et al.*, 2000), long-distance/common/great (Davis & Thompson,

Table 3 Species group membership and classification according to the frameworks defined by Richardson *et al.* (2000), Colautti & MacIsaac (2004), and Davis & Thompson (2000)

Family	Species	Cluster group	Richardson *et al.*	Colautti & MacIsaac	Davis & Thompson
Poaceae	*Urochloa mosambicensis*	1	Transformer	Stage 5	LCG
Poaceae	*Panicum maximum*	2	Transformer	Stage 4b	SCG
Rhamnaceae	*Ziziphus mauritiana*	2	Transformer	Stage 4b	SCG
Poaceae	*Cenchrus ciliaris*	3	Transformer	Stage 5	LCG
Asclepiadaceae/ Apocynaceae	*Cryptostegia grandiflora*	3	Transformer	Stage 5	LCG
Poaceae	*Bothriochloa pertusa*	3	Transformer	Stage 5	LCG
Asteraceae	*Parthenium hysterophorus*	3	Transformer	Stage 5	LCG
Euphorbiaceae	*Jatropha gossypifolia*	3	Naturalized	Stage 4b	SCG
Verbenaceae	*Lantana camara*	3	Naturalized	Stage 4b	SCG
Amaranthaceae	*Alternanthera bettzikiana*	4	Invasive	Stage 4a	LCS
Asteraceae	*Xanthium strumarium*	4	Invasive	Stage 4a	LCS
Malvaceae	*Sida cordifolia*	4	Invasive	Stage 4a	LCS
Apocynaceae	*Carissa ovata*	4	Invasive	Stage 4a	LCS
Poaceae	*Melinis repens*	4	Invasive	Stage 4a	LCS
Asteraceae	*Acanthospermum hispidum*	4	Invasive	Stage 4a	LCS
Amaranthaceae	*Acharanthes aspera*	4	Invasive	Stage 3	SCS
Fabaceae	*Stylosanthes scabra*	4	Invasive	Stage 4a	LCS
Cyperaceae	*Cyperus rotunda*	4	Invasive	Stage 4a	LCS
Tiliaceae	*Triumphetta pentandra*	4	Invasive	Stage 4a	LCS
Poaceae	*Themeda quadrivalvis*	4	Invasive	Stage 3	SCS
Fabaceae	*Crotolaria species*	4	Invasive	Stage 4a	LCS
Euphorbiaceae	*Ricinus communis*	5	Naturalized	Stage 3	SCG
Caesalpiniaceae	*Parkinsonia aculeata*	5	Invasive	Stage 3	SCG
Malvaceae	*Urena lobata*	5	Invasive	Stage 3	SCS
Lamiaceae	*Hyptis suaveolens*	5	Invasive	Stage 3	SCS
Passifloraceae	*Passiflora foetida*	5	Invasive	Stage 3	SCS
Mimosaceae	*Acacia farnesiana*	5	Invasive	Stage 3	SCS
Solanaceae	*Datura stramonium*	5	Invasive	Stage 3	SCS
Euphorbiaceae	*Chamaesyce hirta*	5	Invasive	Stage 3	SCS
Papaveraceae	*Argemone ochroleuca*	5	Invasive	Stage 4a	LCS
Malvaceae	*Sida acuta*	5	Invasive	Stage 4a	LCS
Amaranthaceae	*Alternanthera ficoidea*	5	Naturalized	Stage 3	SCS
Poaceae	*Digitaria ciliaris*	5	Naturalized	Stage 3	SCS
Fabaceae	*Macroptilium atropurpureum*	5	Naturalized	Stage 3	SCS
Malvaceae	*Sida spinosa*	5	Invasive	Stage 3	SCS
Verbenaceae	*Stachytarpheta jamaicensis*	5	Naturalized	Stage 3	SCS
Fabaceae	*Stylosanthes humilis*	5	Naturalized	Stage 3	SCS
Commeliniaceae	*Tradescantia albiflora*	5	Naturalized	Stage 3	SCS
Poaceae	*Urochloa panicoides*	5	Naturalized	Stage 3	SCG
Mimosaceae	*Acacia nilotica*	6	Naturalized	Stage 3	SCG
Fabaceae	*Aeschenomene indica*	6	Naturalized	Stage 3	SCS
Aloeaceae	*Aloe species*	6	Naturalized	Stage 3	SCS
Amaranthaceae	*Alternanthera pungens*	6	Naturalized	Stage 3	SCS
Amaranthaceae	*Amaranthus species*	6	Naturalized	Stage 3	SCS
Asclepiadaceae	*Asclepias currassavica*	6	Naturalized	Stage 3	SCS
Asteraceae	*Bidens bipinnata*	6	Naturalized	Stage 3	SCS
Poaceae	*Bothriochloa humidicola*	6	Naturalized	Stage 3	SCS
Asclepiadaceae	*Calotropis gigantea*	6	Naturalized	Stage 3	SCS
Poaceae	*Cenchrus echinatus*	6	Naturalized	Stage 3	SCS
Poaceae	*Chloris babata*	6	Naturalized	Stage 3	SCS
Solanaceae	*Datura species*	6	Naturalized	Stage 3	SCS
	Delonix regia	6	Naturalized	Stage 3	SCS
Pontederiaceae	*Eichhornia crassipes*	6	Naturalized	Stage 3	SCS
Cactaceae	*Opuntia stricta*	6	Naturalized	Stage 3	SCS
Poaceae	*Paspalum paspaloides*	6	Naturalized	Stage 3	SCS

