CHAPTER 10

Risk Assessment for the Traditional Medicine Trade

ASSESSING EXTINCTION RISK AND CONSERVATION PRIORITY OF SPECIES HARVESTED FOR THE MEDICINAL PLANT TRADE ON THE WITWATERSRAND

CHAPTER SUMMARY

The chapter explores the use of a multivariate methodology for assigning species harvested for the medicinal plant trade to various hierarchies that describe their risks of extinction and conservation priorities. Groups of species of varying risk profiles were identified from the individual surveys of the *muti* shops and Faraday market using cluster analysis and 4-5 trade variables. However, it was from combining the two data sets to form a compound list of 392 ethnospecies that a short-list of 119 higher-risk ethnospecies was identified. This list was further reduced to 87 species for further analysis to ascertain conservation priorities based on the additional inclusion of several biological variables. Species not on the short-lists are not considered to be *as* threatened by harvesting for the medicinal plant trade; however, they may be at risk due to other deterministic and stochastic factors and extra pressures from the *muti* trade may exacerbate the problem. From the list of 87 higher-risk species, ≈31 species were concluded to be of higher conservation priority.

CHAPTER OUTLINE

10.1 Introduction

Over-exploitation is a deterministic factor in the extinction risks to species. Risk assessments (RAs) are necessary so that species can be categorised according to their relative risks and conservation priorities can be set (Frankham *et al.* 2004). There are various tools available to assess the risks and conservation status of species and/or populations. For single species assessments, Population Viability Analysis (PVA) is a form of risk assessment model traditionally used to model the combined impacts of deterministic and stochastic factors to estimate the minimum viable population size and hence predict extinction risks (Burgman *et al.* 1993; Pfab and Witkowski 2000; Frankham *et al.* 2004; O'Grady *et al.* 2004). Additionally, the IUCN Red List categories of threats to species (e.g. Critically Endangered, Endangered or Vulnerable) rank the risks of extinction according to a set of predictive criteria that inform the urgency of conservation measures if conditions that endanger a species prevail (de Grammont and Cuarón 2006; IUCN 2006).

Conventional risk assessment processes do not appear to have been broadly applied to assessing the impact of harvesting plants for the traditional medicine trade. Some species used as traditional medicines have been individually evaluated according to Red List criteria, for example, *Mondia whitei* is categorised as VuA1dD2, i.e. Vulnerable with a high risk of extinction because actual or potential levels of exploitation have resulted in an observed, estimated, inferred or suspected population size reduction of ≥50% in 10 years or 3 generations (Victor 2002). However, protocols such as the Rapid Vulnerability Assessment (RVA) (Wild and Mutebi 1996, Wong *et al.* 2001) and a linear scoring system based on numerical importance values (NIVs) have been used to rank and short-list groups of species vulnerable to over-exploitation.

Risk assessment usually precedes conservation priority setting. However, an approach taken by the IUCN's Medicinal Plant Specialist Group (MPSG) to identify 'flagship' species for conservation priority is the "Top 50" action plan (Cunningham 1996a, 1997). The protocol is based on a five step system that progressively short-lists large groups of species until the top 50 are determined. This procedure was followed by Özhatay *et al.* (unpublished report cited in Lange 1998) to identify the top 50 native Turkish medicinal and aromatic plants most threatened by collection owing to their demand in trade.

The aim of my research was to develop and recommend a hybrid risk assessment method for medicinal plants that incorporated aspects of multivariate statistical models traditionally used in ecological risk assessment, with the RVA and 'Top 50' methods which incorporate indigenous knowledge and scientific data, to identify groups of species that have similar profiles of relative risk, threat and vulnerability to over-utilisation and over-harvesting by the trade. The assessment process integrates features of the trade with selected biological and ecological traits of the species being exploited. Given that some of the criteria used for assessing the risks are similar to criteria used for setting CP, the assessments of species are also partly descriptive of the priorities that should be given to their conservation. While the risk or threat levels proposed for species do not specify the time-line for extinction, they do categorise the probability of partial or total loss of a population if current harvesting trends continue. This process could therefore be used to support the Red List and PVA procedure, and *vice versa*. A further goal was to use variables for which data were accessible and readily available for most species in order to make expeditious assessments.

The chapter begins by describing the statistical, multivariate approaches that have been used to evaluate risk. Thereafter, variables used in other assessments are examined and the variables used in this study are described. The risk analysis then follows a five step process that progressively short-lists and analyses groups of species using various multivariate techniques and variables (a summary of this process is shown later on in Boxes 1 and 2).

10.2 Statistical, Multivariate Approaches for Evaluating Risk

Difficulties with linear ranking schemes for assessing threats and setting priorities include: (a) artificial linearity is imposed onto naturally non-linear biological systems; (b) the schemes are statistically unsatisfactory; (c) many species tend to rank together, making it difficult to separate them; (d) the reasons for a particular species being threatened are concealed; and (e) the total scores may be artefacts since the individual criteria used to derive the scores are not necessarily equivalent or independent (Given and Norton 1993). However, the individual scores do provide a 'threat or priority' profile for each species that would make it possible to *group* species together based on similar profiles (Given and Norton 1993). If threatened species are regarded as lying somewhere in a multidimensional or multivariate 'web' or hyperspace of factors, it is often the job of conservationists to detect where a species is in that hyperspace and to try and shift them out of their respective danger zones by applying management techniques (Hall 1993). Use of multivariate techniques to assess the relative arrangement/position of species in the hyperspace (Given and Norton 1993, Hall 1993) facilitates identification of groups of species with similar risk/threat/ priority profiles, as well as the factors that reveal that risk.

There are a variety of multivariate techniques and strategies that have been proposed for comparative risk assessment and priority setting for threatened species. Cluster analysis and ordination (e.g. Principal Components Analysis) are regarded as two of the most effective ways of identifying groups of species with similar risk profiles in relation to the criteria characterising the risk (Given and Norton 1993, Hall 1993, Ding *et al.* 1996, Myers 1999, Simeonov 2003). In addition, the methods provide an opportunity to explore the influence of, and the relationships between, the various criteria and their value in informing risk identification and characterisation. Because several variables can be considered simultaneously, interpretations can be made that are not possible with univariate statistics (James and McCulloch 1990).

10.2.1 Cluster Analysis

Cluster analysis is a classification method used to obtain groups of objects (cases, species or samples) that are internally homogeneous and distinct from other groups (Lepš and Šmilauer 2003). Homogeneous clusters then become the unit of analysis rather than individual observations (Punj and Stewart 1983). The methods are appropriate as a data reduction technique where there are large numbers of cases (e.g. species), no real dependent variable and it is desirable to determine whether groups of similar cases exist. The cluster sequencing process is either *hierarchical* or *non-hierarchical*. In *hierarchical* classifications, groups (clusters) are formed that contain sub-groups so that there is a hierarchy of levels (Lepš and Šmilauer 2003) that can be displayed by a dendrogram [a two-dimensional hierarchical tree diagram representing the complex multivariate relationships among the objects (James and McCulloch 1990)]. The aim of *non-hierarchical* clustering is to classify the objects into a predetermined number of internally homogenous groups (clusters) without the relationships being established between them, i.e. there is no hierarchy and no dendrogram can be produced (Höft *et al.* 1999). *K-means clustering* is a non-hierarchical classification method.

Hierarchical classifications are either *divisive* or *agglomerative*. In *divisive* cluster analysis, the classification starts with the division of the data set into two groups, which are further split and redivided until all the objects have been grouped (Höft *et al.* 1999, Lepš and Šmilauer 2003, Simeonov 2003). TWINSPAN is an example of divisive analysis (Höft *et al.* 1999). In *agglomerative* cluster analysis, groups are formed sequentially by joining the two most similar objects to form hierarchies of successively larger groups until all the objects are joined. Agglomerative hierarchical cluster analyses are also defined by the distance measure used to gauge object dissimilarities or distance (e.g. Euclidean distance) and the hierarchical algorithm used to link/amalgamate the groups (e.g. single linkage, complete linkage, Ward's minimum-variance method).

In applying these analyses to risk assessment (RA), and specifically to species with different levels of threat to over-harvesting, relatively homogeneous groups are produced where species within a group/cluster have similar risk profiles and/or are at risk for similar reasons (or, have similar conservation priority profiles if using criteria in the analyses that characterize CP levels e.g. Hall 1993). What matters most in the resulting dendrograms is which species are members of which clusters (*not the order on the base line of the dendrogram*) and the degree of difference between clusters (given by the distance or similarity measure).

In addition to describing the pattern of relationships among objects (species and variables) by *cluster analysis* (the classification of the objects into hierarchical or non-hierarchical categories on the basis of a matrix of inter-object similarities), patterns by *ordination* can also be described, i.e. the reduction of a matrix of distances or similarities among the attributes or among the objects into one or a few dimensions (James and McCulloch 1990).

10.2.2 Ordination

When faced with data that contain sets of correlated variables, the variables could be analysed separately using regression analysis – the goal being to find the dependence of individual responses on environmental variables (Lepš and Šmilauer 2003). Often, however, the joint and simultaneous consideration of the variables can produce interpretations that are not possible from sets of univariate statistics (James and McCulloch 1990). Ordination is the collective term for linear- and non-linear multivariate techniques that arrange objects (samples, species and environmental variables) along axes based on the correlation of the variables with the axis. The objects are usually displayed in a two-dimensional ordination diagram, such as a biplot, constructed in such a way as to optimize the linear or non-linear (unimodal) fit to multiple variable data along axes that are gradients of attribute combinations. The importance of an axis is indicated by its eigenvalue, and the length and direction of the resulting vectors (in PCA) indicate the direction of maximum variation in the value of the corresponding variable (ter Braak 1988). By reducing the dimensions of multivariate data, redundant variables can be screened out or more insightful variables can be found as a preliminary step to further analyses (Höft *et al.* 1999). Although the abstract patterns that may emerge from ordination may suggest discrete assemblages of objects, the product of ordination is not usually one of classification (Ludwig and Reynolds 1988). Classification may, however, secondarily follow from the results of an ordination.

As indicated, ordination methods may be linear (e.g. Principal Components Analysis or PCA) or non-linear (unimodal) (e.g. Detrended Correspondence Analysis or DCA). PCA reduces a matrix of object similarities and allows identification of optimal linear combinations of the variables (the 'principal components'). In addition, PCA is a method for partitioning a resemblance matrix into a set of perpendicular axes or components (Ludwig and Reynolds 1988). Each PCA axis corresponds to an eigenvalue of the matrix – an eigenvalue being the measure of importance of an axis, usually expressed as a fraction or percentage of the total variance of the variable data (ter Braak 1987). The higher the eigenvalue, the more important the axis.

The choice between linear and non-linear ordination methods is not a matter of personal choice (ter Braak and Prentice 1988). If the relationships between the variables are linear, then PCA is advised. Similarly, if the relationships are non-linear then unimodal methods such as DCA and Correspondence Analysis (CA) should be used. When deciding which ordination method should be used, Lepš and Šmilauer (2003) recommend first carrying out a DCA of the data. If, at the end of the analysis, the largest value of the longest gradient is >4.0, then unimodal methods (DCA, CA) should be used. Use of a linear method would not be appropriate since the data are too heterogeneous and too many species deviate from the assumed linear response model. If, however, the longest gradient is <3.0 then PCA is a better choice (Lepš and Šmilauer 2003). If the length of the longest gradient ranges between 3 and 4 then both ordination methods will work.

One purpose for using multivariate linear ordination techniques in RA, specifically PCA, is to identify which species are positively or negatively correlated with the criteria selected to evaluate risk, as well as the importance of the criteria. Ordination has the potential to show that not all criteria used to score species are equally important and that some criteria might be redundant (Given and Norton 1993).

10.3 Selection of Suitable Variables and the Source of Data used in the Risk Evaluation Process

There are three stages to analysing potential risks once the stressor (in this case, the commercial harvesting of plants) has been identified, namely: input, process and output (Ding *et al.* 1996). In the input stage, evaluative variables/criteria/factors are chosen that will best quantify the relative risks between the species. Thereafter, the appropriate univariate or multivariate statistical methods are chosen to 'process' the variables in order to determine the output – namely the level of risk to species harvested for the medicinal plant trade. The multivariate statistical methods often used in risk assessment have been described in the previous section. Here the evaluative criteria that will be used to establish the risks are considered.

Part of the challenge of the RA procedure is selecting suitable parameters and criteria that will elucidate the potential risks to species. In a study on German pteridophytes for example, the capability of ferns and fern allies to colonise secondary habitats was assessed to be an important factor in predicting the chance for survival of a species under a permanently changing environment (Bennert 2000). However, a further challenge is making the trade-off between (a) capturing all unique functions and values to categorise the risks and (b) using the variables to group species into a *few* assessment endpoints (Myers 1999). The former can be prohibitively complex (and costly) and the latter ignores biological complexities that may be crucial for maintaining biodiversity and optimal ecosystem functioning (Myers 1999). One must also be cognisant of not using criteria in an RA that are only appropriate for assessing CP, albeit there is overlap in the parameters used to categorise both.

In an ideal situation, there are many kinds of information that can be used in the RA process, including: demographic, habitat and spatial data; functional and physiological data; market indicators; perception and knowledge of resource users; utilisation data; genetic and stochastic data. Stochastic data include information on the variability of age, stage and sex-specific fecundities (Burgman *et al.* 1993). However, the factors selected depend upon the requirements of the study (i.e. a threat and/or priority assessment), the methods used to assess the risks, and the availability and cost of the information. Ideally, an aim should be to select variables that maximise the information that can be gained from variables that are not too costly or time consuming to collect. Variables that are easily and readily available enable expeditious risk assessments of large numbers of species. Thereafter, criteria for studying a smaller number of species in more detail can be selected to determine the exigency of conservation action.

Cunningham (1996a, 1997) proposed a five step process for determining which medicinal plants require conservation action. These steps are recommended by the IUCN MPSG in the "Top 50" plant campaign. The steps progressively short-list species with increasing conservation priority. The steps are: (1) identify the major sales sites or markets (e.g. local, regional, international); (2) identify medicinal species in trade; (3) prepare a short-list of species in trade which are slow-growing, destructively harvested, considered scarce and which are popular and/or expensive and/or sold in large volumes/numbers; (4) short-list these species further on the basis of commonness or rarity using characteristics of their geographic distribution (wide vs. narrow), habitat specificity (broad vs. restricted) and population size (somewhere large vs. everywhere small); and (5) set conservation priorities within the short-list on the basis of the phylogenetic distinctness rating system proposed by Farjon (1995), e.g. monotypic family or genus.

Steps 1 and 2 have already been accomplished for this study, and the species traded in the Witwatersrand for traditional medicines have been identified (e.g. Williams *et al.* 2001, Williams 2003). The next step would be to prepare a short-list of species that are slow-growing, destructively harvested and perceived as scarce and/or popular. However, given the large number of species in trade in the region (>500) it is impractical to obtain data on growth rates for all these species, especially those that are at low risk for exploitation. Therefore, an additional step is proposed that short-lists species that are destructively harvested, perceived as scarce and/or popular *and* are very prevalent in the markets. This step would objectively partition the plants into smaller groups with different relative risks. Thereafter, variables such as growth rate, geographic distribution, habitat specificity and phylogenetic distinctness could be used to further short-list species for conservation priority.

Before discussing the selection of the variables, it is beneficial to review the methodology and variables used to assess threat and/or conservation priority of plants in other studies in both South Africa and abroad in order to identify criteria that could be used in this study.

10.3.1 Criteria Selected for Priority Evaluation in Previous Studies

In endeavouring to assess the threats and/or CP of species that are vulnerable to over-exploitation and population decline, several studies have used a point scoring or rating system to rank taxa and place them into categories of risk, threat or CP (Tables 1 and 2). The various studies incorporate both biological and/or socio-economic factors into the assessment methodologies. However, the criteria for dividing plants into management categories according to their priority scores are sometimes *ad hoc* (with the exception of Given and Norton 1993, which used a multivariate approach).

The criteria selected for measuring risk, and the scoring system used to rank the species, very much depends upon the size of the market being studied (i.e. is the study on a local, regional or national scale), and the number of species being investigated. For example, in the assessment of CP for plants in the Abe Bailey Nature Reserve, South Africa (Dzerefos and Witkowski 2001) it was necessary to use the density of plants in the reserve as an assessment variable. Selection of the appropriate criteria is therefore essential for the purposes of the assessments. Inappropriate criteria and methods could make a regional model too 'coarse' so that subtle differences in the risk between species are not detected.

Several variables should be used with discretion. First is *provincial distribution* (i.e. the number of provinces in which a species occurs). Political boundaries can change and render such measures as redundant or over-estimated. In South Africa, the number of provinces changed from four before the 1994 elections to nine after the elections. Second is *population size*. These data are not cost-effectively available for large data sets and often only known for species that are currently threatened. The variable is best used in the Red List process or, in comparative assessments for species in specified areas such as nature reserves or communal land. Third is *growth rate* or *period*. This is the time in years until utilisation or controlled exploitation of the species can commence (Loxton *et al.* 1994). This is not necessarily the age at which the individual becomes reproductive. Similar life forms (such as trees, geophytes and herbs) have similar values and any numerical or clustering threat assessments would be skewed towards species with the highest values, e.g. trees. Fourth is *diversity of use* or *number of uses*. This can be a nebulous measure because quantification of the number of uses may depend on sample size and effort (i.e. the number of users interviewed), the reliability of the information, how the diseases/ailments/uses are classified (i.e. the 'nosology') and, the availability of literature. While the number of uses can indicate the importance of a species to a user group (such as 'keystone species', Garibaldi and Turner 2004), it does not necessarily follow that more uses makes the species more at risk of over-utilisation and therefore decides conservation priorities at a national level. Given the effort required to obtain the use data, it is not a variable that can be cost-effectively used to analyse a large number of species. It is therefore more appropriate as a variable for smaller data sets, but then the purpose of using it to short-list species is subsequently obviated. Also to be used with caution are trade variables such as value per kilogram (e.g. Rand/kg) and the mass sold per sale, as these variables are very sensitive to plant part types (Chapter 9, Williams *et al.* in press).

Farjon (1995) proposed additional criteria for assessing conservation priorities beyond the IUCN's 1994 Red List Categories for the conservation status of species. The scheme uses a point system that values each taxon according to a point scale under three criteria. The three criteria are: *phylogenetic distinctness*, where monotypic families and genera rank highest and infraspecific taxa rank lowest; *level of endemism* and rarity, which ranges from a few known plants in one location to species that are rare and scattered over a considerable area; *ecological tolerance*, which ranges from taxa restricted to specialised habitats to ones that are widespread or occur in a variety of habitats.

Table 1: Examples of variables used in various studies in South Africa to determine threat and conservation priorities for indigenous plants used for traditional medicine. (CP= conservation priority)

Author	Cunningham (1988)	McKean (1993)	Mander & Quinn (1997)	Dzerefos & Witkowski (2001)
Study aim	Threat or vulnerability assessment for medicinal species over-exploited in KZN	Identify important utilised indigenous plant species in KZN	Create a priority list of species which are of conservation concern in South Africa	Assess the CP of medicinal plants present in the Abe Bailey Nature Reserve, South Africa
Method	Not specified, but takes into account all the variables below	Weighted ranking method based on numerical importance values. Species assigned ranked point scores.	Weighted numerical rating system. Variables assigned ranks and weighted scores, depending on the criteria. 3 broad criteria categories: utilisation, biology and market indicators.	Weighted numerical rating systems. Variables were assigned ranks and weighted scores in two categories: utilisation and biology
Variables / Factors / Criteria	- Raunkaier life form categories (e.g. geophyte, phanerophyte etc.) - Plant part used (e.g. bark, root, bulb etc.) - Demand (based on the estimated quantity of 50 kg-size bags sold annually by traders) - Distribution in the Province Two additional factors were used for trees, namely: - coppicing ability - sensitivity of species to bark removal	- Present utilisation of species (largely based on no. of units sold per year) - Unit value (Rand per unit in which species is sold) - Resilience to utilisation/ harvesting (related to plant part harvested and known harvesting effects) - Conservation importance rating based on criteria by Scott-Shaw (1992). Criteria include: local population size, geographic range, habitat specificity, threats, level of protection, endemicity	<u>Utilisation</u> - Harvesting risk (consequences of plant part type harvesting e.g. mortality or inhibition of growth) - Extent of use (commonness of the ailments treated) - Diversity of use (number of uses) <u>Biology</u> - Distribution (number of provinces in which a species occurs) - Phylogenetic distinctness - Conservation status (IUCN 1996 RL categories) <u>Market indicators</u> - Extent of use (localised, moderate or widespread) - Price (low, moderate or high) - Quantities traded (low, moderate or high) - Reported scarcity (never, seldom, periodic, frequent)	<u>Utilisation</u> - Harvesting risk (determined by plant part) - Use value (% traders listing species as used) - Diversity of use (number of uses) <u>Biology</u> - Density of plants in the reserve
Result	Six categories of vulnerability to exploitation were defined ranging from Extinct in the wild to Indeterminate.	An importance value for each species and 8 recommendations for further research, monitoring, management & public awareness action.	Separate priority lists of species based on utilisation and biology scores, as well as an integrated list based on the cumulative scores for all criteria.	Weighted conservation priority scores for species based on the cumulative scores for all criteria. 3 categories were defined requiring different management strategies.