Table 3 *Continued*

Family	Species	Cluster group	Richardson *et al.*	Colautti & MacIsaac	Davis & Thompson
Portulacaceae	*Portulaca pilosa*	6	Naturalized	Stage 3	SCS
Caesalpiniaceae	*Senna obtusifolia*	6	Naturalized	Stage 3	SCS
Caesalpiniaceae	*Senna occidentalis*	6	Naturalized	Stage 3	SCS
Malvaceae	*Sida subspicata*	6	Naturalized	Stage 3	SCS
Solanaceae	*Solanum nigrum*	6	Naturalized	Stage 3	SCS
Solanaceae	*Solanum torvum*	6	Naturalized	Stage 3	SCS
Poaceae	*Sporobolus indicus*	6	Naturalized	Stage 3	SCS
Zygophyllaceae	*Tribulus terrestris*	6	Naturalized	Stage 3	SCS
Asteraceae	*Tridax procumbens*	6	Naturalized	Stage 3	SCS
Poaceae	*Urochloa mutica*	6	Naturalized	Stage 3	SCS
Poaceae	*Chloris gayana*	7	Naturalized	Stage 3	SCG

2000), or Stage 5 species (Colautti & MacIsaac, 2004). Similarly, 31 minor species were all classified as naturalized (Richardson *et al*., 2000), short-distance/common/great (Davis & Thompson, 2000), or Stage 3 species (Colautti & MacIsaac, 2004). However, it was difficult to summarize the classifications for the remaining intermediate 30 species. For example *Ziziphus mauritiana* was classified as a transformer species (Richardson *et al*., 2000), a short-disperser, common with great impact (SCG) (Davis & Thompson, 2000) and at Stage 4b (Colautti & MacIsaac, 2004). In contrast, *Jatropha gossypifolia* was classified as naturalized (Richardson *et al*., 2000), while classified as a short-disperser, common with great impact SCG (Davis & Thompson, 2000) and at Stage 4b (Colautti & MacIsaac, 2004) (Table 3).

Non-metric multidimensional scaling

The first two dimensions accounted for 98.8% of the variation with a stress of 0.094. The first dimension primarily separated species on total cover, where the score was negatively correlated with total cover, cover per site, cover per transect and cover per quadrat (r = –0.86, –0.77, –0.95 and –0.95, respectively). *Urochloa mosambicensis* was the most abundant species. Other species with negative scores for the first dimension included the woody species, *Ziziphus mauritiana*, *C. grandiflora* and *Lantana camara*. The grasses *Panicum maximum*, *C. ciliaris* and *B. pertusa* all had negative scores for the first dimension, as did the forb *Parthenium hysterophorus*. The small forbs, *Sida cordifolia*, *Alternanthera bettzikiana*, *Xanthium strumarium* and a native perennial shrub *Carissa ovata* were the only other species with negative scores for the first dimension.

The second dimension separated species on site frequency. *Ziziphus mauritiana* and *Panicum maximum* had low scores for the second component, suggesting they occupied relatively few sites. Conversely, *S. cordifolia*, *C. grandiflora* and *C. ovata* had high scores on the second dimension, corresponding to a high site frequency (Fig. 2a,b).

The first and second dimensions captured the main effects present in the data. Minor species had small positive scores on the first axis and small negative scores on the second. It is apparent from the correlations between the attributes and the axis scores that these species had low total cover and low site frequency.

Agglomerative Clustering

Seven groups were identified by the agglomerative hierarchical cluster analysis (Table 4). The groups agglomerated species based on their total cover and site frequency and the groupings concur with the output derived from the non-metric multi-dimensional scaling (Figs 2c & 3). Detailed characteristics and species membership of each group are shown in Table 3. The first three groups, with 1, 2, and 16 species had high total cover and were present in 69, 27, and 39 sites, respectively. All three

Table 4 Summary of species group characteristics derived from the hierarchical cluster analysis

Group	No. spp.	Total cover (%)	Cover/ quad (%)	Cover/ transect (%)	Cover/ site (%)	Quadrat frequency/site	Transect frequency/site	Site frequency
1	1	7.50%	0.19	0.29	0.09	42	2.68	69
2	2	1.90%	0.23	0.23	0.06	23	2.33	27
3	6	1.16%	0.10	0.10	0.02	23	2.16	39
4	12	0.16%	0.02	0.01	0.00	12.57	1.88	48
5	18	0.02%	0.02	0.01	0.00	7.24	1.60	20
6	26	0.001%	0.00	0.00	0.00	1.72	1.10	4
7	1	0.02%	0.29	0.06	0.01	2.00	1.00	3