Table 2: Examples of variables used in selected studies outside of South Africa to determine threat and conservation priorities for indigenous, mainly medicinal, plants.

Author	Given & Norton (1993)	Wild & Mutebi (1996)	Kala *et al.* (2004)	Evans *et al.* (2004)
Study aim	Assess threat and set conservation priorities for selected species in New Zealand	Rapidly assess the vulnerability of species used by people at two sites in Uganda	Compile a priority list of medicinal plants in India based on identified features	Assess priorities for conservation action of medicinal plants in an area in DPR Korea
Method	Species were scored based on 17 criteria. Cluster and DCA analyses were then used to divide the species into similar groups with similar threats and conservation priority.	Rapid Vulnerability Assessment: information rapidly (not numerically) evaluated according to the criteria	Numerical importance value based on weighted scores for the criteria.	Compile a list of candidate species believed to be harvested in significant quantities. Each species was then subjectively assigned to one of five priority classes based on the criteria.
Variables / Factors / Criteria	<u>Distinctiveness</u> - Taxonomic - Geographic <u>Population features</u> - No. of populations - Mean pop. size - Largest pop. size - Condition of largest population - Wild pop. decline rate <u>Vulnerability</u> - Legal protection of habitat - Habitat loss rate - Predator/harvest impact - Competition - Habitat specificity - Reproductive specificity - Other factors affecting survival <u>Potential</u> - Propagation / protection *ex situ* <u>Values</u> - Maori cultural values - Pakeha cultural values	- Raunkaier life form - Plant part used - Demand (indicated by the quantity harvested and frequency thereof) - Abundance and distribution (high vs. low and wide vs. narrow) - Habitat specificity (broad vs. restricted) - Response to harvesting (ability to regrow) - Growth rate - Pattern of selection and use (size/age specific harvesting) - Seasonal harvesting - Traditional conservation practises - Substitute availability - Commercial or subsistence use	- Endemism - Mode of harvesting (e.g. shoot and/or root) - Use value (no. of ailments treated) - CAMP status (e.g. rare, vulnerable, endangered or critically endangered)	<u>Threat status</u> -Global and national threat <u>Characteristics of the harvest</u> - size of market - observed decline in availability - unavailability (inferred from product substitution) - No. of different significant uses <u>Quantitative records of collection or confiscation</u> - known targets and harvest ceilings - opinions on relative size of unofficial harvest <u>Ecological indicators</u> - Life form, perennial vs. annual - organ harvested and re-growth potential - preferred habitat and successional stage - rarity (range and abundance)
Result	Species clustered into 5 groups with similar priority profiles	Species are assigned to one of eight management categories based on the criteria. Each category has a set of management recommendations.	The scores for each category were summed and the species prioritised accordingly into 3 priority groups.	Species assigned to one of 5 priority classes ranging from 'highest priority' to 'insufficient data'.

In addition to the studies listed in Tables 1 and 2, the Department of Environmental Affairs and Tourism (DEAT) has proposed a draft risk assessment framework for species threatened by various potentially harmful activities that threaten biodiversity. This framework has been proposed in accordance with Section 89 of the new National Environmental Biodiversity Act (NEMBA) (Act 10 of 2004), which makes provision for risk assessments to be conducted. The methodology for estimating and evaluating the risks to species have yet to be finalised, but some of the proposed variables include: the nature of the threat and the number of species and individuals affected by the activity; the status of species (e.g. national status, Red List etc.); population trends and numbers of the affected species; geographic distribution (current and extra-limital); habitat type and niche; known threats to the species both nationally and to the target population. DEAT's proposal is contentious, primarily because the framework would make the IUCN Red List process redundant and South Africa would be forced to use DEAT's criteria for assessing risk. Furthermore, the framework does not take into account risks due to land transformation and also proposes to use the same terminology as the IUCN criteria (e.g. Vulnerable, Critically Endangered), but with different meanings (W. Foden, pers. comm.).

10.3.2 Description of the Criteria Selected for this Study

There are often limitations to the amount and type of information that can be obtained due to constraints of data availability, resources, time and the type of research undertaken. Selection of the appropriate variables for assessing risk and determining conservation priority will also depend upon the size of the market studied (i.e. local, regional or national) or the study area (e.g. a nature reserve) and the number of species being investigated.

A notable limitation in choosing certain variables was the availability of that information for *all* the species under investigation; hence, the intention was to select variables for which data were available for most species. Furthermore, using too many variables increases the 'noise' and variance in the multivariate analysis. Fourteen variables were finally considered for the risk and priority assessment process. Some of the variables were only used in the latter stages of the analyses. The trade variables were used in the preparation of the short-list of species from the total list of species traded to indicate risk hierarchies. Thereafter, biological variables were included to identify species of greater conservation priority (See Boxes 1 and 2, page 16-17).

a) Trade variables

The trade variables represent quantitative data that can easily be collected during surveys of *muti* shops or street markets. These variables (used throughout the analyses) describe the prevalence, popularity, scarcity and volume of species sold by traders, namely:

- The number of traders selling a plant [shops (S) and/or Faraday street traders (F)];
- The number of traders citing a plant as scarce;
- The number of traders citing a plant as popular;
- The total volume present per plant during a survey (i.e. number of 50kg-size bags) (data available for the 2001 Faraday survey only).

The reason for using citations of plant scarcity and popularity as a model variable is that the perceptions of local resource users are known to provide valuable insights into the actual scarcities of useful plant species, and can thus be used in the development of conservation and resource management protocols for key species identified as vulnerable (Cunningham 1996b). Plant scarcities reflect the traders' real or perceived difficulties in obtaining the resources. If scarcity increases, then the distance travelled to harvest the resources will usually increase and there will sometimes be corresponding increases in the wholesale and retail price of plants.

Data were also collected for four other trade variables, but were only used to analyse a group of 22 specially selected species (see also Chapter 9 and Section 10.5.9):

- Mean Rand per kilogram selling value (*R/kg*)
- Mean quantity sold per sale (*kg/S*)
- Mean number of sales per annum (*S/a*)
- Mean number of bags bought per annum (*Bags/a*).

These variables require an adequately large sample size, thereby limiting the number of species for which data can be collected.

b) Biological variables

Except for *Plant part*, which was used in nearly all of the analyses, the biological variables were used in the analysis of conservation priority for the short-listed species. The six biological variables are:

- Plant part, as an indicator of the degree of destructiveness of harvesting
- Phylogenetic distinctness (after Farjon 1995)
- Endemism (after Farjon 1995)
- Number of quarter-degree grids ('QGrids')
- Habitat specificity
- Red List categories

The variable *plant part* depends upon the part of the plant that is utilised and harvested. The plant part used significantly affects harvesting sustainability (Wild and Mutebi 1996a) and hence plant risk. Harvesting fruits and leaves usually has a lower impact on population persistence than harvesting a whole plant such as a bulb (Cunningham 2001). Highest priority should therefore be given to species that are destructively harvested and where the whole plant is removed.

Phylogenetic distinctness indicates the degree of monotypy at the level of genus and family (Cofré and Marquet 1999). The more genetically unique a taxon is (e.g. a monotypic family), the greater its conservation value and priority (Millsap *et al.* 1990).

Endemic species are usually of conservation value because of their reduced geographic range. Greater conservation priority is usually awarded to 'narrow' endemic species (i.e. specialised species that tend to be confined to a small geographic area) and less priority is given to cosmopolitan species with a multi-national distribution.

Geographic distribution can also be represented by the number of *quarter-degree grid* squares (*QGrids*). Each QGrid for a species corresponds to a voucher specimen that was collected in a particular area (*e.g*. 2929CB) and deposited with the National Herbarium (PRE). One QGrid is equivalent to an area of 15'x15' (≈650km^2). Lists of QGrids for species are only available from the Pretoria Computerised Information System (PRECIS) database, maintained by the South African National Biodiversity Institute (SANBI) in Pretoria. These data included only localities of specimens in the national herbarium records and not from any of the almost 100 other herbaria in the country. However, PRECIS is the biggest and most readily available source of computerized locality data for a large number of species. One limit to the use of these data is that QGrids quantify the total extent of the *known* distribution and not the *current* distribution – i.e. local extinctions will be masked. The smaller the distribution, the greater the conservation priority.

Habitat requirements define the *habitat specificity* of species. The degree of habitat specialisation informs the conservation priority. Species found in more than one habitat are less of a conservation priority than species restricted to specialised or specific habitats. For example, *Myrothamnus flabellifolia* is restricted to shallow soils on rocks in the full sun in the bushveld, whereas *Talinum caffrum* is widespread in woodland and grassland (von Ahlefeldt 2003).

The *Red List (RL)* categories are based on the 2001 IUCN Red List categories and criterion for South African species. The information was obtained from Victor (2002) and J. Victor *(pers. comm.*). *Red List* was selected as a variable to ensure that species with a national conservation priority rating were not demoted in rank in the assessments for medicinal plants.

In general, species characterized by small geographic range, high habitat specialisation and low abundance are at higher risk of extinction, and consequently of over-utilisation, than species that are widely distributed and less habitat specific (Cofré and Marquet 1999). Similarly, conservation priorities are usually higher for higher risk species.

10.4 Methods of Analysis

10.4.1 Market Surveys and Quantitative Resource Inventories

Between February 1994 and January 2001, two semi-quantitative market surveys were conducted within the Witwatersrand. The first survey, based on a stratified random sample of 50 *muti* shops, appraised the characteristics of the trade in the formal sector (Williams *et al.* 2000). The second survey, a stratified random survey of 100 street traders in the informal Faraday market, was conducted on behalf of the provincial Directorate for Nature Conservation (Williams 2003).

In both surveys, an inventory of all common names of plants ('ethnospecies', or ESP) sold by each trader was compiled (Williams *et al.* 2005). The data were quantified based on the frequency of occurrence of each ethnospecies between the traders to obtain the total number of citations per species in each market. In addition, the number of times species were cited as 'scarce' and/or 'popular' was also recorded. During the Faraday survey, the volume of plant material sold per species per trader was also recorded and then summed to obtain the total number of 50 kg-size bags present in the market per species at the time of the survey (Williams 2003, 2004).

10.4.2 Biological Variable Scores

Data for the trade variables were quantitative and continuous and derived from the market surveys. Data for the biological variables are discrete and ordinal. Each biological variable was given a score ranging from 3 to 5. Larger variable values imply a higher risk and/or conservation priority.

Plant part was divided into five ranked classes depending on the degree of destructiveness of harvesting to the plant. The scores for each plant part code, in ascending order of harvesting risk, are:

1. fruits, seeds and/or flowers
2. leaves and/or stems
3. bark
4. roots
5. whole plants, bulbs, geophytes, tubers or rhizomes.

For *phylogenetic distinctness,* the following scores, listed in ascending order of species conservation priority, were given to species (modified after Farjon 1995):

1. Species in a large genus (≥16 species)
2. Distinct species in a medium genus (6-15 species)
3. Distinct species in a small genus (2-5 species)
4. Species in a distinct monotypic genus (i.e. genus with 1 species)
5. Species in a monotypic family (i.e. family with one genus).

The following scores were given to species based on their level of *endemism* (after Farjon 1995):

1. Species occurs scattered or rare over a wide range (≥2 countries)
2. Species is endemic to South Africa, but is scattered or rare over many km^2
3. Taxon is a narrow endemic with one or a few populations (usually an unknown number of populations, but with ≤1,000 individuals at reproductive age).

The scores given to species for *habitat specificity* were:

1. Found in three or more habitats
2. Limited to two habitat types
3. Restricted to one habitat type

The following scores, in ascending order of conservation priority, were given to species assessed according to current IUCN *Red List* categories:

0. Not evaluated or data deficient
1. Least concern
2. Near threatened
3. Vulnerable
4. Endangered
5. Critically endangered

Scores for *QGrids* were on a numerical scale and based on actual counts of the number of Quarter-degree grids listed with PRECIS for a species.

10.4.3 Standardizing the Scores

If variables are not measured on the same scale, then analyses should be performed using standardized scores for the variables (James and McCulloch 1990). The original variable values for the data ranged from, for example, 1-5 for plant part code and 1-61 for the number of Faraday traders selling a species. The variable values were therefore standardized between 0 (lowest score) and 1 (highest score) so that each variable was similarly scaled so as to not distort the results (especially the linkage distances in Ward's method, which is sensitive to the relative magnitude of the scores). To do this, the value for each species in a corresponding variable column was divided by the highest value in that column (with the exception of QGrids). For example, species A sold by 61 traders is presumed to be more vulnerable than species B sold by two traders. The corresponding standardised values for A and B would thus be 1.00 and 0.03 (=2/61) respectively.

Standardised scores for QGrids were calculated in a slightly different way. Unlike the other variables, where a high variable value implies higher priority, the inverse is true for QGrids because higher values imply a broader geographic distribution and hence lower conservation priority. For example, 11 QGrids for *Siphonochilus aethiopicus* signifies a greater chance of quasi-extinction risks to the species than does 282 QGrids for *Scabiosa columbaria*. Hence, standardised QGrid values were obtained by multiplying the inverse of the number of QGrids by the smallest number of QGrids in the data set. In the case of *S. aethiopicus* and *S. columbaria* above, the standardized valued would be 1.00 and 0.04 respectively. In the case of exotics, no values were given and unidentified or multiple species were not used in this stage of the risk evaluation and priority setting process.

Myers (1999) recommended that species be sorted according to their variable values so that similar species are placed nearby in "multivariate space". To do this, the standardized values for each variable were summed, without any weighting, and the species were sorted into descending order of the total score to derive a linear rank used as the species numbers in the PCA. Overall, by creating standardised, relative variable values for each species (by using the largest value), the resultant risks and conservation priorities proposed for species and/or clusters are also relative to each other and members of other clusters.

10.4.4 Regression

To assess whether the presence of species in a market can be positively correlated with volume, and hence the value of using species frequency as a variable, values for the number of traders selling a species were regressed with the total number of bags present per species. Additionally, the number of traders selling a species was correlated with the number of traders citing a species as scarce and/or popular to determine whether 1) popular species are necessarily more prevalent in the market, or 2) whether species regarded as scarce are any less prevalent. To test the validity of local knowledge of plant scarcities, the number of citations of plant scarcity was regressed with the number of QGrids.

10.4.5 Multivariate Analyses

A two-stage clustering method was used in the analyses that integrated the results of Ward's minimum variance method with the results from the K-means method to determine the number of clusters to be retained. *Ward's minimum variance* method on squared Euclidean distances was used to construct hierarchical agglomerative dendrograms of species groupings. Ward's method uses an analysis of variance to evaluate the distance between the clusters and attempts to minimize the sum of squares of any two clusters that can be formed at each step (StatSoft 2001). On each iteration, those groups whose fusion results in the lowest increase in the error of the sum of squares (or the variance) are combined (Kent and Coker 1992). The number of uniquely functioning clusters depends on what distance criteria or cluster dissimilarity is used (Myers 1999). Statistica 6.0 (StatSoft 2001) and Community Analysis Package 2.15 (CAP) (Pisces 2002) were used to conduct the analyses for the different data sets and combination of variables. An advantage of Statistica is that the linkage distances can be displayed on the axis as either actual or relative (%) distances. By comparison, the scale-bars of CAP dendrograms have tick-marks representing relative distance intervals of 25%, but are labelled with the actual distances. A plot of the linkage distances at successive clustering steps was used to decide the optimal cut-off for how many clusters to retain.

The *K-means clustering* or partitioning method involves minimizing the within-group variance, and maximizing the between-group variance, to produce a non-hierarchical partition of the objects into *K* groups or clusters (*K* being determined *a priori* by the user) (Legendre and Legendre 1998, StatSoft 2001). The groups should be internally homogeneous and different from each other (Lepš and Šmilauer 2003). K-means clusters can be further analyzed by comparing the mean values of each variable in the individual clusters. This is best summarized as a graph that plots the mean values of each variable in each cluster. Furthermore, by comparing the between-group variances, the variables most influencing the assigning of species to clusters can be determined. Comparing the Euclidean distances between the groups also facilitates the determination of how similar or different the clusters are. Statistica 6.0 was used to perform the K-means cluster analyses, and the process was repeated by progressively specifying the formation of 2 to 6 clusters (depending on the data set analyzed) and comparing the species within each cluster and with species in the dendrogram clusters. Each cluster was given the following nomenclature: K_2 K_3, K_4 or K_5 indicating the number of user-specified clusters of 2, 3, 4 or 5 per analysis, followed by a number indicating a particular cluster. For example, $K_4$2 indicates cluster number 2 for a pre-specified number of 4 clusters. The plot of means for each cluster is an optional output of the Statistica program.

To select the optimal linkage distance that determines the number of unique clusters representing groups of species with similar risk profiles, species in the hierarchical dendrogram clusters at various linkage distances were compared with the species partitioned by K-means clusters. Using an initial partition of 2 K-clusters (K_2), species in the two clusters at ±100% relative distance[1] (i.e. maximum distance) on the dendrogram were compared with the species in the two K-means clusters. If the species were found to be the same or similar (i.e. ≥75% species correspondence), then 3 K-clusters (K_3) were created and the species were compared with species in the three dendrogram clusters at the corresponding shorter relative distance. If these species were the same or similar, then 4 or more K-clusters (K_4) were created and compared at the corresponding linkage distances until the species in the corresponding clusters were partitioned in groups that were less alike. The linkage distance selected, and hence the final number of clusters accepted for interpretation, was the point at which species in the K-means and hierarchical agglomerative clusters were most congruent – this was usually >40% relative distance.

The data sets of variables in a standardized species attribute matrix (species x variables) were ordinated using DCA and PCA in CANOCO for Windows Version 4.51. First, DCA was carried out to determine whether the data were linear or unimodal. If the length of the longest gradient from the output of the DCA subroutine was <3.0 (which they were), then the data were known to fit the linear response model and PCA analyses were conducted. Because some of the data were presence-absence, Principle Coordinates Analysis (PCoA) was a more appropriate analysis. In CANOCO, PCoA can be obtained as a variant of PCA by selecting options that centre by samples and species, and that do not post-transform species scores. Wherever PCA is mentioned in the results and the discussion, it is to be read that the results are the same as for a PCoA. The data in the matrix was ordered so that the species were in rows and the criteria/variables for evaluating risk were in columns. Hence, in the dialogue boxes of the CANOCO program, ‘SPECIES’ was equivalent to the variables selected to analyze the risk, and ‘SAMPLES’ was replaced by the species being evaluated.

The emphasis of the ordination was on assessing which variables are positively correlated with risk and which variables most influenced the cluster analysis results. The output of the PCA produced a table of eigenvalues (between 0 and 1) for the four axes. The eigenvalues measured the importance of an axis, and expressed this as a fraction of the total variance in the variable data. A biplot was also produced showing a scatter plot of the eigenvector scores (ter Braak, 1988). The variables are represented by arrows that roughly point in the direction of maximum variation in the value of the corresponding variable. The species are represented by points scattered around the arrows. The length and direction of the arrows indicate correlations between the species and the variables.

[1] Relative distance = (linkage distance ÷ maximum distance) * 100

10.4.6 Other Statistical Analyses

The conservation priority categories for 87 short-listed species (see Step 2, following section) were correlated with conservation categories created for species with similar totalled numerical importance value (NIV) scores using Spearman's rank correlation coefficient (r_s).

10.4.7 Summary of the Risk Evaluation Method

The risk evaluation process was divided into several stages, depicted in Boxes 1 and 2.

- **Step 1**: used cluster analysis (Ward's hierarchical and K-means non-hierarchical) on three data sets to identify a short-list of species that are at risk or vulnerable due to harvesting for the medicinal plant trade. The three data sets are:
  - I. 333 ethnospecies, derived from plants sold in 50 *muti* shops (Page 24);
  - II. 315 ethnospecies, derived from plants sold by 100 street traders in Faraday (Page 29);
  - III. 392 ethnospecies, derived by combining data from I and II for 150 traders (Page 35).