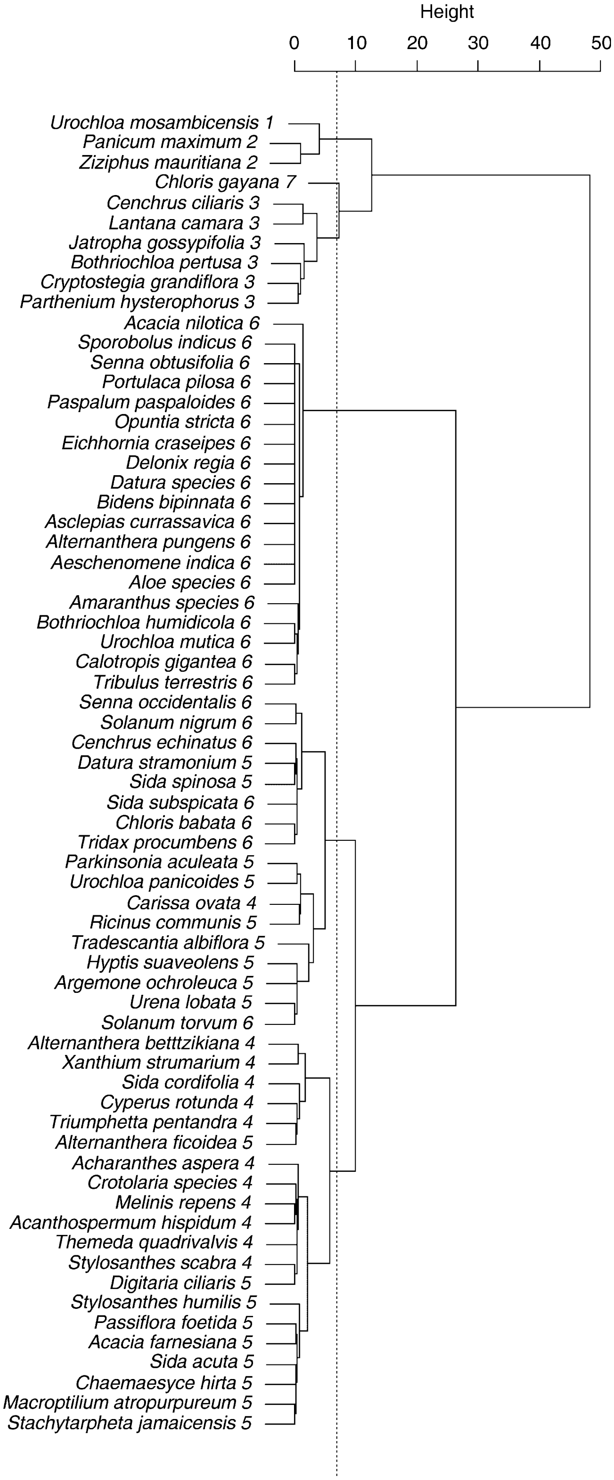

Figure 3 Dendrogram from the hierarchical cluster analysis indicating the cut height and group membership of individual species.

groups were widely dispersed throughout the site. Group 1 species or *U. mosambicensis*, occupied 42 quadrats per site when present, while groups 2 and 3 species occupied 23 quadrats per site when present. Groups 1 and 2 also occupied approximately 20% of the quadrat when present, while species in group 3 occupied 10% of the quadrat.

The remaining four groups, with 12, 18, 26, and 1 species, respectively, occupied negligible proportions of the catchment, although group 4 species were widespread and present in 48 sites. All four groups were rare, even when present at the site, as species were found in few quadrats per site, and groups 6 and 7 species were generally found in just one of the nine transects surveyed per site.

Two species, *U. mosambicensis* and *Chloris babata*, were outliers in this survey, because of their very high site and quadrat per site frequency (*U. mosambicensis*) and very high cover per quadrat but low site frequency (*C. babata*). As a result, they were excluded from all other groups.

Categorical data, such as life form, were ignored in this study and grouped species often had different ecological characteristics. For example, grasses, forbs, vines, and trees were all found in cluster groups 3, 4, 6, and 7 (Table 3). Other information, such as time since introduction, was either unavailable or not applicable to this catchment as species were usually first introduced elsewhere in Australia. We did not consider it necessary to further separate the few species present at a large number of sites and therefore discounted information on the spatial aggregation of individual species.

Relationship between attributes

The two site level measures of frequency, transect frequency per site and quadrat frequency per site, were moderately correlated with the landscape scale measure, site frequency ($r^2 = 0.50$ and 0.36, respectively). Similarly, two of the three site-level measures of cover, cover per transect and cover per site, were highly correlated with total cover ($r^2 = 0.74$ and 0.81, respectively). Cover per quadrat was only moderately correlated with total cover ($r^2 = 0.36$). The relationship between attributes and the dimensions from the multivariate analysis are presented in Fig. 2(b). In general, there was an orthogonal relationship between site frequency and cover per quadrat, cover per transect, and cover per site. This suggests that the area occupied by a species, captured by the various measures of cover, was unrelated to the processes governing its dispersal at the catchment scale, as represented by site frequency. However, dispersal within site, measured by quadrat per site and transect per site, was correlated with cover per site and cover per transect. Collectively, these relationships suggest a dichotomy between the processes involved in long-distance dispersal between sites and shorter dispersal distances for within site measures.