  Short-lists of species categorised as higher risk were identified for data sets I and II. From the combined data set III, a short-list of 119 species (reduced to 87 species after exotics and multiple species cluster were removed), were identified as high to medium high risk and then analysed in more detail in Step 2 using more variables.

- **Step 2**: used cluster analysis on the 87 short-listed species (data set IV, derived from data set III) with additional biological variables to assess vulnerability and conservation priority (Page 43).

- **Step 3**: used PCA to explore the direction and strength of the correlations between variables and species with respect to their relative risks, and to detect the presence of redundant variables (Page 50).

- **Step 4**: assessed the efficacy of the methods (Steps 1-3) for predicting the risks in smaller data sets by using a small group of 22 species (data set V). Additionally, variables that can only be used on smaller data sets were evaluated (e.g. *R/kg*, *kg/sal*e and *sales/annum*) (Page 56).

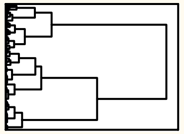
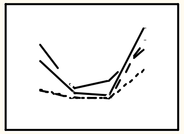
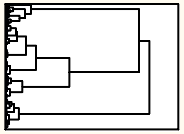
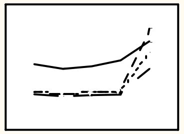
- **Step 5**: compare the results of the cluster analysis for 87 species with their ranked order based on summing the variable values to create NIVs (Page 64)

Steps 1, 2 and 4 essentially examine the relative risk of species in large, medium and small data sets respectively using different sets of appropriate variables.

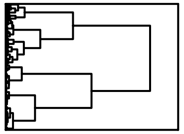
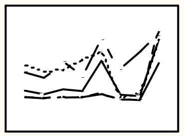
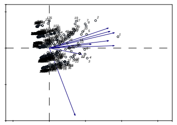
With respect to interpreting the dendrograms and clusters generated with Ward's hierarchical method, the multivariate clusters simultaneously account for multiple factors affecting harvesting vulnerability. Species in the same clusters were assumed to have similar vulnerabilities/risks, and the distances between the clusters indicated how similar or dissimilar those risks and conservation priorities were.

In Step 1, the variables used in the analyses indicate the degree of destructiveness of plant harvesting as well as the prevalence and perceived scarcities and popularities of the plant – all of which are indicators of plant vulnerability. Risk hierarchies were thus created because plants with similarly high variable values were clustered together, implying that they were more at risk to over-harvesting than plants with low values that were further away at greater linkage distances. Therefore, species that were far apart on the vertical axis were assumed to warrant different management actions. Variables used in Step 2, however, primarily assigned species to clusters of similar vulnerability and conservation priority.

While there is an element of subjectivity in proposing a risk characterisation for a cluster (e.g. whether a cluster is high, medium-high, medium or low risk), what is objective, however, is the clusters (and the species therein) that the analyses produce. In previous studies, a challenge with interpreting NIVs and conservation priority indices has been deciding where the numerical cut-off point is between species in order to assign them to different priority and management groups. Cluster analysis removes the numerical ranks, and clusters species into groups with similar characteristics and attributes. These clusters have similar suites of risk factors.

Data set	Analyses		Variables tested
I Shops 333 esp N = 50		Ward's (pg 24)	4 variables • No. shops selling • Scarcity • Popularity • Plant part
		K-means (pg 26)	
II Faraday 315 esp N = 101		Ward's (pg 29)	5 variables • No. traders selling • Scarcity • Popularity • Plant part • No. bags
		K-means (pg 31)	

Risk analysis

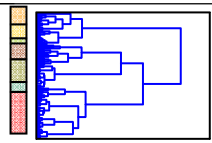
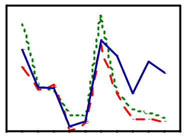
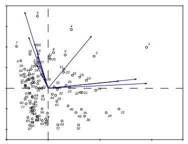
Data set	Analyses		Variables tested
III All 392 esp N = 151		Ward's (pg 35)	8 variables • No. shops selling • No. Faraday traders selling • Scarcity (S and F)* • Popularity (S and F) • Plant part • No. bags
		K-means (pg 37)	
		PCA (pg 50)	

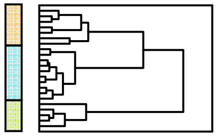
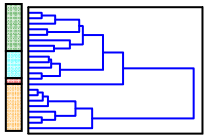
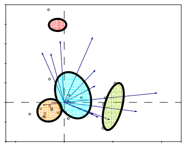
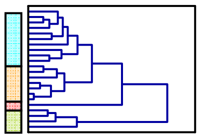
Box 1:

Steps taken in the primary risk analysis. First, data sets I and II were analysed using clustering techniques. Thereafter, I and II were combined to form data set III, which was analysed using clustering and also PCA. A short-list of 119 species (further reduced to 87 species) (data set IV) was derived from III and analysed with more variables to determine conservation priorities for the species.

* S and F represent 'Shops' and 'Faraday' respectively; HS = Habitat specificity, End = Endemism, PD = Phylogenetic distinctness, RL = Red List

Conservation priority assessments

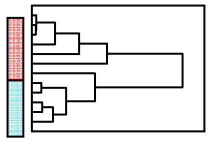
Data set	Analyses		Variables tested
IV Short-list of 119 species. Reduced to 87 spp after exotics and multiple-species clusters were removed		Ward's (pg 43)	10 variables • No. traders • Scarcity • Popularity • Plant part • No. bags • HS • End • PD • RL • QGrids
		K-means (pg 46)	
		PCA (pg 53)	

Data set	Analyses		Variables tested
a) V 22 test species Analysed and compared with same 8 variables used to test data set III ('All, 392 esp)		Ward's (pg 56)	8 variables • No. shops • No. Faraday traders • Scarcity (S and F) • Popularity (S and F) • Plant part • No. bags
	Table 23	K-means (pg 56)	
b) V 22 test species Analysed and compared with the same 11 variables used to test data set IV (87 species short-list)		Ward's (pg 58)	10 variables • No. traders • Scarcity • Popularity • Plant part • No. bags • HS • Endemism • PD • RL • QGrids
	Graph not shown	K-means (pg 58)	
c) V 22 test species Analysed with the same 11 variables used to test data set IV plus 5 additional trade variables. Trade data derived from *muti* shops only.		PCA (pg 60)	12 variables • No. shop traders • Scarcity • Popularity • Plant part • No. bags/a • No. sales/a • R/kg • kg/sale • RL • QGrids • HS • PD
		Ward's (pg 62)	
	Graph not shown	K-means	

Box 2:

Risk and conservation priority analysis of 22 test species (data set V). The species were analysed in a similar way to data sets III and IV (Box 1) to assess the efficacy of the methods in predicting risks/conservation priorities for smaller data sets with fewer species. Additional variables also investigated.

* S and F represent 'Shops' and 'Faraday' respectively; HS = Habitat specificity, End = Endemism, PD = Phylogenetic distinctness, RL = Red List

Data set	Analyses		Variables tested
d) 12 species Bulbs, roots, tubers, rhizomes		Ward's (pg 63)	4 variables • R/kg (Faraday and shops) • Kg/sale (Faraday and shops)

10.5 Results and Discussion

10.5.1 The Value of a Citation

The simplest way to quantify plants in a marketplace is to count the total number of traders selling each species. Similarly, counting citations of scarcity and/or popularity indicates which species may be in demand or are becoming increasingly difficult to obtain. It is thus constructive to question whether plant frequency in a market is a satisfactory indicator of risk and vulnerability to harvesting. In other words, does the increased incidence of a species mean that there is likely to be a corresponding increase in the volume of that species present in the market and therefore an analogous increase in plant risk/vulnerability?

The relationship between the number of traders selling a species and the corresponding volume present in the market is highly positive and significant (r^2=0.920, p<0.0001) (Figure 1a). Therefore, when questioning the value of a citation, it can generally be expected that the more traders that sell a species, the more of it there is likely to be in the market.

Of the plants sold in the Faraday market in 2001, 84% were sold by ≤13 traders with a total individual volume of <6 bags per species (Figure 1, area enclosed in the box; Figure 1b). The relationship between the number of citations and volume for the species sold by ≤13 traders is still positively correlated (r^2=0.648; p<0.0001) (Figure 1b), but the relationship is not as strong (i.e. lower r^2) as the regression of the entire sample. The graph also shows the greater variability in the total quantities available for the less frequently occurring species (Figure 1b).

Another question is whether plants cited by traders as 'popular' occur in larger quantities in the market than so-called 'less popular' species? Seventy-five percent of the plants sold in Faraday were cited at least once as being popular. Citations of popularity per species and the quantity present in the market are positively correlated (r^2=0.815, p<0.0001) (Figure 2a). For species like *Hypoxis* spp. and *Gunnera perpensa* (no's 4p & 9p in Figure 2a) however, the overall quantities present in the market are higher than expected relative to the number of citations of popularity. On the other hand, *Elaeodendron transvaalense*, *Acacia xanthophloea* and *Schotia brachypetala* (no's 5p, 7p & 8p, Figure 2a) have lower than expected quantities relative to their popularity. This could be due to high sales resulting in lower relative availability.

In some cases, an inverse relationship between the number of citations of species scarcity and the quantity found in the market might be expected. However, there is a positive, albeit weaker, correlation between citations of species scarcity and the quantities present (r^2=0.452; p<0.0001) (Figure 2a). Fifty-three percent of all plants sold in Faraday were cited at least once as being scarce, but the maximum number of traders citing any one species was 10 (i.e. 10% of the total number of traders).

The results serve to confirm that plants mentioned by traders as being scarce and/or popular are likely to be present in relatively larger volumes in the markets. This is likely due to the demand generated for certain species that has resulted in some of them becoming scarce, but the same demand causes traders to actively acquire stock even though the quantities they obtain may be smaller and the prices may be higher. Species numbers 1s,p – 7s,p (Figure 2a), for example, have approximately the same rank in terms of both popularity and scarcity and are available in relatively large quantities. Species like *Warburgia salutaris* and *Siphonochilus aethiopicus* (no's 10s & 11s, Figure 2a), however, have a high number of citations of scarcity and are present in quantities indicative of their known scarcity in the wild. The demand is therefore greater than the amount that can be supplied.

The value of a scarcity and popularity citation for a species can also be judged from its relationship with the total number of traders keeping the species (Figure 2b). In general, the more traders that sell a plant, the more likely they are to cite it as popular (r^2=0.853) and, to a lesser extent, scarce (r^2=0.539). Additionally, 10% of the species were cited as both scarce *and* popular, and it is the popularity with consumers that has contributed to decreased availability.

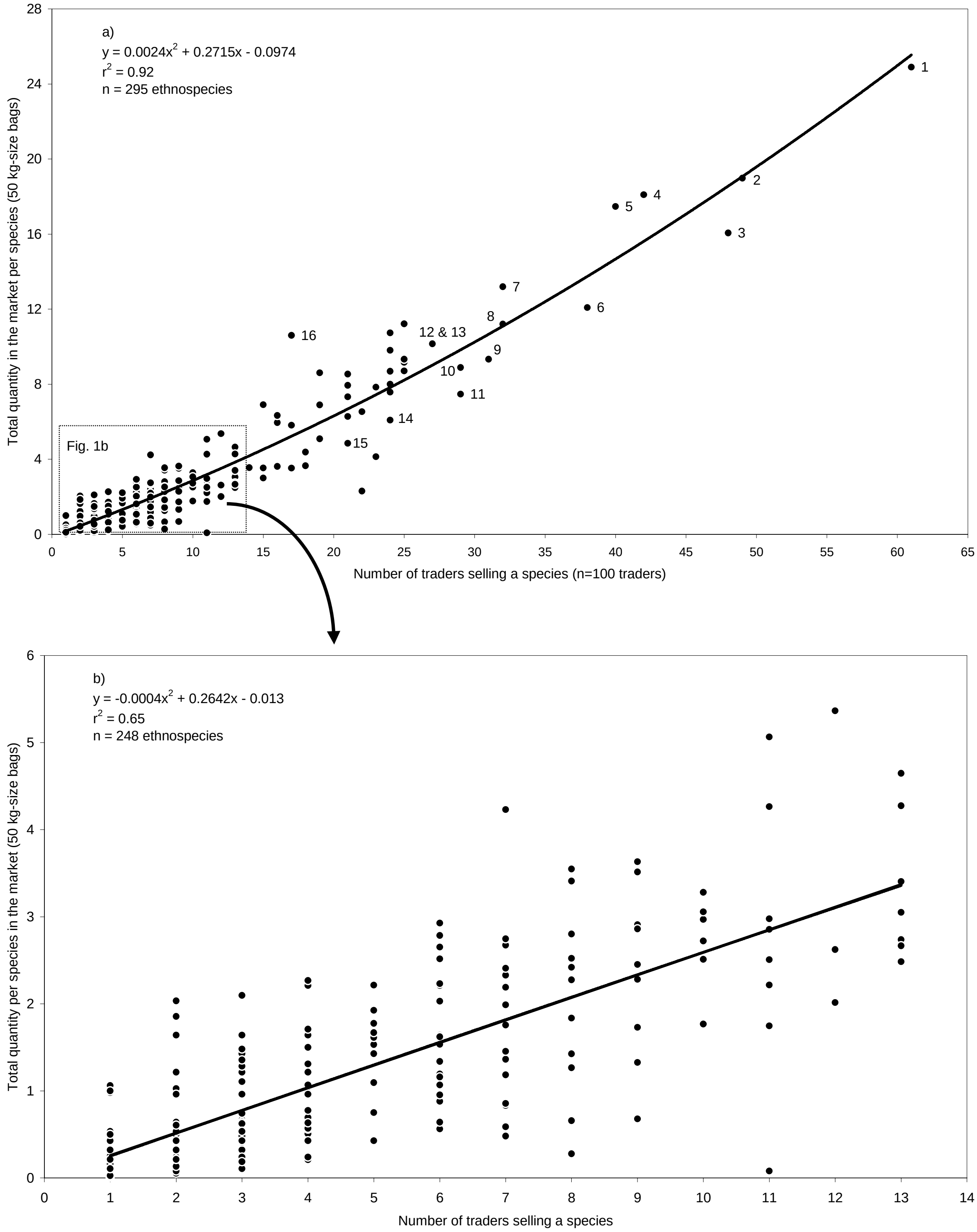


Figure 1: The relationship between the total number of citations per species (i.e. number of recorded incidences of the species in the market) and the total volume present in the market per species in Faraday in 2001. Figure (b) is a subset of (a), and represents ≈84% of the data set. Species in (b) were sold by ≤13 traders with a quantity in the market totalling <6 bags. Notable species: 1=*Drimia* spp.; 2=*Hydnora africana*; 3=*Elaeodendron transvaalense*; 4=*Albizia adianthifolia*; 5=*Hypoxis* spp.; 6=*Acacia xanthophloea*; 7=*Schotia brachypetala*; 8=*Sclerocarya birrea*; 9=*Urginea* spp.; 10=*Adenia gummifera*; 11=*Dioscorea sylvatica*; 12=*Rapanea melanophloeos*; 13=*Merwilla plumbea*; 14=*Eucomis autumnalis*; 15=*Warburgia salutaris*; 16=*Helichrysum* spp.

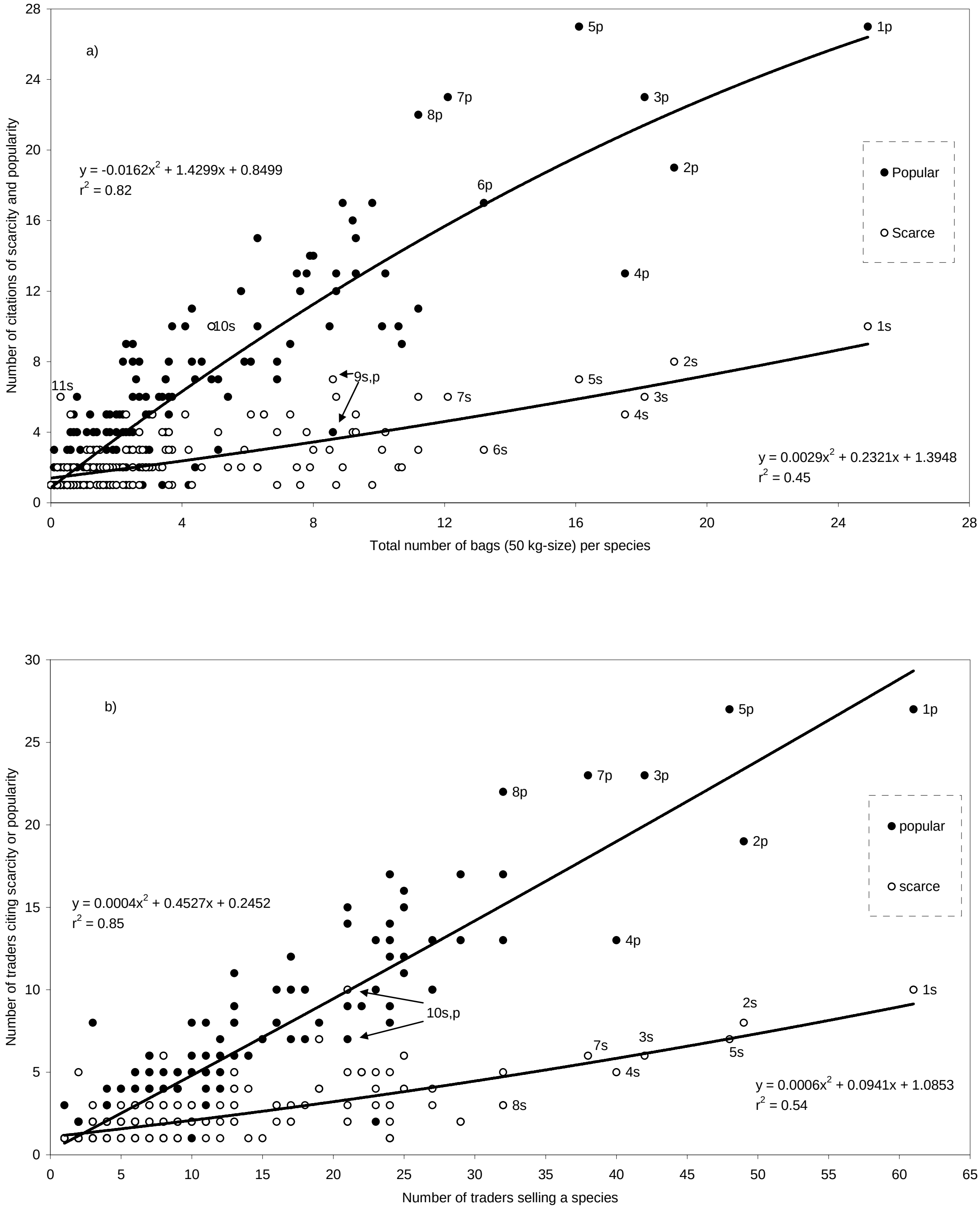


Figure 2: Regression between (a) the number of citations of popularity and scarcity per species and the quantity present in the Faraday market, and (b) the total number of traders selling a species and the total number of citations of scarcity and/or popularity (Faraday 2001; n=100 traders). 1=*Drimia* spp.; 2=*Hydnora africana*; 3=*Albizia adianthifolia*; 4=*Hypoxis* spp.; 5=*Elaeodendron transvaalense*; 6=*Sclerocarya birrea*; 7=*Acacia xanthophloea*; 8=*Schotia brachypetala*; 9=*Gunnera perpensa*; 10=*Warburgia salutaris*; 11=*Siphonochilus aethiopicus*. Suffixes ‘p’ and ‘s’ denote ‘popular’ and ‘scarce’ respectively.

In summary, species that have a high incidence in markets are more likely to be harvested and be present in larger quantities, and are consequently at greater risk of over-harvesting. Therefore, if data on the quantity of plants sold in a market is not collected during an ethnobotanical survey, then the frequency of species occurrences is a reasonable indicator of the relative quantities that would have been present and hence the relative risks.

10.5.2 The Validity of Local Knowledge/Perception as a Model variable

In this section, the validity of the traders' perceptions of plant scarcity was examined against the known distributions of certain species. Local perceptions can provide insight into the actual scarcities of valuable species. Some scarcities may be caused by limited geographical distributions and habitat reduction, while others may be attributed to over-exploitation. Plant popularity is not similarly examined because 'popularity' is usually a statement of opinion subject to consumer behaviour. That said, if there is no demand for species with a limited distribution then traders are unlikely to sell and cite them as scarce.

The cost of obtaining geographic distribution information restricted the number of species that could be examined. The relationship between species scarcity and known distribution (a function of the number of QGrids) was only tested with the group of 87 short-listed ethnospecies that emerged as high to medium risk from the cluster analyses (Section 10.5.7, page 43). It is to be remembered that QGrids show the total extent of the known historical distribution and not the current distribution.