DISCUSSION

Weed classification schemes are designed to provide clarity and consistency to the field of invasion ecology through the use of broad, generally applicable terminology and concepts for species invasions. These frameworks put forward a series of hypotheses about the process of invasion and are intended to enable ecologists to qualitatively assess characteristics of invasive species and group them concordantly. The three frameworks require different criteria to perform the classifications. Application of the frameworks is further complicated by the subjective interpretation of classification criteria and implementation of the classification systems. Taxonomic classification systems, such as soil classifications (e.g. Isbell, 1996) or land systems (e.g. Tongway & Hindley, 1995), traditionally use a key that can be repeatedly applied to the soil or to a landscape. The classification systems employed in the present study have not been developed to the extent of soil or landscape classification systems, possibly because the data necessary to evaluate their performance was unavailable. Their application is invariably subjective. For example, if we had considered species occupying more than 20 sites to be 'widespread' (Colautti & MacIsaac, 2004) and 'long-distance dispersers' (Davis & Thompson, 2000), then the number of species categorized as Stage 5 widespread and abundant (Colautti & MacIsaac, 2004) and long-distance, common with great impact, would have increased from five to eight species. Every application of these frameworks is likely to be different and arguably site and scale specific, with different interpretations of what is frequent, abundant, local, common, high-impact, or long-distance dispersal. The frameworks were also unable to make use of additional available data, such as the cross-scale indices used in this study. Given these limitations, it seems unlikely that widespread application of these frameworks, in their current form, can advance the field of invasion ecology. This may be because each framework has a discrete set of hypotheses about what constitutes a serious invader. The suitability and application of these hypotheses to a localized problem, with unique or inappropriate data, may not always be relevant for managerial purposes.

There are elements of subjectivity in application of the clustering algorithms to the frequency and abundance data. This subjectivity exists at two levels. First, the decision to cut the dendrogram at a height of seven is arbitrary. Cutting at a height of 12 would have resulted in just three groups and yielded groupings synonymous with those derived from the classification frameworks. For example, groups 2 and 3 would have been amalgamated, even though group 3 species were present, on average in 12 more sites but occupied just 10% of the quadrat when present. In contrast, group 2 species occupied 23% of the quadrat. We consider it productive to preserve this variation in the data since it potentially reflects key differences in modes of invasive spread that should be considered in management. Thus, group 2 species, while having the capability to become regionally common, probably through a greater capacity for long-distance spread, rarely become locally dominant. In contrast, group 3 species appear to spread well locally and become dominant where they occur but at present have not dispersed as widely. In practical terms, management strategies could target group 3 species because they have yet to disperse widely. They may also pose a greater threat to the ecosystem simply because they occupy a larger portion of the quadrat and transect when present. Alternatively, group 2 species have demonstrated proficiency for dispersal. If this trend continued, these species could threaten a much larger area, but with slightly lower levels of cover than group 3 species. Group 4 species displayed a similar penchant for dispersal, but occupied very minor portions of the transect or quadrat. These are arguably the hardest species to classify with confidence. Because they are successful dispersers, they have the potential to threaten the ecosystem, although at this point in time their effect is minor. Our interpretations of the effect of species from group 5 and group 6 are also based on the frequency and abundance data at this point in time. Their presence and overall effect in the catchment are minor. However, our analyses and interpretations are limited by the data available. Additional information, such as time since introduction into the catchment may alter the assessment, but these data were unavailable for most species. Some species may still be adapting, or may not have had the opportunity to establish a sufficiently large population to create a problem. Individual species in groups 4, 5, and 6 may indeed be sleeper weeds and become a greater problem than they are. This concept was recently reviewed by Grice & Ainsworth (2003).

In addition, the cluster analysis, detected two extreme outlier species, group 1, *U. mosambicensis*, and group 7, *C. babata*. *Urochloa mosambicensis* was introduced as a pasture grass in the 1950s and the observed frequency and abundance reflect its importance to the grazing industry as it would have been deliberately planted by graziers since its introduction. *Chloris babata* was only found in three sites, but, on average occupied 29% of the quadrat and 6% of the transect, suggesting that this species could have a dramatic impact on the landscape.

The other element of subjectivity lies with the algorithms used to assess the dissimilarity between species attributes and the algorithms used to combine species based on their dissimilarities (e.g. Pielou, 1984). In this study, Euclidean metric scaling was applied to normalized ordinal data, ensuring that each attribute would contribute equally to the group structure. Alternative distance measures would be required if data were categorical. Clustering algorithms have been widely used and although elements of subjectivity remain, they can be applied to a weed data set and discriminate between species at some level. They use whatever data are available and help formulate hypotheses about the invasion in the landscape of interest.

We used the classification frameworks in the first instance to provide baseline parameters for a mixture method of clustering using the software developed by Fraley & Raferty (2002). Data were not multivariate normal and the resulting classifications were influenced by the initial group structure and often identified local optima (Fraley & Raferty, 2002). This approach may be adopted in the future if the methods are modified to cope with Poisson distributions, such as those encountered here.

The extensive survey data used to derive the landscape indices may be time consuming to collect in some landscapes. The similarities between the seven variables imply that meaningful classifications of invasive species can be achieved with a reduced sampling effort. The simple site level indices derived in this study also provide insights into how invasive a species may become at a larger scale. The particularly strong relationship between cover per transect and total cover in the landscape suggests that, if a species consistently occupies more than 10% of a transect in a specific habitat, then it has the potential to impact and invade that habitat. The link between these scales is worthy of further study to determine the relationship between local and regional impact by species at a certain stage in the invasion process. It may be argued that a species that occupied even a small portion of the landscape in one location has overcome many of the barriers to colonization identified in Richardson *et al.*, 2000's framework and has the potential to invade similar habitats, providing a dispersal mechanism exists.

Both quantitative and qualitative forms of classification are developed to enable researchers and managers to classify an invasion and determine which species have dispersed widely and had the greatest impact on the ecosystem. Both have limitations. The classification techniques provide a means for classifying an invasion based on a series of hypotheses that describe an invasion. In contrast, the multivariate techniques enabled us to partition species into meaningful groups, based on observed patterns within the data set. This data set contained information on species frequency and cover at a range of spatial scales. We suggest that the subjective, qualitative approaches developed by Richardson *et al.*, 2000, Davis & Thompson, 2000, and Colautti & MacIsaac, 2004 be used in conjunction with multivariate techniques developed here to derive classifications that are data driven, but spatially and temporally specific.

A classification scheme should ensure that species within a group have attributes that are more similar than those in an adjacent group. This concept may be extended further, where the species in the same group should be subjected to similar levels of management. The subjective use of thresholds that often separate species using a single attribute, arrived at non-sensical classifications when all the data are considered and viewed in multivariate space (Fig. 2c–f). This may be because the thresholds chosen were inappropriate, or the frameworks lacked the capacity to make the best use of the data available. The frameworks had a tendency to under-classify the data.

The other purpose of these analyses was to assess the characteristics of the data to determine which species have similar levels of frequency and abundance across a range of spatial scales. From this information, management strategies could be developed for as few or many groups as necessary. Outliers may emerge in this analysis and warrant preferential treatment.

CONCLUSION

Our use of a simple cluster analysis to objectively group species with similar population-level distributional characteristics has several advantages over classification using any of the three individual frameworks. Most importantly, it is repeatable, using data alone to drive the classification and assigning species to like groups independently of preconceived notions of their relative invasiveness. Furthermore, it is regionally and temporally specific, lending itself well to a management application. A strong advantage also lies in the potential for integration of the outcomes of our analyses with current frameworks for assessing regional-scale dynamics of native species.

ACKNOWLEDGEMENTS

We thank Lindsay Whiteman, Mike Nicholas and Brett Abbott for assisting with fieldwork and Professor Iain Gordon for constructive comments on earlier versions of the manuscript. We also thank the CRC for Australian Weed Management and CRC for Tropical Savannas for funding various components of the study.

REFERENCES

Caley, P. & Kuhnert, P. (2006) Application and evaluation of classification trees for screening unwanted plants. *Austral Ecology*, in press.

Colautti, R.I. & MacIsaac, H.J. (2004) A neutral terminology to define invasive species. *Diversity and Distributions*, **10**, 135–141.

Davis, M.A. & Thompson. K. (2000) Eight ways to be a colonizer, two ways to be an invader: a proposed nomenclature scheme for invasion ecology. *ESA Bulletin*, **81**, 226–230.