There was a slight decreasing trend in the number of citations of scarcity with increased number of QGrids (r^2=0.017 – 0.029, Figure 3 a,b,c for Shop, Faraday and combined data sets respectively). Most noticeable in terms of high citations of scarcity for species occurring in $\leq$50 QGrids were *S. aethiopicus*, *W. salutaris* and *Ocotea bullata*. Scarcities for these species are attributable to a combination of limited distribution (in South Africa) and over-harvesting. *Drimia* spp. and *Eucomis autumnalis*, however, have a high number of scarcity citations relative to their distribution, indicating that they are becoming increasingly vulnerable due to the threats of over-harvesting. Scott-Shaw (1999) describes *E. autumnalis* as being "critically over-exploited over most of its distribution range".

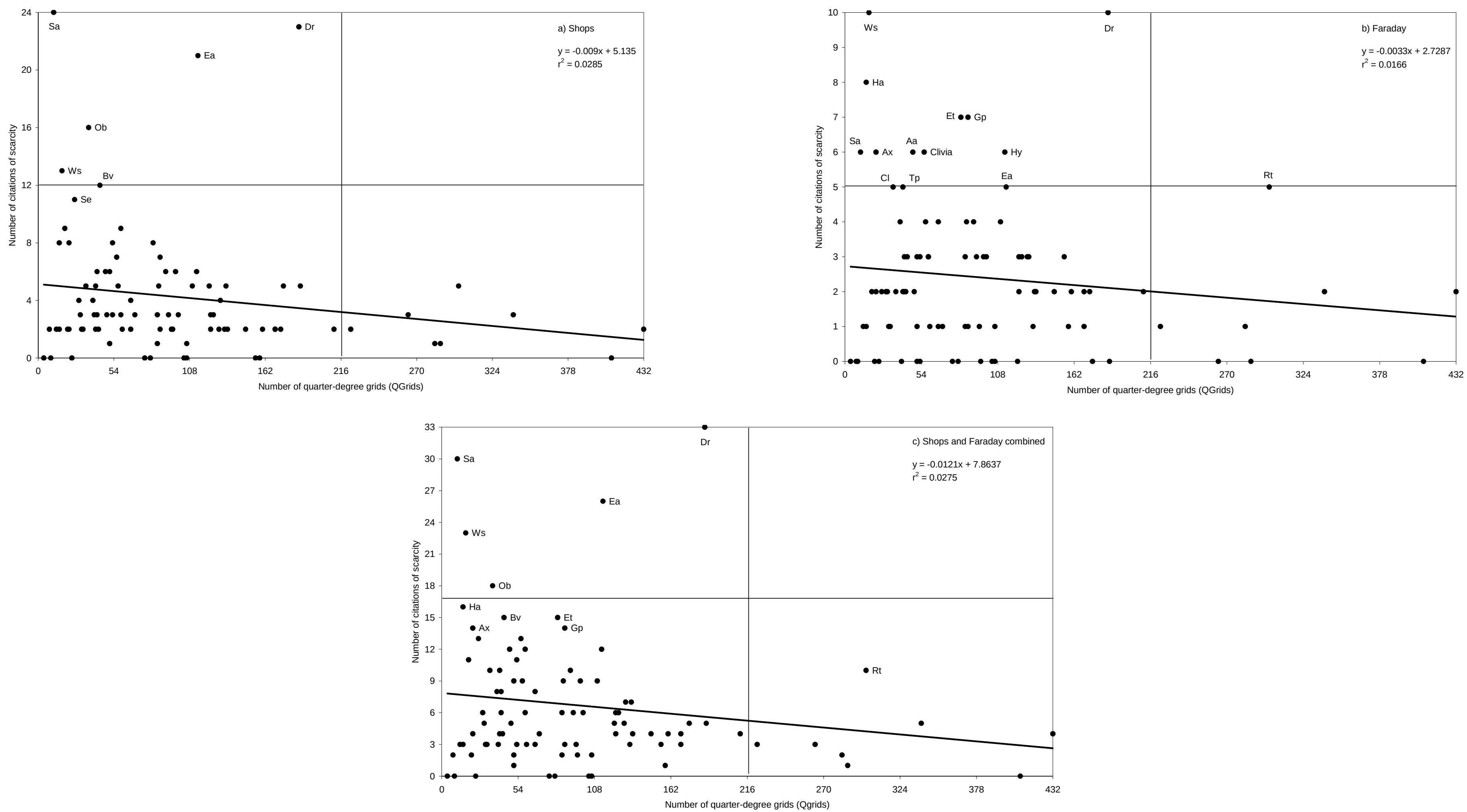


Figure 3: Relationship between the geographical distribution of plants (in number of quarter-degree grids) and the number of citations for plant scarcity by shop and street traders (n=87). Aa=*Albizia adianthifolia*; Asp=*Asparagus* spp.; Ax= *Acacia xanthophloea*; Bv=*Bowiea volubilis*; Cl=*Clivia* spp.; Cr=*Croton* spp.; Dr=*Drimia* spp; Ea=*Eucomis autumnalis*; Er=*Eriosema* spp.; Et=*Elaeodendron transvaalense*; Gn=*Gnidia* spp.; Gp=*Gunnera perpensa*; Ha=*Hydnora africana*; Hel=*Helichrysum* spp.; Hy=*Hypoxis* spp; Ob=*Ocotea bullata*; Pt=*Pterocelastrus* spp.; Sa=*Siphonochilus aethiopicus*; Sc=*Scabiosa columbaria*; Se=*Stangeria eriopus*; Ws=*Warburgia salutaris*.

10.5.3 Bivariate risk characterization model

While QGrid data may mask local extinctions, it enables comparisons on the relative extent of the known occurrence of species. The bivariate, schematic risk characterization model presented in Figure 4 can be used to illustrate the relative risks of species using QGrids on the x-axis and a range of other quantitative variables on the y-axis, such as the number of traders selling a species, the number of sales per annum and volume.

In Figure 3a-c, the mid-points of the axes are approximately half the maximum variable value for a group of species. Species lying close to the x-axis have an inferred low risk because their y-values are lower, compared to species that are close to the y-axis because their distributions are restricted. Should populations decline in number and area of occupancy, then QGrid values would shift to the left and the risks to species would increase (Figure 4). If, for some reason, the demand for a species increased and prevalence in the market simultaneously increased, then the risk profile of a species could change from low to medium or high risk. Species that would be found in block 2b are worth noting (Figure 4). For some, population size and prevalence in the market has already declined because of increased harvesting pressures. If the number of QGrids has also decreased significantly, then species would be reassigned to a higher risk category (Figure 4).

While bivariate correlations are informative, the multivariate consideration of several variables is most useful. The risk evaluation method presented in the remainder of this chapter explores a multivariate approach to assessing risks and conservation priorities for species.

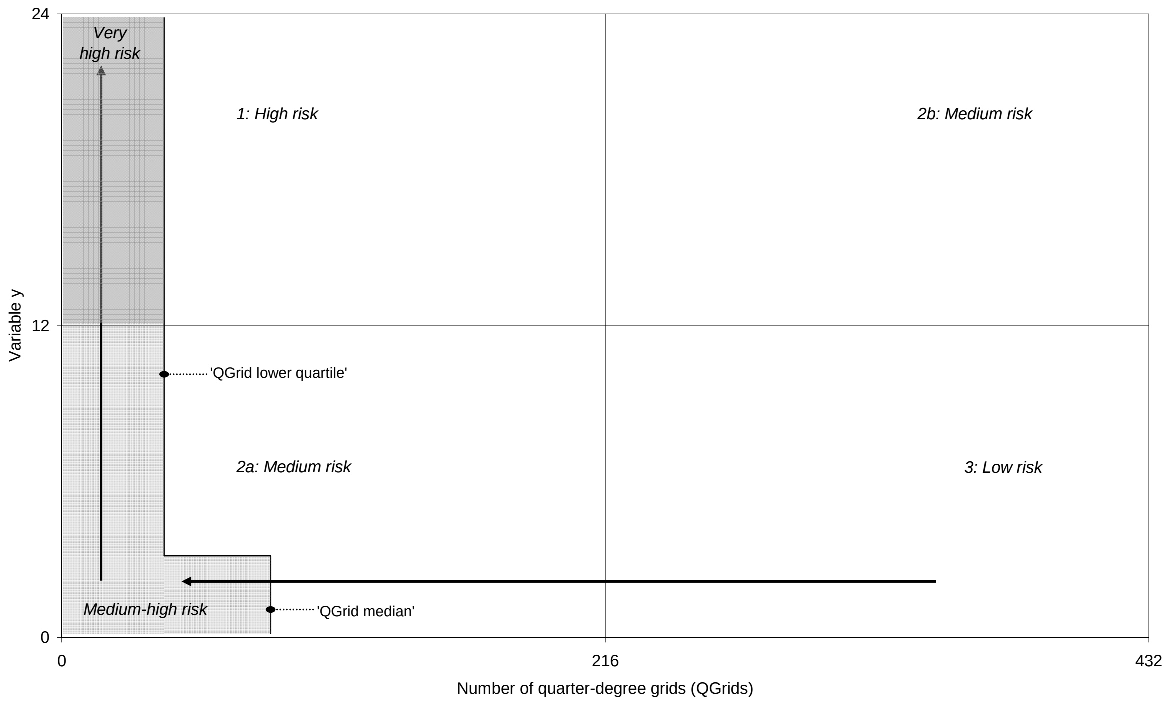


Figure 4: Proposed risk characterizations for bivariate analyses where the variable on the y-axis is, for example, the prevalence of species or the number of citations of scarcity/popularity or any other variable that indicates increasing risk with increasing value of the variable. The long arrows indicate the direction of increasing risk along the x or y axis if the variable value for a species had to change. ‘QGrid lower quartile’ and ‘QGrid median’ indicate where the lower quartile and median values for the number of QGrids are.

10.5.4 Plant risk analysis: 333 ethnospecies sold in 50 *muti* shops

a) Hierarchical clustering: Ward's method.

In this sequence for 333 ethnospecies inventoried in 50 *muti* shops, clustering resulted in a risk hierarchy of species clusters shown in the dendrogram of Figure 5. The variable attributes are summarised in Table 3, and the expanded dendrogram showing the species in each cluster is in Appendix 1. Altogether, three main cluster groups were identified (A, B & C) at a linkage distance of ≈21 (57% level of relative distance). These clusters were further aggregated into two larger groups (1 and 2). Given the characteristics of the species within the clusters, and the distances between them, A and B/C can be broadly categorised as higher and lower risk respectively. Cluster A contains a mixture of highly traded and cited species of different plant part types. Clusters B and C contain species that are bark/leaves/stems/fruits and whole plants/roots respectively that are not as highly traded and/or cited.

The attributes of the species in the clusters were as follows:

- **Cluster A** (29.4% of ethnospecies): species have a high prevalence in the shops and tend to have a high number of citations of scarcity and/or popularity. There is a range of plant part types. Species in this cluster include *Drimia* spp, *W. salutaris*, *O. bullata* and *S. aethiopicus* (sub-cluster A1). Suggested risk categorisation: high.
- **Cluster B** (40.8% ethnospecies): species in this cluster are species used for bark, leaves/stems and fruits. They have a medium to low prevalence in the shops and a low number of citations of scarcity and popularity. Sub-cluster B2 contains only species harvested for leaves/stems and fruit, the harvesting impact of which is generally low. Hence, species in B2 sub-cluster would have an over-all lower risk to harvesting than most of the rest of B1. Species in B include *Dombeya rotundifolia*, *Schotia brachypetala* and *Rhus chirindensis* Suggested risk categorisation: low.
- **Cluster C** (29.7% ethnospecies): species in this cluster are all whole plants or roots - the harvesting impact on which is generally higher. However, these species have a medium to low prevalence in the shops and are generally not regarded as scarce and/or popular. Species in C include *Ornithogalum* spp., *Agapanthus* spp. and *Typha capensis.* Suggested risk categorisation: low.

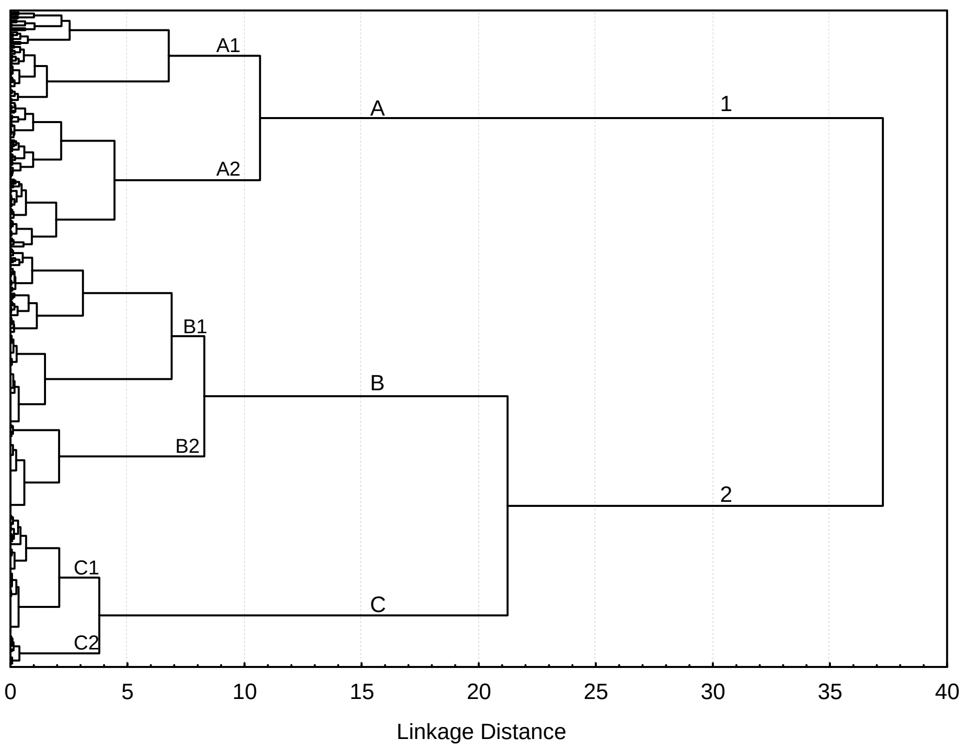


Figure 5: Dendrogram based on Euclidean distances and Ward's agglomerative clustering method for 333 ethnospecies inventoried in 50 *muti* shops. The species names are present in the extended dendrogram in Appendix 1. Attributes of the species in the clusters are listed in Table 3.

Table 3: Cluster attribute summary for 333 ethnospecies (esp) sold in 50 *muti* shops corresponding with the dendrogram in Figure 5 and Appendix 1.

Group	Cluster	Sub-cluster	No esp. (%)	Number of esp. per plant part category: Whole plants	Roots	Bark	Leaves/ stems	Fruit	Number of traders selling	Scarce citations	Popular citations
1	A	A1	47 (14.1%)	8	4	24	11	-	$6 \le x \le 45$	0 – 18	0 – 7
		A2	51 (15.3%)	19	32	-	-	-	$13 \le x \le 33$	0 – 5	0 – 2
		Subtotal	*98 (29.4%)*	*27*	*36*	*24*	*11*		*$6 \le x \le 45$*	*0 – 18*	*0 – 7*
98 (29.4%)											
2	B	B1	90(27.0%)	-	-	69	15	6	$1 \le x \le 22$	0 – 5	0 – 4
		B2	46 (13.8%)	-	-	-	39	7	$1 \le x \le 4$	0 – 1	0
		Subtotal	*136 (40.8%)*	-	-	*69*	*54*	*13*	*$1 \le x \le 22$*	*0 – 5*	*0 – 4*
	C	C1	37 (11.1%)	37	-	-	-	-	$1 \le x \le 21$	0 – 5	0 – 1
		C2	62 (18.6%)	-	62	-	-	-	$1 \le x \le 11$	0 – 3	0 – 1
		Subtotal	*99 (29.7%)*	*37*	*62*	-	-	-	*$1 \le x \le 21$*	*0 – 5*	*0 – 1*
235 (70.6%)											
Total:			333 esp.	64	98	93	65	13	1 – 45	0 - 18	0 - 7

b) Non-hierarchical clustering: K-means

The goal of K-means is to find the optimum 'partition' for dividing a number of species into K-clusters (i.e. K_2, K_3, K_4 etc. clusters), minimize the within-cluster variance and maximize the between-cluster variance. Using Ward's method, three clusters in the *muti* shop data set were identified. The kinds of solutions K-means clustering suggests for three or more clusters are examined here.

i) 3 clusters (K_3)

The results for 3 (K_3) clusters were very similar to the clusters found with Ward's method in terms of the species within a cluster and the number of species per cluster:

- **$K_3$1** (93 esp) corresponds with Ward's cluster **A** (98 esp)
- **$K_3$2** (131 esp) corresponds with **B** (136 esp)
- **$K_3$3** (109 esp) corresponds with **C** (99 esp)

Furthermore, all species in $K_3$1 were in cluster A, and the same plant part types were found in $K_3$2 and B as well as $K_3$3 and C respectively.

A useful result to examine is the Euclidean distances between the clusters (Appendix 8). The distances don't specifically correspond to risk, but show how different the clusters are and how the cluster differences are related to the differences in the mean values per cluster. Hence, the shorter the distance, the more similar the variable means are for one cluster compared with another. For 3 clusters, clusters $K_3$2 and $K_3$3 were closer together (Euclidean distance = 0.192) than $K_3$1 was from clusters $K_3$2 and $K_3$3 (Euclidean distance = 0.298 and 0.255 respectively) (Appendix 8 Table a). This result would further support categorising clusters B and C as lower risk versus cluster A as higher risk, given the shorter Euclidean distance between B and C.

Another way of identifying the nature of each cluster is to examine the standardized and transformed means for each cluster variable (Figure 6a and Table 4 respectively). Species grouped in $K_3$1 were sold by a mean of 28 shops, compared to means of 5 and 7 for $K_3$2 and $K_3$3 respectively. In addition, the mean plant part code for $K_3$3 was between 4 and 5, indicating the group consists of species sold as roots and whole plants. Conversely, species in $K_3$2 were predominantly bark and leaves/stems. Judging from the magnitude (and significance levels) of the *F* values, the variables *no. of shops* and *plant part* were the major criteria for assigning the species to clusters (Table 4).

Table 4: Mean values per cluster (transformed from standardized values in Figure 6a to mean actual values) corresponding with 3 clusters. Shaded values were the highest in the row, thereby implying higher associated risks. [*See Page 14 for plant part code] [df = 2, 330; $P < 0.0001$ for all variables]

	$K_3$1	$K_3$2	$K_3$3	F
Mean no. shops selling species	28	5	7	552.7
Mean no. scarce	2	0	0	21.9
Mean no. popular	1	0	0	43.0
Mean plant part code*	3.8	2.4	4.3	222.0

ii) 4 clusters (K_4)

The results for 4 (K_4) clusters did not correspond as well with clusters formed by Ward's method as K_3 did. There were, however, some similarities:

- **$K_4$4** (47 esp) was 91% similar in terms of species composition as Ward's sub-cluster **A1** (47 esp). All plant part types were represented.
- **$K_4$1** (76 esp) corresponds with sub-cluster **A2** (51 esp), with the balance of the species from **$K_4$1** mainly in **C1**. All species were whole plants and roots
- **$K_4$2** (143 esp, all whole plants, roots and bark) and **$K_4$3** (67 esp., all leaves/stems and fruit) were similar in composition to clusters **B** and **C** respectively. The difference between the clusters produced by the two different methods was the cluster in which bark species were present.

Clusters $K_4$1 and $K_4$4 were relatively close together (Euclidean distance = 0.197) compared to the distance of $K_4$1 from clusters $K_4$2 and $K_4$3 (Appendix 8 Table b). The most dissimilar clusters were $K_4$4 and $K_4$3 (Euclidean distance = 0.338).

Results for the standardized and transformed means for each cluster variable show that species in $K_4$4 and $K_4$1 were sold by a mean of 31 and 21 shops respectively, compared to means of 5 for $K_4$2 and $K_4$3 (Table 5 and Figure 6b). Cluster $K_4$4 also had higher mean values for scarcity and popularity, and could be considered higher risk than cluster $K_4$1. While $K_4$2 and $K_4$3 have the same mean trade values, the species in $K_4$3 generally had lower harvesting risks due to the less destructive nature of plant part harvesting. Judging from the magnitude (and significance levels) of the *F* values, variables *no. of shops* and *plant part* were, once again, the major criteria for assigning the species to clusters (Table 5).

Table 5: Mean values per cluster (transformed from standardized values in Figure 6b to mean actual values) corresponding with the 4 clusters. Shaded values were the highest in the row. [*See Page 14 for plant part code]. [df = 3, 329; P < 0.0001 for all variables]

	$K_4$1	$K_4$2	$K_4$3	$K_4$4	F
Mean no. shops selling species	21	5	5	31	447.0
Mean no. scarce	1	0	0	2	15.9
Mean no. popular	0	0	0	2	49.8
Mean plant part code*	4.5	3.7	1.8	3.1	203.7

The overall results of Ward's and K-means clustering suggest that 3 clusters were the optimum number, and that species in cluster A/$K_3$1 are higher risk than species in B/$K_3$2 and C/$K_3$3. There was further evidence from the K_4 results that species in A1/$K_4$4 are more vulnerable than the other species in A2/$K_4$1.

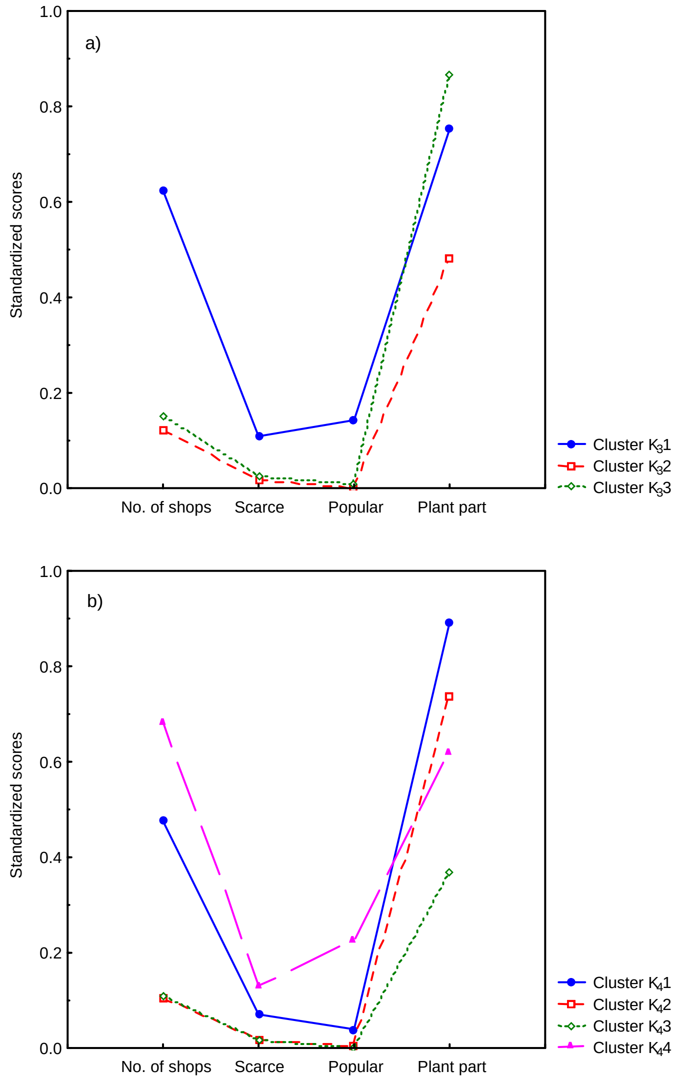


Figure 6: Plot of standardised cluster means for each k-means cluster for (a) 3 clusters and (b) 4 clusters for species sold in 50 *muti* shops. See Tables 5 and 7 for the transformed standardised mean scores into actual mean values for (a) and (b) respectively.

c) Priority species sold by muti shops

By comparing and combining the species in clusters A1 and $K_4$4, a short-list of 51 species identified as higher risk/vulnerable due to their trade in *muti* shops was compiled (Table 6).

Table 6: Fifty-one species identified as higher risk by Ward's and K-means cluster analyses. The species correspond to those in clusters A1 and $K_4$4. The species are listed in descending order of the number of shops selling the species and the number of scarcity citations. Plant part code: 1=fruit; 2=leaves/stems; 3=bark; 4=roots; 5=whole plants (incl. bulbs, tubers etc.). ESP numbers listed in Appendix 4. * exotic.

Probable species	**No. of shops selling species (N_{max}=50)**	**No. cited as Scarce**	**No. cited as Popular**	**Plant part**
Drimia spp.	45	13	5	5
Rapanea melanophloeos	42	2	3	3
Eucomis autumnalis ssp.	40	16	7	5
ESP 17 (uVelabahleka)	37	1	3	4
ESP 14 (uBangalala)	36	7	2	4
Merwilla plumbea	36	6	7	5
Adenia gummifera var.	36	1	2	2
Helichrysum spp.	36	0	6	2
Acacia xanthophloea	35	2	0	3
Ocotea bullata	34	14	2	3
Alepidea amatymbica var.	34	4	2	5
Croton spp.	34	3	1	3
Warburgia salutaris	34	3	4	3
ESP 11 (umVuthuza)	34	2	1	2
ESP 22 (uSehlulamanye)	34	0	1	3
Elaeodendron transvaalense	33	1	4	3
Balanites maughamii	33	0	0	3
Bowiea volubilis	32	9	0	5
Calodendrum capense	32	3	0	3
Dianthus mooiensis sspp.	32	1	6	5
Ekebergia capensis	32	1	0	3
ESP 20 (uMabophe)	32	1	1	4
Gunnera perpensa	32	0	3	4
Bersama spp.	31	0	0	3
Cryptocarya spp.	31	0	0	3
Pittosporum viridiflorum	31	0	0	3
Pappea capensis	30	5	1	4
Cassipourea spp.	30	3	1	3
ESP 16 (umLulama)	30	0	0	3
Trichilia spp.	30	0	1	3
Stangeria eriopus	29	9	0	5
ESP 2 (uMadlozana)	29	2	2	4
ESP 3 (umKhwangu)	29	2	0	3
Maytenus undata	29	0	2	3
*Cinnamomum camphora**	28	7	1	3
Bridelia cathartica	28	1	0	3
ESP 6 (isiPhahluka)	28	1	0	3
Albizia adianthifolia	27	0	0	3
Sclerocarya birrea ssp.	27	0	0	3
ESP 1 (iLetha)	26	2	0	3
Diospyros spp.	25	0	0	2
ESP 15 (unSukumbili)	24	2	0	2
Myrothamnus flabellifolia	24	2	0	2
ESP 10 (uBhubhubhu)	24	1	0	2
Vernonia adoensis	24	0	0	2
Senecio coronatus	23	0	1	2
Cymbopogon spp.	22	1	4	2
Curtisia dentata	19	0	1	3
Siphonochilus aethiopicus	14	18	2	5
Brackenridgea zanguebarica	6	9	0	4

10.5.5 Plant risk analysis: 315 ethnospecies sold by 100 Faraday market traders

a) Hierarchical clustering: Ward's method.

In this sequence for 315 ethnospecies inventoried at the stalls of 100 Faraday market traders, four main cluster groups were identified (A, B1, B2 & C) at a linkage distance of ≈13 (45% level of relative distance) (Figure 7 and Appendix 2). These clusters were further aggregated into two larger groups (1 and 2). Given the characteristics of the species within the clusters, A and B1/B2/C are broadly categorised as higher and lower risk respectively (Table 7). Cluster A contains a mixture of highly traded and cited species of different plant part types that were present in the market in medium to large quantities. Cluster B and C contain species that are whole plants/roots/bark that are traded in smaller quantities but have higher harvesting risks than the species sold as leaves/stems/fruits in cluster C. Like the dendrogram for the *muti* shops (Figure 5), cluster A contains a mixture of all plant part types, whereas B and C are divided based on plant part type.

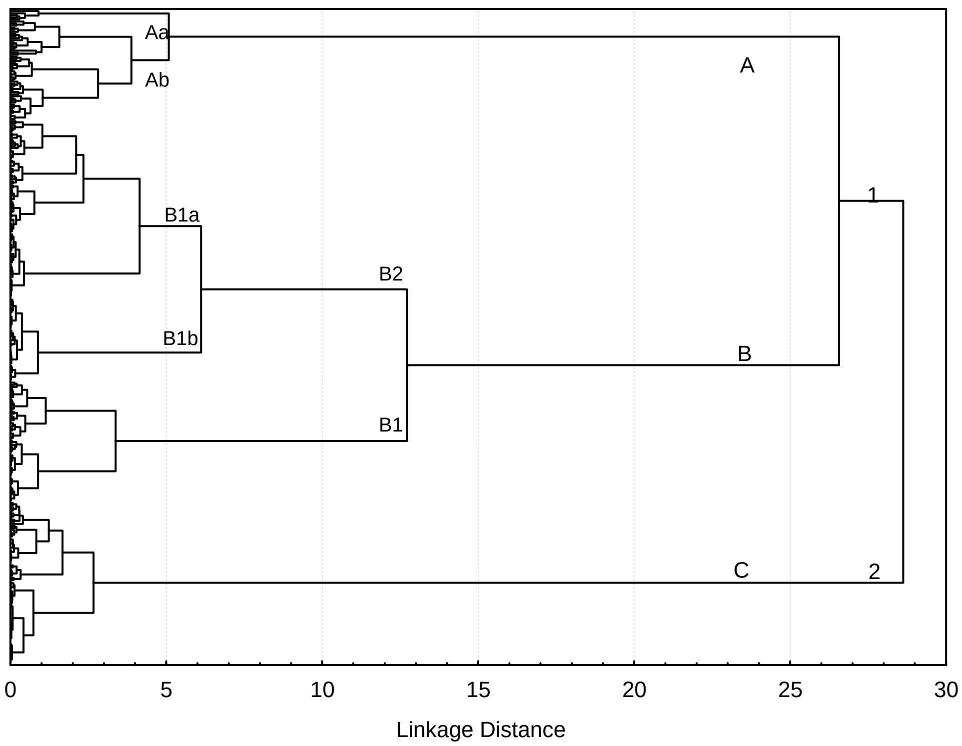


Figure 7: Dendrogram based on Euclidean distances and Ward's agglomerative clustering method for 315 ethnospecies inventoried with 100 Faraday street traders. The species names are present on the extended dendrogram in Appendix 2. Attributes of the species in the cluster are given in Table 9.

The attributes of the species in the clusters are as follows:

- **Cluster A** (16.8% of ethnospecies): species had a high prevalence in the market (especially sub-cluster Aa) and a high number of citations of scarcity and/or popularity. There was a range of plant part types. The five species in sub-cluster Aa were: *Drimia* spp. *H. africana*, *E. transvaalense*, *A. adianthifolia* and *A. xanthophloea*. Species in sub-cluster Ab include *W. salutaris*, and *S. aethiopicus*. Suggested risk categorisation: high
- **Cluster B1** (40.0% ethnospecies): species in this cluster were species sold for roots and bark. They had a medium to low prevalence in the shops and a low number of citations of scarcity and popularity. Species include *O. bullata*. Suggested risk categorisation: low
- **Cluster B2** (18.4% ethnospecies): species in this cluster were only whole plants. They had a medium to low prevalence in the market and a lower number of citations of scarcity and/or popularity. Suggested risk characterisation: low
- **Cluster C** (24.8% ethnospecies): species in this cluster were all leaves/stems and fruits - the harvesting impact of which is generally low. These species had a low prevalence in the market and were generally not regarded as scarce and/or popular. Suggested risk categorisation: low.

Table 7: Cluster attribute summary for 315 ethnospecies (esp.) sold by 100 Faraday traders corresponding with the dendrogram in Figure 7 and Appendix 2.

Group	Cluster	Sub-cluster	No esp. (%)	Number of esp. per plant part category					Total volume (bags)	Number of traders selling	Scarce citations	Popular citations
				Whole plants	Roots	Bark	Leaves/ stems	Fruit				
1	A	Aa	5 (1.6%)	1	1	3	-	-	12.1 – 24.9	38 ≤ x ≤ 61	6 – 10	19 – 27
		Ab	48 (15.2%)	11	9	25	3	-	0.3 – 17.5	8 ≤ x ≤ 40	1 – 10	1 – 22
		Subtotal	*53 (16.8%)*	*12*	*10*	*28*	*3*	-	*0.3 – 24.9*	8 ≤ x ≤ 61	*1 – 10*	*1 – 27*
	B1	B1a	86 (27.3%)	-	68	18	-	-	0.1 – 5.1	1 ≤ x ≤ 18	0 – 3	0 – 11
		B1b	40 (12.7%)	-	-	40	-	-	0.1 – 1.6	1 ≤ x ≤ 6	0 – 1	0 – 3
		Subtotal	*126 (40.0%)*	-	*68*	*58*	-	-	*0.1 – 5.1*	*1 ≤ x ≤ 18*	*0 – 3*	*0 – 11*
	B2		58 (18.4%)	58	-	-	-	-	0.1 – 6.9	1 ≤ x ≤ 15	0 – 6	0 – 9
		Subtotal	*184 (58.4%)*	*58*	*68*	*58*	-	-	*0.1 – 6.9*	*1 ≤ x ≤ 18*	*0 – 6*	*0 – 11*
237 (75.2%)												
2	C		78 (24.8%)	-	-	-	69	9	0.1 – 4.4	1 ≤ x ≤ 11	0 – 3	0 – 8
78 (24.8%)												
Total:			315 esp.	70	78	86	72	9	0.1 – 24.9	1 – 61	0 - 24	0 - 32

b) Non-hierarchical clustering: K-means.

i) 3 clusters (K_3)

The results for 3 (K_3) clusters were similar to the clusters found with Ward's method in terms of the species within a cluster and the number of species per cluster:

- **$K_3$1** (46 esp) corresponds with Ward's cluster **A** (53 esp). All the species in $K_3$1 are in cluster A.
- **$K_3$2** (139 esp) and **$K_3$3** (130 esp) combined correspond with **B** (184 esp) and **C** (78 esp) combined. However, the bark species are clustered with the leaves/stems/fruits in the K-means results instead of with whole plants/roots as in the Ward's results.

Results for the Euclidean distances between the clusters show that for 3 clusters, clusters $K_3$2 and $K_3$3 were closer together (Euclidean distance = 0.188) compared to the distance of $K_3$1 from clusters $K_3$2 and $K_3$3 (Appendix 8 Table c). This result would further support categorising clusters B and C as lower risk versus cluster A as higher risk in Ward's analysis.

Species grouped in $K_3$1 were sold by a mean of 25 Faraday traders, compared to means of 4 and 5 for $K_3$2 and $K_3$3 respectively (Table 8). Cluster $K_3$1 also had higher mean values for the number of bags present in the market and cited scarcity and popularity. This further validates the assumption that species in cluster $K_3$1 are higher risk. The distinctive nature of $K_3$1 was evident from the graph of standardized means in Figure 8a. While $K_3$2 and $K_4$3 had similar mean trade values, the species in $K_3$2 generally had lower harvesting risks due to the less destructive nature of plant part harvesting. Judging from the magnitude (and significance levels) of the *F* values, the *plant part* variable was the major criteria for assigning the species to clusters, followed by the *no. of traders* (Table 8). *Scarcity* had the least influence.

Table 8: Mean values per cluster (transformed from standardized values in Figure 8a to mean actual values) corresponding with the 3 clusters. Shaded values were the highest in the row, thereby implying higher associated risks. [*See Page 14 for plant part code]. [df = 2, 312; $P < 0.0001$ for all variables]

	$K_3$1	$K_3$2	$K_3$3	F
Mean no. traders selling	25	4	5	325.8
Mean no. bags	8.6	0.9	1.3	258.3
Mean no. scarce	4	1	1	154.9
Mean no. popular	12	2	2	274.8
Mean plant part*	3.6	2.4	4.5	383.7

ii) 4 clusters (K_4)

The results for 4 (K_4) clusters did not entirely correspond with clusters formed by Ward's method:

- **$K_4$1** (40 esp) corresponded with Ward's sub-cluster **A** (53 esp). All plant part types were represented.
- **$K_4$2** (131 esp) and **$K_4$3** (66 esp) corresponded with **B1** and **B2**, the only difference was the cluster that the roots were placed in.
- **$K_4$4** (78 esp,) is exactly the same as cluster **C**.

Euclidean distances between the clusters show that for 4 clusters, clusters $K_4$3 and $K_4$4 were the closest together (Euclidean distance = 0.103) (Appendix 8 Table d). $K_4$1 was the most distant from $K_4$4 (Euclidean distance = 0.379) and was also quite dissimilar to $K_4$2 and $K_4$3

Figure 8b and Table 9 shows the standardized and transformed actual mean values per variable per cluster respectively. Results show that species in $K_4$1 are sold by a mean of 27 street traders, compared to means of ≤ 5 for $K_4$2, $K_4$3 and $K_4$4. It is evident from Figure 11 that $K_4$1 is distinctively different from $K_4$2, $K_4$3 and $K_4$4 (the latter differing mainly in plant part composition), further justifying the risk categorization of species in $K_4$1 as high, and the remainder of the species as low. From the magnitude of the *F*-values, *plant part* is the major criterion for assigning species to clusters, and *Scarcity* as the least important criteria (Table 9).

Table 9: Mean values per cluster (transformed from standardized values in Figure 8b to mean actual values) corresponding with the 4 clusters. Shaded values were the highest in the row. [*See Page 14 for plant part code]. [df = 3, 311; P < 0.0001 for all variables]

	$K_4 1$	$K_4 2$	$K_4 3$	$K_4 4$	F
No. traders selling	27	5	5	3	227.6
No. bags	9.4	1.3	1.4	0.7	213.8
No. scarce	4	1	1	0	83.1
No. popular	13	2	2	1	194.7
Plant part*	3.7	4.5	3.0	1.9	457.5

iii) 5 clusters (K_5)

The results for 5 (K_5) clusters also corresponded with the clusters formed by Ward's method. $K_4 1$ divided into 2 new clusters. The first cluster, $K_5 3$ (6 esp) corresponds with Ward's Aa (5 esp) and $K_5 4$ (44 esp) corresponded with Ab (48 esp). The six species in $K_5 3$ were *Drimia* spp., *H. africana*, *E. transvaalense*, *Hypoxis* spp. and *A. xanthophloea*. In most of the previous and following cluster analyses, these six species repeatedly reoccur as priority species in the Witwatersrand *muti* trade and are therefore higher risk than species in $K_5 3$/Ab. Clusters $K_5 2$, $K_5 5$ and $K_5 1$ correspond with B1, B2 and C respectively. Again, the main differences were the clusters in which bark or roots are placed.

The Euclidean distances between the clusters indicated that for 5 clusters, clusters $K_5 2$ and $K_5 5$ were the closest together (Euclidean distance = 0.095) (Appendix 8 Table e). The most distant clusters were $K_5 3$ and $K_5 2$ (Euclidean distance = 0.64).

Species in $K_5 3$ and $K_5 4$ were sold by a mean of 46 and 21 street traders respectively, compared to means of ≤6 for $K_5 1$, $K_5 2$ and $K_5 3$ (Table 10). It was evident from Figure 8c that $K_5 4$ and $K_5 3$ were distinctively different from $K_5 1$, $K_5 2$ and $K_5 3$ (the latter differing mainly in plant part composition), further justifying the risk categorization of species in $K_5 4$ and $K_5 3$ as high, especially $K_5 4$. The species in the remainder of the clusters were lower risk and lower priority. The major criteria for assigning species to clusters in this analysis were the *no. of traders* and the *no. of bags* (Table 10). Least important was *scarcity*.