Fraley, C. & Raftery, A.E. (2002) *MCLUST: Software for model-based clustering, density estimation and discriminant analysis*. Technical Report, Department of Statistics, University of Washington, Seattle, WA. URL: http://www.stat.washington.edu/mclust, accessed 3 February 2006.

Gordon, A.D. (1999) *Classification*. 2nd edn. Chapman & Hall, London.

Grice, A.C. & Ainsworth, N. (2003) Sleeper weeds — a useful concept? *Plant Protection Quarterly*, **18**, 35–39.

Grice, A.C., Radford, I.J. & Abbott, B.N. (2000) Regional and landscape-scale patterns of shrub invasion in tropical savannas. *Biological Invasions*, **2**, 187–205.

Isbell, R.F. (1996) *The Australian soil classification*. CSIRO Publishing, Collingwood, Australia.

Lonsdale, M. (1999) Global patterns of plant invasions and the concept of invasibility. *Ecology*, **80**, 1522–1536.

Mack, R.N. (1996) Predicting the identity and fate of plant invaders: emergent and emerging approaches. *Biological Conservation*, **78**, 107–121.

Murphy, H.T., VanDerWal, J., Lovett-Doust, L. & Lovett-Doust, J. (2006) Invasiveness in exotic plants: immigration and naturalization in an ecological continuum. *Conceptual ecology and invasions biology: reciprocal approaches to nature* (ed. by M.W. Cadotte, S.M. McMahon and T. Fukami), pp. 65–105. Springer, Dordrecht.

Perrins, J., Williamson, M. & Fitter, A. (1992) Do annual weeds have predictable characteristics? *Acta Oecologia*, **13**, 517–533.

Pheloung, P.C., Williams, P.A. & Halloy, S.R. (1999) Weed risk assessment. A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *Journal of Environmental Management*, **57**, 239–251.

Pielou, E.C. (1984) *The interpretation of ecological data*. John Wiley and Sons, New York.

Pyšek, P., Richardson, D.M., Rejmánek, M., Webster, G.L., Williamson, M. & Kirschner, J. (2004) Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. *Taxon*, **53**, 131–143.

Rabinowitz, D. (1981) Seven forms of rarity. *The Biology Aspects of Rare Plant Conservation* (ed. by H. Synge), pp. 205–217. John Wiley and Sons, New York.

Rejmánek, M. (1995) What makes a species invasive? *Plant Invasions — General Aspects and Special Problems* (ed. by P. Pyšek, K. Prach, M. Rejmánek and M. Wade), pp. 3–13. SPB Academic Publishing, Amsterdam.

Rejmánek, M., Richardson, D.M., Barbour, M.G., Crawley, M.J., Hrusa, G.F., Moyle, P.B., Randall, J.M., Simberloff, D. & Williamson, M. (2002) Biological invasions: politics and the discontinuity of ecological terminology. *ESA Bulletin*, **83**, 131–133.

Richardson, D.M. & Pyšek, P. (2006) Plant invasions: merging the concepts of species invasiveness and community invasibility. *Progress in Physical Geography*, **30**, 409–431.

Richardson, D.M., Pyšek, P., Rejmánek, M., Barbour, M.G., Panetta, F.D. & West, C.J. (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distribution*, **6**, 93–107.

Rossiter, N.A., Setterfield, S.A., Douglas, M.M. & Hutley, L.B. (2003) Testing the grass-fire cycle: alien grass invasion in the tropical savannas of northern Australia. *Diversity and Distributions*, **9**, 169–176.

Tongway, D. & Hindley, N. (1995) *Manual for Assessment of soil condition of tropical grasslands*. CSIRO Division of Wildlife and Ecology, CSIRO Australia, Canberra, Australia.

Williamson, M.H. (1999) Invasions. *Ecography*, **22**, 5–12.

Williamson, M.H. & Fitter. A. (1996) The characters of successful invaders. *Biological Conservation*, **78**, 163–170.