Table 10: Mean values per cluster (transformed from standardized values in Figure 8c to mean actual values) corresponding with the 4 clusters. Shaded values were the highest in the row. [*See Page 14 for plant part code]. [df = 4, 310; P < 0.0001 for all variables]

	$K_5 1$	$K_5 2$	$K_5 3$	$K_5 4$	$K_5 5$	F
Mean no. traders selling	4	4	46	21	6	335.8
Mean no. bags	0.9	1.0	17.9	6.9	1.5	289.8
Mean no. scarce	1	1	7	3	1	105.7
Mean no. popular	1	2	22	10	2	220.8
Mean plant part*	2.4	4.0	3.8	3.5	5.0	286.3

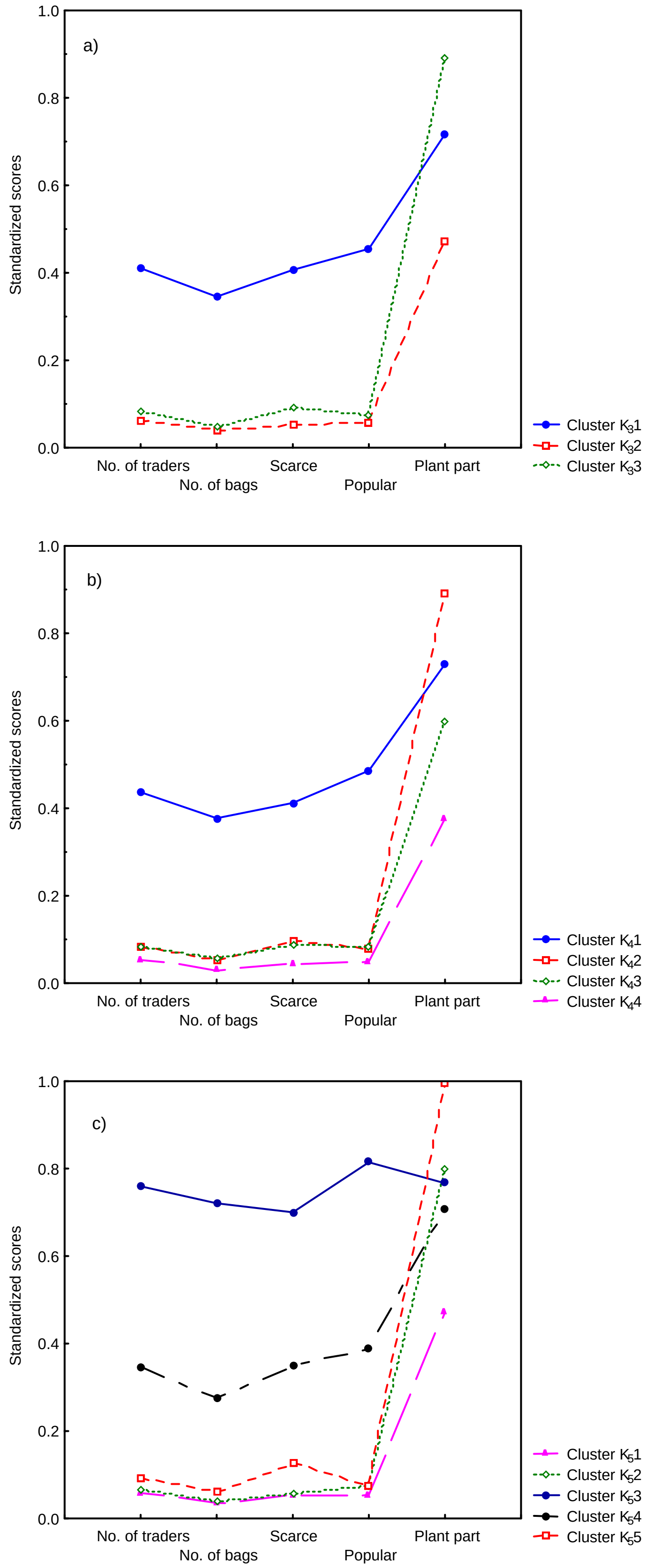


Figure 8: Plot of standardised cluster means for each k-means cluster for a) 3 clusters, b) 4 clusters and c) 5 clusters for species sold in the Faraday market. See Tables 8, 10 and 11 for the transformed standardised mean scores for a), b) and c) respectively.

c) Priority species sold by the Faraday street traders

By comparing and combining the species in clusters A and $K_3$1, a short-list of 53 species were identified as higher risk/vulnerable due to the extent of trade in the Faraday street market.

Table 11: Species sold at the Faraday market flagged as high risk/vulnerability by K-means and Ward's cluster analysis. Species are listed in descending order of the number of traders selling the species and the number that cited it as scarce. [See Page 14 for plant part code]

Probable species	No. of traders selling (N_{max}=100)	Bags (50 kg-size)	Scarce	Popular	Plant part
Drimia spp.	61	24.9	10	27	5
Hydnora africana	49	19	8	19	4
Elaeodendron transvaalense	48	16.1	7	27	3
Albizia adianthifolia	42	18.1	6	23	3
Hypoxis spp.	40	17.5	5	13	5
Acacia xanthophloea	38	12.1	6	23	3
Urginea spp.	32	9.3	5	13	5
Schotia brachypetala	32	11.2	3	22	3
Sclerocarya birrea ssp *birea*	32	13.2	3	17	3
Adenia gummifera var.	29	8.9	2	17	2
Dioscorea sylvatica	29	7.5	2	13	5
Rapanea melanophloeos	27	10.2	4	13	3
Merwilla plumbea	27	10.1	3	10	5
Clivia spp.	25	11.2	6	11	5
ESP 20 (uMabophe)	25	8.7	6	12	4
Curtisia dentata	25	9.2	4	16	3
ESP 16 (umLulama)	25	9.3	4	15	3
Rhoicissus tridentata	24	6.1	5	8	5
ESP 29 (iNyazamgoma-ebomvu)	24	8	3	14	3
ESP 6 (isiPhahluka)	24	10.7	2	9	3
ESP 5 (umLahlankosi)	24	8.7	1	13	3
Ornithogalum longibracteatum	24	7.6	1	12	5
Trichilia spp.	24	9.8	1	17	3
Callilepis laureola	23	4.1	5	10	5
Ekebergia capensis	23	7.8	4	13	3
ESP 14 (uBangalala)	23	2.3	3	2	4
Eucomis autumnalis ssp.	22	6.5	5	5	5
Thesium pallidum	22	2.3	5	9	4
Warburgia salutaris	21	4.9	10	7	3
ESP 17 (uVelabahleka)	21	7.3	5	9	4
Dombeya rotundifolia var.	21	8.5	3	10	3
ESP 1 (iLetha)	21	7.9	2	14	3
Maytenus undata	21	6.3	2	15	3
Gunnera perpensa	19	8.6	7	4	4
Balanites maughamii	19	6.9	4	8	3
Talinum caffrum	19	5.1	4	7	5
Euclea spp.	18	3.7	3	10	3
Raphionacme sp.	17	3.5	3	7	4
Helichrysum spp.	17	10.6	2	10	2
Ptaeroxylon obliquum	17	5.8	2	12	3
ESP 2 (uMadlozana)	16	5.9	3	8	4
ESP 22 (uSehlulamanye)	16	3.6	3	8	3
ESP 3 (umKhwangu)	16	6.3	2	10	3
Macaranga capensis	14	3.6	4	6	3
Olinia radiata	13	3.1	5	5	3
Rauvolfia caffra	13	3.4	4	6	3
ESP 12 (umDlavusa)	13	2.7	3	8	3
Justicia capensis	13	4.6	2	8	2
ESP 31 (umVangasi)	12	5.4	2	6	3
Boscia spp.	11	5.1	0	3	4
Acalypha villicaulis	10	2.5	3	8	4
Garcinia spp.	10	3.3	2	6	3
Siphonochilus aethiopicus	8	0.3	6	1	5

10.5.6 Plant risk analysis: 392 ethnospecies sold by 150 traders (‘All’, Shop and Faraday combined)

a) Hierarchical clustering: Ward’s method.

In this sequence of 392 ethnospecies resulting from combining the Shop and Faraday data sets, four main cluster groups were identified (A, B, C & D) at a linkage distance of ± 19.5 (45% level of relative distance) (Figure 9 and Appendix 3). These clusters were further aggregated into two larger groups (1 and 2). Sub-clusters Ba/Bb and Da/Db were also recognized at 10 (24%) and 9 (20%) linkage distances respectively. Given the characteristics of the species within the clusters, A was characterized high risk, B as medium risk (with Ba being medium-high), and C/D as low risk (Table 18).

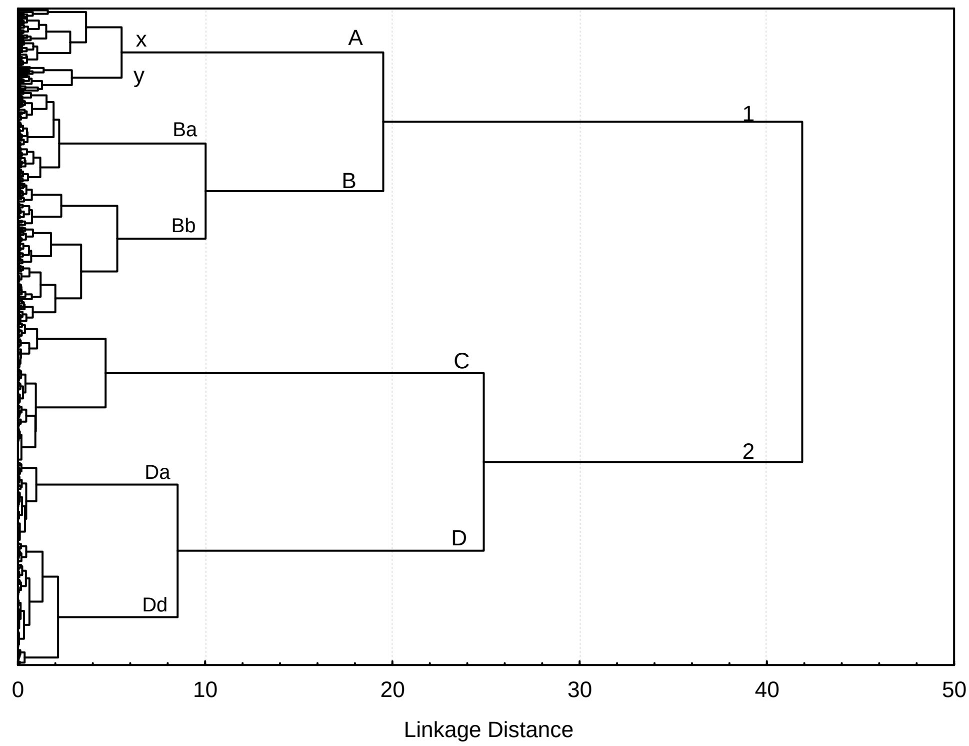


Figure 9: Dendrogram based on Euclidean distances and Ward’s agglomerative clustering method for 392 ethnospecies sold between 150 traders. The species names are present on the extended dendrogram in Appendix 3. Attributes of the species in the cluster are given in Table 12. Labels X and Y are discussed in Section 10.5.9a.

The attributes of the species in the clusters were as follows:

- **Cluster A** (12.5% of ethnospecies): species had a medium to high prevalence and a higher number of citations of scarcity and popularity. There was a range of plant part types, and species were present in the market in mostly large quantities (with the exception of *S. aethiopicus*). Suggested risk categorisation: high
- **Cluster B** (35.2% ethnospecies): species in this cluster also ranged in plant part type and had a medium to low prevalence in the market. Citations of scarcity and popularity were low and the quantities present were medium to low. Sub-cluster Ba contained species that generally had a higher harvesting risk due to the plant part harvested, and their prevalence in the markets was generally higher than sub-cluster Bb. Suggested risk categorisation: medium, with Ba possibly being medium-high.
- **Cluster C** (21.4% ethnospecies): species in this cluster were only whole plants and roots. They had a low prevalence in the market, were traded in small quantities and had almost no citations of scarcity and popularity. Suggested risk characterisation: low
- **Cluster D** (30.98% ethnospecies): species in this cluster were all bark (sub-cluster Da) and leaves/stems and fruits (Db) - the harvesting impact of which is generally low. These species also had a low prevalence in the market, and were generally not regarded as scarce and/or popular. Suggested risk categorisation: low.

Table 12: Cluster attribute summary for 392 ethnospecies (esp.) sold by 150 traders corresponding with the dendrogram in Figure 9 and Appendix 3.

Group	Cluster	Sub-cluster	No esp. (%)	Number of esp. per plant part category: Whole plants	Roots	Bark	Leaves/ stems	Fruit	Total volume for sale (bags)	Number of traders selling	Scarce citations	Popular citations
1	A	A	49 (12.5%)	14	8	24	3	-	0.3‡ – 24.9†	22 ≤ x ≤ 106	1 – 24†	1 – 32†
	B	Ba	55 (14.0%)	32	20	3	-	-	0.1 – 7.6	16 ≤ x ≤ 59	0 – 6	0 – 8
		Bb	83 (21.2%)	-	25	32	22	4	0.1 – 8.7	3 ≤ x ≤ 41	0 – 9	0 – 14
		Subtotal	*138 (35.2%)*	*32*	*45*	*35*	*22*	*4*	*0.1 – 8.7*	*3 ≤ x ≤ 59*	*0 – 9*	*0 – 14*
187 esp. (47.7%)												
2	C	C	84 (21.4%)	30	54	-	-	-	0.1 – 2.8	1 ≤ x ≤ 10	0 – 2	0 – 4
	D	Da	48 (12.2%)	-	-	48	-	-	0.1 – 1.6	1 ≤ x ≤ 11	0 – 1	0 – 3
		Dd	73 (18.6%)	-	-	-	64	9	0.1 – 1.9	1 ≤ x ≤ 10	0 – 2	0 – 3
		Subtotal	*121 (30.9%)*	-	-	*48*	*64*	*9*	*0.1 – 1.9*	*1 ≤ x ≤ 11*	*0 – 2*	*0 – 3*
205 esp. (52.3%)												
Total:			392 esp.	76	107	107	89	13	0.1 – 24.9	1 – 106	0 – 24	0 – 32

†: *Drimia* spp; ‡ *Siphonochilus aethiopicus*

b) Non-hierarchical clustering: K-means.

A general observation from the species partitions resulting from Ward's and K-means clustering for the combined data set was, that there was a lower level of correspondence when compared with the results for the individual Shop and Faraday data sets. Eight variables were used to partition and cluster the species in this data set (compared to 4 and 5 variables in the Shop and Faraday data sets respectively). Hence, the algorithm used more variables to produce internally homogenous clusters.

i) 3 clusters (K_3)

The results for 3 (K_3) clusters were similar to the clusters found with Ward's method in terms of the species within a cluster and the number of species per cluster:

- **$K_3$1** (62 esp) mainly corresponded with Ward's cluster **A** (49 esp), but also included species in **Ba**.
- **$K_3$2** (162 esp) mainly corresponded with **C** (84 esp), but also contained species from **B** and **Da**. All species that were whole plants and roots were present in this cluster, but there were also a few bark species as well.
- **$K_3$3** (168 esp) mainly corresponded with **D** (121 esp), but contained species present in **B**. All species that were leaves/stems and fruits were present in this cluster, with the remainder of the bark species present as well.

Clusters $K_3$2 and $K_3$3 are closer together (Euclidean distance = 0.145) compared with the distance of $K_3$1 from clusters $K_3$2 and $K_3$3 (Euclidean distance >0.26) (Appendix 8 Table f). Given the similarities of $K_3$2 and $K_3$3 with C and D, these clusters could be classified as lower risk and $K_3$1 as higher risk.

Species grouped in $K_3$1 were sold by a mean of 21 Faraday traders and 29 shop traders (Table 13). Additionally, the mean plant volume present per species in Faraday was 6.8 bags, and citations of scarcity and popularity were higher than for the other two clusters. $K_3$2 and $K_3$3 were similar in terms of their citations for scarcity, popularity and the mean number of street traders selling the species. $K_3$2, however, had a higher mean number of shop traders selling the species and plants that were mainly sold whole or as roots. The distinctive nature of the clusters was evident in Figure 10a. Judging from the magnitude (and significance levels) of the *F* values, *plant part* was the major criteria for assigning the species to clusters, followed by the *no. of Faraday traders* and the *no. of shops*. *Shop scarcity* had the least influence (Table 13).

Table 13: Mean values per cluster (transformed from standardized values in Figure 10a to mean actual values) corresponding with 3 clusters. Shaded values are the highest in the row. [*See Page 14 for plant part code]. [df = 2, 389; P < 0.0001 for all variables]

	$K_3$1	$K_3$2	$K_3$3	F
Mean no. Faraday traders selling	21	4	3	276.6
Mean no. bags	6.8	1.0	0.7	184.8
Mean no. scarce (Faraday)	4	1	0	202.5
Mean no. popular (Faraday)	10	2	1	196.7
Mean no. shops selling	29	10	4	226.8
Mean no. scarce (Shops)	2	1	0	36.5
Mean no. popular (Shops)	1	0	0	67.0
Mean plant part*	3.7	4.3	2.3	359.3

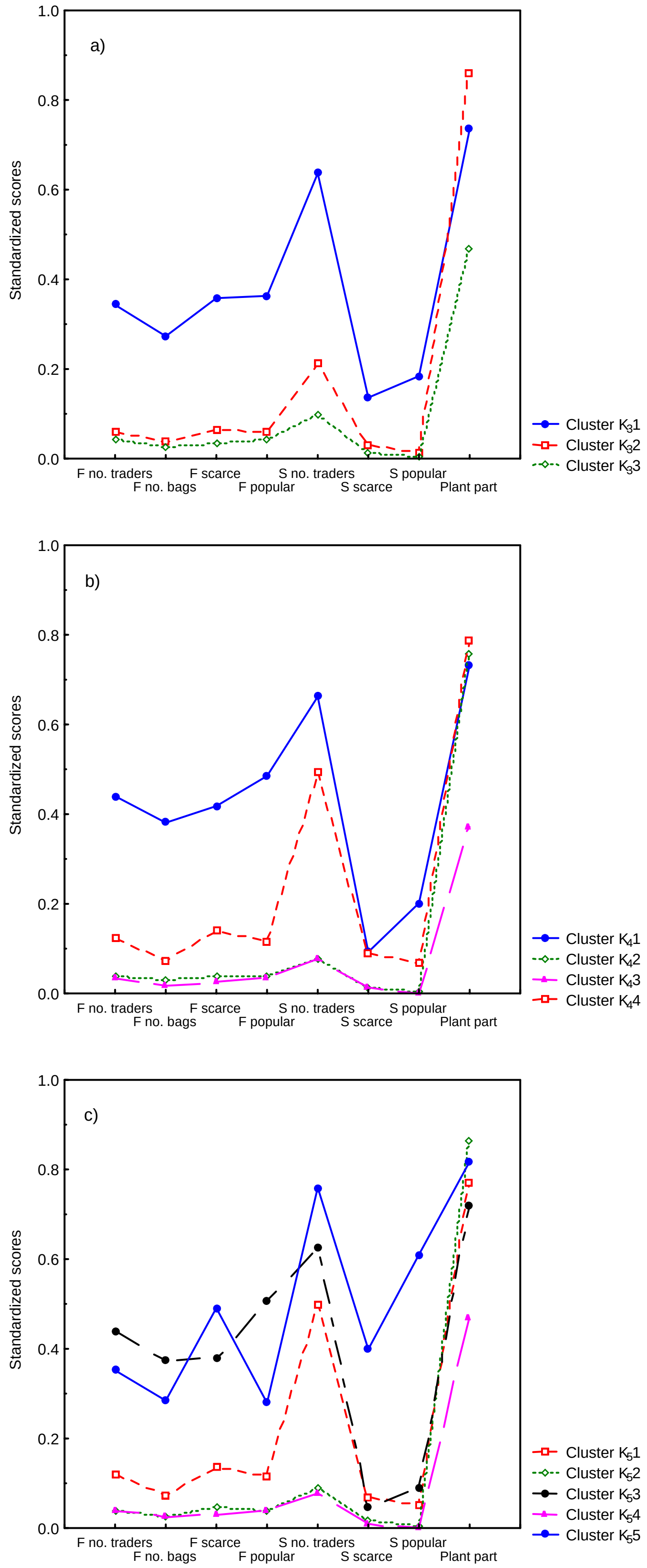


Figure 10: Plot of standardised cluster means for each k-means cluster for a) 3 clusters, b) 4 clusters and c) 5 clusters for the combined data sets of Faraday and Shops ('All'). See Tables 13, 14 and 15 for the transformed standardised mean scores for a), b) and c) respectively. (F=Faraday; S=Shops)

ii) 4 clusters (K_4)

Results for 4 (K_4) clusters have a lower level of correspondence with clusters formed by Ward's method:

- **$K_4$1** (39 esp) and **$K_4$4** (90 esp): all species in these clusters were present in Ward's **A** (49 esp) and **B** (138 esp). The remainder of species from A and B not present in $K_4$1 and $K_4$4 were present in **$K_4$2** and **$K_4$3**. All plant part types were represented. This is a cluster of high to medium-high risk species.
- **$K_4$2** (172 esp): all species were whole plants, roots and bark. The cluster was more like **C** (84 esp), but also contained species from **Bb** and **Da**.
- **$K_4$3** (91 esp): all species were leaves/stems and fruits. The cluster was more like **D** (121 esp), but also contained species from **Bb**.

Clusters $K_4$2 and $K_4$3 were the closest together (Euclidean distance = 0.136) (Appendix 8 Table g). $K_4$1 was the most distant from $K_4$3 (Euclidean distance = 0.383) and was similarly dissimilar to $K_4$2. Results suggest species in $K_4$1 and $K_4$4 to be high and medium-high risk respectively, compared to low risk for $K_4$2 and $K_4$3.

Species in $K_4$1 were sold by a mean of 27 street traders and 30 shop traders, compared to means of ≤3 for $K_4$2 and $K_4$3 (Table 14). Species in cluster $K_4$4, while present in a relatively high number of shops (mean=22), were sold by a relatively low number of street traders. The number of citations of scarcity and popularity, as well as the volume, for species in clusters $K_4$2, $K_4$3 and $K_4$4 were low compared to species in $K_4$1. It is evident from Figure 10b that $K_4$1 was distinctively different from $K_4$2 and $K_4$3, whereas $K_4$4 was intermediate between them. Risk categorizations for $K_4$1, $K_4$4 and $K_4$2/ $K_4$3 are thus proposed as high, medium and low risk respectively. From the magnitude of the *F*-values, *no. of shops* followed by number of *Faraday traders,* was the major criterion for assigning species to clusters (Table 14). *Shop scarcity* was the least important criterion.

Table 14: Mean standardised values per cluster (transformed from standardized values in Figure 10b to mean actual values) corresponding with 4 clusters. [*See Page 14 for plant part code] [df = , 388; P < 0.0001 for all variables]

	$K_4$1	$K_4$2	$K_4$3	$K_4$4	F
Mean no. Faraday traders selling	27	2	2	8	355.8
Mean no. bags	9.5	0.7	0.4	1.8	294.8
Mean no. F. scarce	4	0	0	1	144.0
Mean no. F. popular	13	1	1	3	260.0
Mean no. shops selling	30	3	3	22	513.1
Mean no. S. scarce	2	0	0	2	16.5
Mean no. S. popular	1	0	0	0	38.8
Mean plant part*	3.7	3.8	1.9	3.9	153.1

iii) 5 clusters (K_5)

The results for 5 (K_5) clusters also have a lower level of correspondence with the clusters formed by Ward's method:

- **$K_5$5** (11 esp) and **$K_5$3** (32 esp) corresponded with species in Ward's cluster **A** (49 esp).
- **$K_5$1** (80 esp) corresponded with species in cluster B (138 esp).
- **$K_5$2** (110 esp) contain only species that were whole plants and roots. All species from **C** and some from **D** were in this group.
- **$K_5$4** (159 esp) contains only species that were bark, leaves/stems and fruits. All species from **D** and some from **Bb** were in this group. $K_5$3 probably indicates species that are higher risk than $K_5$4.

In the formation of 5 clusters, $K_5$5 represented 11 species known to be high risk and/or vulnerable, namely: *Drimia* spp., *M. plumbea*, *E. autumnalis*, *W. salutaris*, *Helichrysum* spp., *G. perpensa*, *O. bullata*, *D. mooiensis* and *S. aethiopicus*. These species are repeatedly flagged in this study and others. $K_5$3 represents a group of 32 medium-high risk species (including *E. transvaalense*, *Hypoxis* spp., *H. africana*, *D. sylvatica*, *Clivia* spp. *et al.*). $K_5$1 corresponds to a group of 80 medium-low risk species, and there were 2 groups of lower risk species ($K_5$2 and $K_5$4) partitioned on the basis of plant part.

Species in $K_5$3 and $K_5$5 were sold by a mean of 27 and 21 street traders respectively, as well as a mean of 28 and 34 shops respectively (Table 15). The volume of plants present in the Faraday market was relatively high, as well as the number of citations of scarcity and popularity (Table 15). By comparison, species in $K_5$1 had a medium prevalence in Faraday, but a relatively high prevalence in the shops (means of 7 and 23 respectively). Means for variables in $K_5$3 and $K_5$4 indicate these species are lower risk, and that the clusters were mainly partitioned based on plant part (Figure 10c). Based on the significance and magnitude of the *F* values, the major criteria for assigning species to clusters in this analysis was the *no. of shop traders*, the *no. of street traders* and the *no. of bags* (Table 15). Least important was *scarcity*.

Table 15: Mean standardised values per cluster (transformed from standardized values in Figure 10c to mean actual values) corresponding with 5 clusters. [*See Page 14 for plant part code] [df = 4, 387; $P < 0.0001$ for all variables]

	$K_5$1	$K_5$2	$K_5$3	$K_5$4	$K_5$5	F
Mean no. Faraday traders selling	7	2	27	2	21	241.0
Mean no. bags per species (Faraday)	1.8	0.7	9.3	0.6	7.1	171.1
Mean no. scarce (Faraday)	1	0	4	0	5	114.3
Mean no. popular (Faraday)	3	1	14	1	8	183.5
Mean no. shops selling	23	4	28	3	34	411.6
Mean no. scarce (shops)	1	0	1	0	7	55.0
Mean no. popular (shops)	0	0	1	0	4	171.3
Mean plant part*	3.9	4.3	3.6	2.3	4.1	138.4

A general comment on K-means results for all data sets is that the analysis of variance indicates *scarcity* to be the least significant criterion for partitioning species in the clusters. *Plant part*, the *no. of traders* selling a species and the *no. of bags* were the most important criteria.

c) Priority species sold by 150 Witwatersrand traders

Comparing species from Ward's clusters A and Ba with species from $K_5 5$, $K_5 3$ and $K_5 1$, a group of 119 species were short-listed as high to medium-high risk (Table 16). It is from this list of species in Table 16 that 87 species were selected for further study (see following Section 10.5.7). The criteria for excluding species from the final group were exotics and ethnospecies with multiple genera.

Table 16: Short-list of 119 species sold by shop and street traders flagged as high risk/vulnerable by K-means and Wards cluster analysis. Species are listed in descending order of the number of traders selling the species and the number that cited it as scarce. *exotic [See Page 14 for plant part code]

Probable species	No. of traders selling (N_{max}=150)	Total Scarce (M_{max}=24)	Total Popular (N_{max}=32)	Bags (50 kg-size)	Plant part
Drimia spp.	106	23	32	24.9	5
Elaeodendron transvaalense	81	8	31	16.1	3
Acacia xanthophloea	73	8	23	12.1	3
Hypoxis spp.	73	6	13	17.5	5
Hydnora africana	71	8	19	19	4
Albizia adianthifolia	69	6	23	18.1	3
Rapanea melanophloeos	69	6	16	10.2	3
Adenia gummifera var.	65	3	19	8.9	2
Merwilla plumbea	63	9	17	10.1	5
Urginea spp.	63	9	13	9.3	5
Eucomis autumnalis	62	21	12	6.5	5
EP 14 (uBangalala)	59	10	4	2.3	4
Dioscorea sylvatica	59	5	13	7.5	5
Sclerocarya birrea ssp. *birrea*	59	3	17	13.2	3
Clivia spp.	58	7	11	11.2	5
EP 17 (uVelabahleka)	58	6	12	7.3	4
EP 20 (uMabophe)	57	7	13	8.7	4
Warburgia salutaris	55	13	11	4.9	3
Ekebergia capensis	55	5	13	7.8	3
EP 16 (umLulama)	55	4	15	9.3	3
Talinum caffrum	54	5	8	5.1	5
Trichilia spp.	54	1	18	9.8	3
Schotia brachypetala	53	3	22	11.2	3
Helichrysum spp.	53	2	16	10.6	2
Callilepis laureola	52	5	10	4.1	5
Balanites maughamii	52	4	8	6.9	3
EP 6 (isiPhahluka)	52	3	9	10.7	3
Croton spp.	52	3	8	4.4	3
Gunnera perpensa	51	7	7	8.6	4
EP 22 (uSehlulamanye)	50	3	9	3.6	3
Maytenus undata	50	2	17	6.3	3
Rhoicissus tridentata	49	5	8	6.1	5
Thesium pallidum	48	5	10	2.3	4
EP 1 (iLetha)	47	4	14	7.9	3
EP 2 (uMadlozana)	45	5	10	5.9	4
EP 3 (umKhwangu)	45	4	10	6.3	3
Curtisia dentata	44	4	17	9.2	3
Ocotea bullata	43	16	6	1.7	3
Alepidea amatymbica	43	6	6	1.3	5
Dombeya rotundifolia	43	3	10	8.5	3
Raphionacme sp.	43	3	8	3.5	4
Ornithogalum longibracteatum	43	2	12	7.6	5
EP 11 (umVuthuza)	41	3	3	0.6	2
Elephantorrhiza elephantina	41	2	6	3	5
Bowiea volubilis	40	12	1	2.8	5
Calodendrum capense	40	6	3	2.3	3
Dianthus mooiensis	40	2	10	0.7	5
Tulbaghia spp.	39	5	3	2.7	5
Boophone disticha	39	2	7	2.9	5
Stangeria eriopus	38	11	3	2.9	5
EP 29 (iNyazamgoma-ebomvu)	38	3	14	8	3
Secamone gerrardii	37	4	2	1.8	5
Dietes iridioides	37	1	1	0.7	5

Bridelia cathartica	36	2	2	2.7	3
Hippobromus pauciflorus	36	0	7	2.6	4
EP 66 (uMusa)	35	2	6	0.7	4
Pentanisia prunelloides	35	2	1	1.7	5
EP 5 (umLahlankosi)	35	1	14	8.7	3
EP 41 (inhlanhlomhlope)	35	1	3	4.4	2
Scabiosa columbaria	35	1	3	0.9	5
*Cinnamomum camphora**	34	9	2	1.5	3
EP 15 (unSukumbili)	34	5	3	2.7	2
Gnidia kraussiana var.	34	2	3	3	5
Vernonia tigna	34	2	3	0.6	4
Bersama spp.	34	0	1	0.5	3
Capparis spp.	34	0	7	3.5	4
Cassipourea spp.	33	6	1	1.4	3
Eulophia speciosa	33	3	5	3.6	5
Aster bakerianus	33	2	2	2.2	4
Cymbopogon spp.	33	2	9	1.7	2
Synaptolepis kirkii	33	2	0	0.5	4
EP 4 (inGcino)	33	1	4	2.3	5
Cryptocarya spp.	33	0	2	0.1	3
EP 10 (uBhubhubhu)	32	2	3	2.4	2
Pappea capensis	31	5	1	0.1	4
Eriospermum mackenii	31	3	5	2.1	5
Acalypha villicaulis	30	4	8	2.5	4
Agathosma ovata	30	2	6	2	4
Cephalaria humilis	30	2	4	2.4	5
EP 24 (unDwendweni, umLunge))	30	2	6	3.7	5
EP 32 (uBhoqo)	30	0	5	1.8	4
Senecio coronatus	29	3	1	1.2	2
Diospyros spp.	29	2	2	1.3	2
Myrothamnus flabellifolia	29	2	2	1.6	2
Aristea spp.	29	1	2	0.6	5
*Acorus calamus**	28	3	1	1.1	4
EP 12 (umDlavusa)	28	3	8	2.7	3
EP 45 (umLomomnandi)	28	2	4	0.5	4
Zanthoxylum spp.	28	2	5	3.1	4
Dioscorea dregeana	27	8	1	4.2	5
Stapelia gigantea	27	3	9	2.5	5
EP 30 (uDabulamafu)	27	1	0	0	5
EP 36 (iShongwe)	27	1	4	1.1	4
EP 47 (abaNqonqosi)	27	1	0	0.5	3
Asparagus spp.	27	0	0	0.5	4
Rubia cordifolia	25	3	1	0	4
Polygala spp.	24	3	2	1.3	5
Eucomis bicolor	24	2	7	6.9	5
Gnidia spp.	24	2	2	0.8	5
Pleurostylia capensis	24	2	3	1.9	3
Crinum spp.	23	2	1	2.9	5
Siphonochilus aethiopicus	22	24	3	0.3	5
Haworthia spp.	22	4	4	0.8	5
Andrachne ovalis	22	1	1	0.2	4
EP 65 (iZaza)	22	0	2	0.3	4
Euphorbia woodii	22	0	3	0.6	5
Anemone spp.	21	2	2	1.4	4
Pelargonium luridum	21	2	5	1.7	5
Plectranthus sp.	20	5	2	2.2	4
Vitex spp.	20	3	2	1.5	3
EP 34 (unGibonisele)	20	1	1	1.5	5
Agapanthus spp.	19	1	2	0.5	5
EP 27 (iNyongwane)	19	0	3	0.7	4
EP 33 (uVimbokalo)	19	0	1	2.3	4
Eriosema sp.	18	1	3	0.6	4
Pinus sp.*	17	3	0	0	3
Ansellia africana	17	2	1	0.7	5
Silene bellidioides	16	3	3	0.5	5
EP 25 (uPhamapuce)	16	2	0	0.6	5

10.5.7 Plant risk analysis: top 87 species

In the previous sections, cluster analysis was used as a 'statistical filter' to group species into broad risk hierarchies and to prepare a short-list of species for further analysis. Species in the different clusters and risk hierarchies require different management actions. From the group of 119 ethnospecies flagged as high to medium-high risk, a final group of 87 species was selected. Ethnospecies were excluded from the final list if they were exotics and/or from a group of multiple genera (i.e. the 'ESP' numbers). The purpose was to propose threat and conservation priority profiles for the species by grouping species with similar priorities into hierarchical clusters.

In addition to the trade variables, biological variables that can be used to assess conservation priorities were incorporated in the analyses. These variables were: *phylogenetic distinctness, habitat specificity*, *endemism* and *QGrids*.

a) Hierarchical clustering: Ward's method.
The hierarchical cluster analysis results of the top 87 ethnospecies are shown in the dendrogram of Figure 11. The attributes of the species in the cluster sequence are summarised in Table 17, and the species-attribute matrix is shown in Appendix 5.

Three main cluster groups were identified (A, B and C) at a linkage distance of ≈5.75 (74%) (Figure 11), and four clusters were identified at ≈4.5 (58%) (A, B1, B2 and C). Seven sub-clusters were also recognized (Aa, Ab, B1a, B1b, B2, Ca, Cb), but these were formed at a linkage distance of <30%. Given the characteristics of the species within the clusters, A/B and C were characterised as higher and lower conservation priority respectively. However, given that the 87 species were short-listed from the original group of 392, their conservation priority is higher overall than the 305 species not short-listed for further investigation.

The attributes of the species in the clusters were as follows:

- **Cluster A**: species in this cluster had higher values for the trade variables and lower scores for the biological variables. The prevalence and volume of these species in the market is high, as were the number of citations for scarcity and popularity (especially Aa). Overall, however, the species were less habitat specific (i.e. found in ≥2 habitat types), were not endemic to South Africa, and had lower levels of *phylogenetic distinctness*. Conservation priorities proposed for species in this cluster are high. There is only one species with a Red List score, namely *Dombeya rotundifolia* (RL=1 i.e. 'Least concern').
- **Cluster B**: while species in this cluster exhibited lower volumes in trade and were sold by comparatively fewer traders, some of these species are known from other studies to be vulnerable to trade and extensive harvesting is known to have diminished their availability, e.g. *W. salutaris* and *S. aethiopicus*. Scores for biological variables indicate that species have a higher risk and conservation priority than species in cluster C. Species in this group also tended to have smaller distributions (*QGrids*), habitats that are more restricted and, a greater level of *endemism* and *phylogenetic distinctness* than species in clusters A or C. It is within this cluster that most of the RL species are located.
- **Cluster C**: species in this cluster have over-all lower trade factor values, larger distributions and fewer biological factors indicating higher vulnerability to harvesting. Conservation priorities for these species are lower than A and B. There are four species with a Red List score of 1 (i.e. 'Least concern'), namely *Agathosma ovata*, *Bridelia cathartica*, *Crinum moorei* and *Schlechterina mitostemmatoides*.

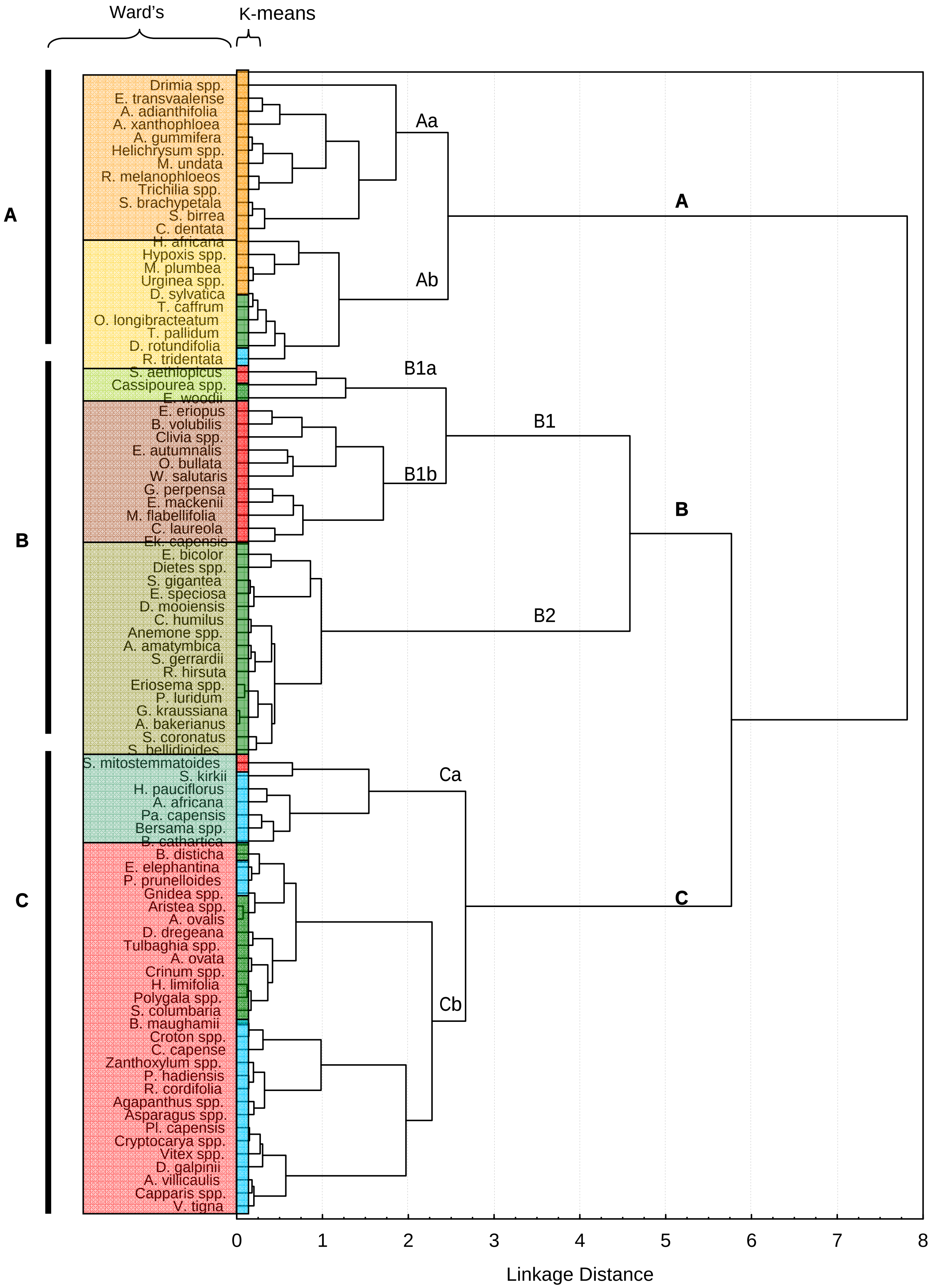


Figure 11: Dendrogram based on Euclidean distances and Ward's agglomerative clustering method for 87 short-listed species. Attributes of the species in the cluster are given in Table 17. The narrow shaded areas indicate the clusters created by K-means analysis for 4 clusters (see Table 19, page 48).

Table 17: Cluster attribute summary for 87 short-listed species (esp corresponding with the dendrogram in Figure 11 and the species-attribute matrix in Appendix 5.

			Trade factors				Biological factors				
Cluster	Sub-cluster	No species (%)	No. of traders selling	Scarce citations	Popular citations	No. bags	QGrids	Habitat specificity	Endemism	Phylogenetic distinctness	Red List
A	Aa	12 (13.8%)	44 ≤ x ≤ 106	1 – 23	16 – 32	6.3 – 24.9	22 – 432	1 – 2	1	1 – 4	0
	Ab	10 (11.5%)	43 ≤ x ≤ 73	0 – 9	8 – 19	2.3 – 19.0	15 – 300	1 – 3	1	1 – 2	1
Subtotal A		***22 esp (25.3%)***	*43 – 106*	*0 – 23*	*8 – 32*	*2.3 – 24.9*	*15 – 432*	*1 – 3*	*1*	*1 – 4*	*1*
B	B1a	3 (3.4%)	22 ≤ x ≤ 33	6 – 24	1 – 3	0.3 – 1.4	9 – 51	2 – 3	1 – 3	1 – 2	3 – 5
	B1b	9 (10.3%)	29 ≤ x ≤ 62	2 – 21	1 – 13	1.6 – 11.2	17 – 114	2 – 3	1 – 2	1 – 5	0 – 2
		Subtotal B1	*22 – 62*	*2 – 24*	*1 – 13*	*0.3 – 11.2*	*9 – 114*	*2 – 3*	*1 – 3*	*1 – 5*	*0 – 5*
	B2	16 (18.4%)	16 ≤ x ≤ 43	1 – 6	1 – 10	0.5 – 6.9	13 – 283	3	1 – 2	1 – 2	0 – 1
Subtotal B		***30 esp (34.5%)***	*16 – 62*	*1 – 24*	*1 – 13*	*0.3 – 11.2*	*9 – 283*	*2 – 3*	*1 – 3*	*1 – 5*	*0 – 5*
C	Ca	7 (8.0%)	16 ≤ x ≤ 36	0 – 5	0 – 7	0.1 – 2.7	4 – 187	1 – 2	1	2 – 4	0 – 1
	Cb	28 (32.2%)	20 ≤ x ≤ 52	0 – 8	0 – 8	0.1 – 6.9	21 – 4090	1 – 2	1	1 – 2	0 – 1
Subtotal C		***35 esp (40.2%)***	*16 – 52*	*0 – 8*	*0 – 8*	*0.1 – 6.9*	*4 – 409*	*1 – 2*	*1*	*1 - 4*	*0 – 1*
Total		**87 esp**	16 – 106	0 – 24	0 – 32	0.1 – 24.9	4 – 432	1 – 3	1 – 3	1 – 5	0 – 5

b) Non-hierarchical clustering: K-means.

i) 3 clusters (K_3)

The results for 3 (K_3) clusters are similar to the partition of species resulting from Ward's method in terms of the species within a cluster:

- **$K_3$1** (38 esp) corresponded with species in Ward's cluster **B1a**, **B2**, species from **Ab** that have intermediate trade variable values, as well as the bulb, root and rhizome species of **Cb**
- **$K_3$2** (28 esp) mainly corresponded with species in **C**.
- **$K_3$3** (21 esp) corresponded with species in **Aa**, **Ab** (species with high trade values) and **B1b**

From the Euclidean distances between the clusters, $K_3$1 and $K_3$2 were closer together than $K_3$2 is to $K_3$3 (Euclidean distance = 0.142 and 0.227 respectively) (Appendix 8 Table j). Given the attributes of the species in the clusters, $K_3$3 and $K_3$1 could be considered higher conservation priority than $K_3$2.

Figure 12a and Table 18 show the standardized and transformed actual mean values per variable per cluster respectively. Species in $K_3$3 (similar to Ward's A and B1b) had higher means for trade variables and lower means for biological variables as opposed to species in $K_3$1 (similar to Ward's B1a and B2) where the opposite was true (Table 18, Figure 12a). Species in $K_3$2 had greater *phylogenetic distinctness* (i.e. they tended to be distinct species in small genera or from a monotypic genus). Species in $K_3$3 were sold by more traders, were cited more often as scarce and/or popular, and tended to be found in larger quantities in the markets.

It is clear that species in $K_3$3 and $K_3$1 have higher vulnerabilities to harvesting for different sets of reasons. In addition, many species in $K_3$1 are known by other regional medicinal plant trade studies to be highly exploited and hence have become highly threatened by this activity (e.g. *S. aethiopicus* and *S. eriopus*). As a result, their prevalence and quantity in the markets are reduced. Therefore, species in $K_3$3 and $K_3$1 have similarly higher conservation priorities compared to species in $K_3$2.

The distinctive nature of the clusters is evident in Figure 12a. Judging from the magnitude (and significance levels) of the *F* values (Table 18) and the distance between the mean value on the graph, *popularity*, *no. of bags, no. of traders*, *habitat specificity* and *geographic distribution* variables were the major criteria for assigning the species to clusters. *Phylogenetic distinctness*, *endemism* and *QGrids* had the least influence. Six of the ten variables used were significant for assigning the species to clusters.

Table 18: Mean values per cluster (transformed from standardized values in Figure 12a to mean actual values) corresponding with 3 clusters. Shaded values are the highest in the row. [See Page 14 for biological variable codes] [df = 2, 84]

	$K_3$1	$K_3$2	$K_3$3	F	P
Phylogenetic distinctness	1.7	1.9	1.7	0.4	0.68
Habitat spec.	2.6	1.5	2.0	28.4	<0.0001
Endemism	1.1	1.0	1.0	1.2	0.32
Red List	0.7	0	0.2	4.5	0.0140
QGrids*	32	57	51	1.6	0.21
Plant part	4.6	3.4	3.6	17.7	<0.0001
No. traders sell	33	32	64	6.4	<0.0001
No. scarce	4	2	7	8.2	0.0006
No. popular	5	3	18	10.2	<0.0001
No. bags	2.7	2.0	11.6	77.4	<0.0001

* Species with smaller QGrid values have higher conservation priorities

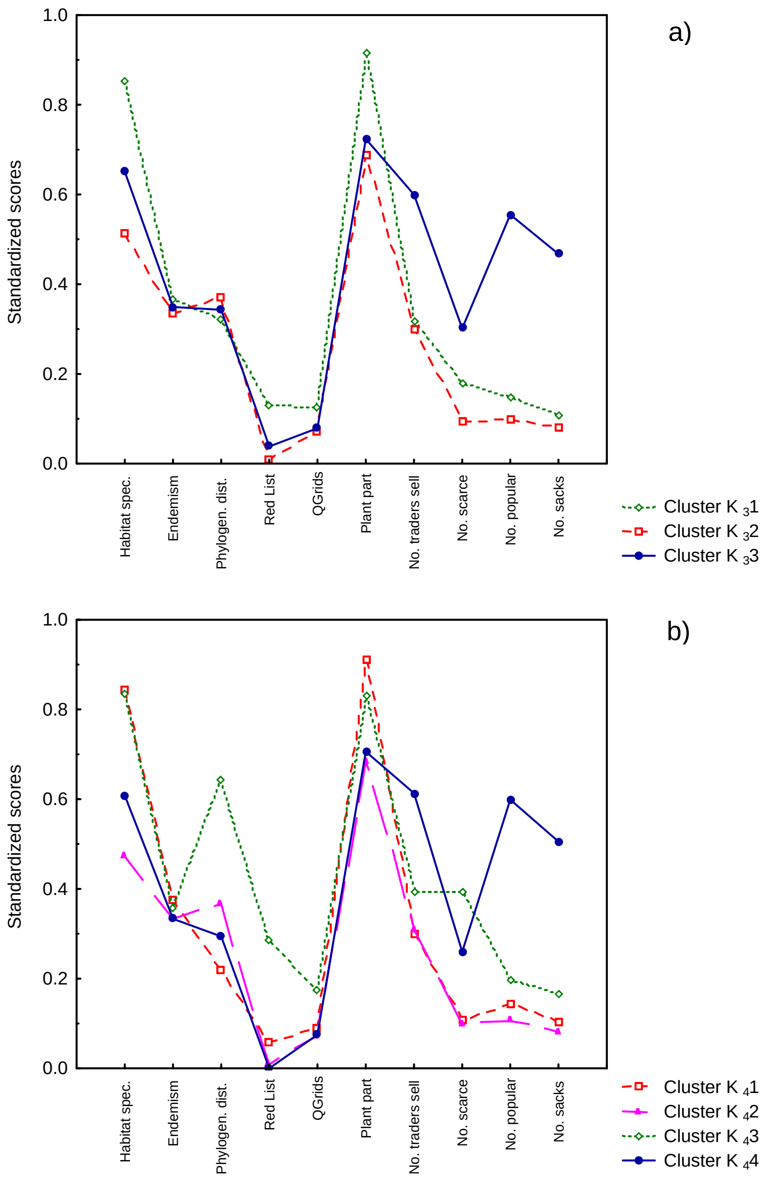


Figure 12: Plot of standardised cluster means for each K-mean cluster. Results for a) 3 and b) 4 clusters for the 87 short-listed ethnospecies. See Tables 18 and 19 for the transformed standardised mean scores for a) and b).

ii) 4 clusters (K_4)

Increasing K-means to four clusters resulted in $K_3$1 being split into two smaller clusters ($K_4$1 and $K_4$3) based mainly on biological variables. The results for 4 (K_4) clusters were also similar to the partition of species resulting from Ward's method in terms of the species within a cluster (see also Figure 11):

- **$K_4$1** (32 esp) corresponded with species in Ward's cluster **B2** and the bulb, root and rhizome species of **Cb**
- **$K_4$2** (24 esp) corresponded with species in **Ca** and the remainder of **Cb** not in $K_4$1.
- **$K_4$3** (14 esp) corresponded with species in **B1**.
- **$K_4$4** (17 esp) corresponded with species in **A**, except for four species which were clustered in **$K_4$1**.

Two groups of species emerged as higher priority (Table 19). First, **$K_4$3** because of greater phylogenetic distinctness, more Red List species, smaller geographic distribution and relatively more species cited as scarce. Species in the second group, **$K_4$4**, are also of a higher conservation priority because the species ranked first in terms of the values of the trade variables (Table 19). Species in **$K_4$1** are of medium conservation priority because they rank high in terms of habitat specificity, endemism and plant part (i.e. there are more whole plants and roots within this group); however, they tended to be ranked third in terms of trade variables. $K_4$2 is of lower conservation priority as the values for the variables are lower than the other clusters.

Table 19: Mean values per cluster (transformed from standardized values in graphs to mean actual values) corresponding with the 4 clusters in Figure 12b. [See Page 14 for biological variable codes] Yellow-shaded values are the highest in the row, and blue-shaded variables are the second highest. The species corresponding to each cluster are also listed below. The species order in a column is irrelevant and corresponds only with the order in the Figure 11 dendrogram. The coloured boxes correspond with the narrow shaded areas in Figure 11. [df = 3, 83]

	$K_4$1	**$K_4$2**	**$K_4$3**	**$K_4$4**	F	P
Habitat specificity	2.53	1.4	2.50	1.8	26.3	<0.001
Endemism	1.13	1.0	1.07	1.0	1.2	0.32
Plant part	4.6	3.4	4.1	3.5	10.0	<0.001
Phylogenetic distinctness	1.1	1.8	3.2	1.5	18.2	<0.001
Red List	0.3	0.0	1.4	-	11.3	<0.001
QGrids	44	55	23	53	2.1	0.11
No. scarce	3	2	9	6	13.9	<0.001
No. traders sell	32	33	42	65	36.1	<0.001
No. popular	5	3	6	19	66.9	<0.001
No. sacks	2.5	2.0	4.2	12.6	51.9	<0.001

$K_4$1	**$K_4$2**	**$K_4$3**	**$K_4$4**
Medium conservation priority	Lower conservation priority	Higher conservation priority	Higher conservation priority
Talinum caffrum	Rhoicissus tridentata	Siphonochilus aethiopicus*	Drimia spp.
Ornithogalum longibracteatum	Synaptolepis kirkii	Cassipourea spp.*	Acacia xanthophloea
Thesium pallidum	Ansellia africana	Stangeria eriopus*	Elaeodendron transvaalense
Dombeya rotundifolia	Hippobromus pauciflorus	Bowiea volubilis*	Albizia adianthifolia
Euphorbia woodii*	Pappea capensis	Clivia spp. *	Rapanea melanophloeos
Eucomis bicolor	Bersama spp.	Eucomis autumnalis	Trichilia spp.
Dietes spp.	Bridelia cathartica	Ocotea bullata	Adenia gummifera
Stapelia gigantea	Elephantorrhiza elephantina	Warburgia salutaris	Helichrysum spp.
Eulophia speciosa	Pentanisia prunelloides	Gunnera perpensa	Maytenus undata
Dianthus mooiensis	Balanites maughamii	Eriospermum mackenii	Schotia brachypetala
Cephalaria humilis	Croton spp.	Myrothamnus flabellifolia	Sclerocarya birrea
Anemone spp.	Calodendrum capense	Callilepis laureola	Curtisia dentata
Alepidea amatymbica	Zanthoxylum spp.	Ekebergia capensis	Hydnora africana
Secamone gerrardii	Plectranthus hadiensis	Schlechterina mitostemmatoides	Hypoxis spp.
Raphionacme hirsuta	Rubia cordifolia		Merwilla plumbea
Eriosema spp.	Agapanthus spp.		Urginea spp.
Pelargonium luridum	Pleurostylia capensis		Dioscorea sylvatica
Gnidia kraussiana	Cryptocarya spp.		
Aster bakerianus	Vitex spp.		
Senecio coronatus	Diospyros galpinii		
Silene bellidioides	Acalypha villicaulis		
Boophone disticha	Capparis spp.		
Gnidia spp.	Vernonia tigna		
Aristea spp.			
Andrachne ovalis			
Dioscorea dregeana			
Tulbaghia spp.			
Agathosma ovata			
Crinum spp.			
Haworthia limifolia /			
Aloe aristata			
Polygala spp.			
Scabiosa columbaria			

* Species currently classified as 'threatened' by IUCN Red Lists

From the magnitude and significance levels of the *F* values, eight of the ten variables were significant for assigning species to clusters, especially *popularity*, *no. of bags,* and *no. of traders* (Table 19). The only non-significant variables were *QGrids* and *endemism*.

iii) ≥5 clusters

Increasing K-means to five clusters resulted in no substantial improvement to the clustering results and there appeared to be no value in increasing K-means beyond 4 clusters, hence 4 clusters were retained. Six of the ten variables were significant in assigning species to the clusters.

c) Implications of the results

Cluster analysis provided a relatively simple and objective way to explore the risk and conservation priorities of selected species by systematically reducing the list of 392 ethnospecies found to be sold in the region to a core group of 87 higher-risk species. Species not on the final short-list are not considered to be as threatened by commercial harvesting for the medicinal plant trade as the 87 species are. The excluded species may, however, be at risk of extinction from other deterministic and stochastic risk factors. At the level of 87 species, it was feasible to add more variables to the analysis to elucidate which species are of greater conservation priority within the higher-risk group. The Ward's and K-means cluster analyses optimally divided the species into 4 main synonymous clusters, with a few variations in the placement of species.

Species in A/$K_4$4 are highly traded and are threatened with quasi-extinction if current harvesting trends continue. Some of their counter-parts in B1/$K_4$3 are already known to be threatened with extinction, have restricted distributions and greater *Phylogenetic distinctness*. Thus, species in A/$K_4$4 and B1/$K_4$3 have high conservation priorities with respect to their presence in the medicinal plant trade in the Witwatersrand compared to all other species. Their status can therefore be described as potentially or currently declining, and would warrant consideration on the SANBI's Orange List (e.g. Victor & Keith) early warning system (if they aren't already Red Listed). Many of these species are known to be unsustainably harvested.

Species in clusters B2 and $K_4$1 can be described as medium conservation priority. Factors such as their higher level of endemism and habitat specificity (in $\leq$2 habitat types) make them a broadly national conservation priority, but the level of trade is not as extensive as A/$K_4$4 and B1/$K_4$3 and therefore priority levels based on harvesting for the medicinal plant trade for these species are lower. There are, however, a few species in this group that could be shifted to a higher conservation priority group and/or monitored, for example *E. woodii* (already classified as Vulnerable according to the Red List) and *A. amatymbica*. Populations of *A. amatymbica* are known to be declining as the habitat of the species is reduced (Diederichs 2005). Furthermore, the plants are being unsustainably harvested.

Species in clusters C and $K_4$2 are lower conservation priority compared to the rest of the top 87 species. However, their presence in the top 87 species already places them in a conservation priority category higher than the species that were excluded from the list.

10.5.8 Plant risk analysis: Principle Components Analysis (PCA) of the Variables

The emphasis of the ordination was to assess the direction and strength of the correlations between the variables and species used in the cluster analysis, and hence the relative importance of the variables in evaluating risk and determining conservation priority. Results from the PCA[2] showed which species were positively correlated with which variables and, that not all the variables were equally important in categorising the risks or threats to species.

As described in the methods, species were given standardized scores for each variable – the same scores were used in the cluster analyses to group the species into risk hierarchies. For the PCA, the variable scores were totalled and the species sorted in descending order of the total score so as to rank the species from 1 (species with the highest total score) to 392 (species with the lowest total score). These numbers appear on the resultant biplots. In addition, species at the edge of a biplot (furthest from the origin) have the highest risks and/or conservation priorities compared to species nearer the origin. Similarly, the longest arrows indicate variables of most importance in the analyses.

a) Ordination of 392 ethnospecies

Ordination of the data for 392 ethnospecies using all the variables in the PCA explained 76.3% of the variance with the first two axes (Table 20i). In the biplot of the first and second axes, the distinctive influence of the *plant part* variable in grouping the species into plant part types is very evident (Figure 13a). From the ordination, it is also apparent that the most important criteria are prevalence in the shops and market, and the quantity present (*no. of bags*). Scarcity of species in the shops is less strongly associated with the position of species in the biplot. Overall, species with the higher ranks (i.e. have higher scores) are found towards the edge of the biplot, are thus more strongly correlated with the variables and are at greater risk to over-harvesting, especially *Drimia* spp., *E. transvaalense, E. autumnalis* and *M. plumbea* (no's 1-4 respectively, Figure 13a).

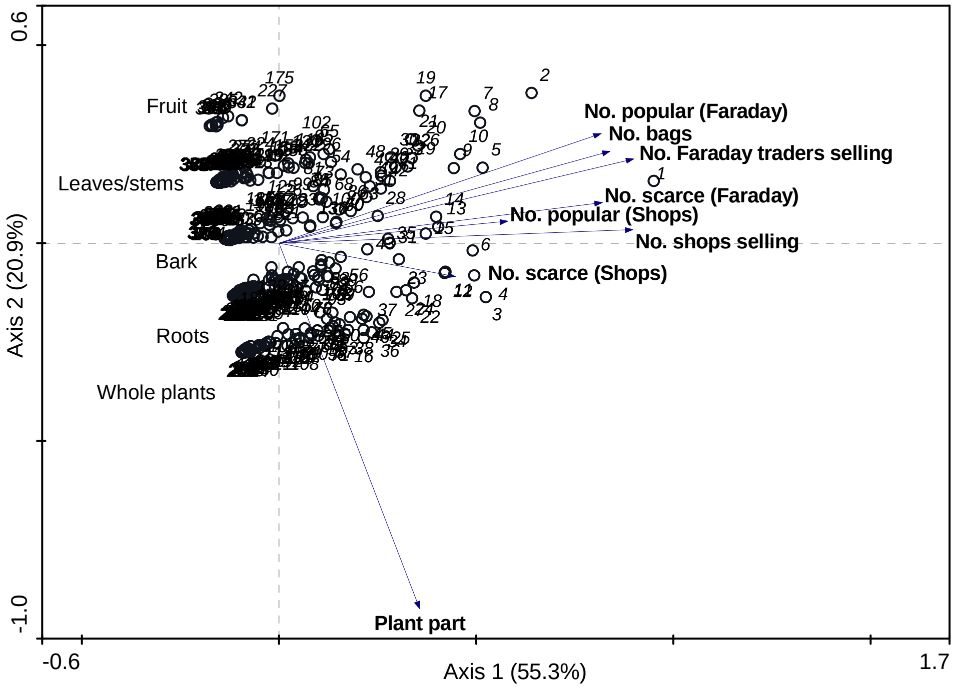
Table 20: Axis eigenvalues for the PCA ordination of 392 ethnospecies sold by 150 traders (Shop and Faraday data combined).

	(i)		(ii)		(iii)	
392 ethnospecies	• Corresponds with biplot Figure 13a • 8 variables		• Corresponds with biplot Figure 13b • 7 variables (excluding *plant part*)		• Corresponds with biplot Fig 14 • 6 variables (excluding *scarcity* for shops and Faraday)	
Axes	Eigenvalue	Cumulative variance	Eigenvalue	Cumulative variance	Eigenvalue	Cumulative variance
1	0.553	55.3%	0.689	68.9%	0.567	56.7%
2	0.209	76.3%	0.139	82.8%	0.248	81.6%
3	0.105	86.7%	0.073	90.2%	0.114	93.0%
4	0.049	92.4%	0.049	95.1%	0.057	98.6%

When *plant part* is removed as a variable from the ordination, the species do not cluster into distinct groups on the biplot (Figure 13b). Furthermore, some species are seen to be more positively correlated with quantitative indicators of threat from the shop survey (e.g. no's 3,4 and 17) than for the Faraday survey. Ordination of the data for 392 ethnospecies using PCA and *excluding* plant part as a variable explained 82.8% of the variance with the first two axes (Table 20ii).

[2] Given the qualitative and presence-absence nature of some of the data, Principle Coordinates Analysis (PCoA) was run as a variant of the PCA analysis. The results are, however, described as PCAs.

a)



b)

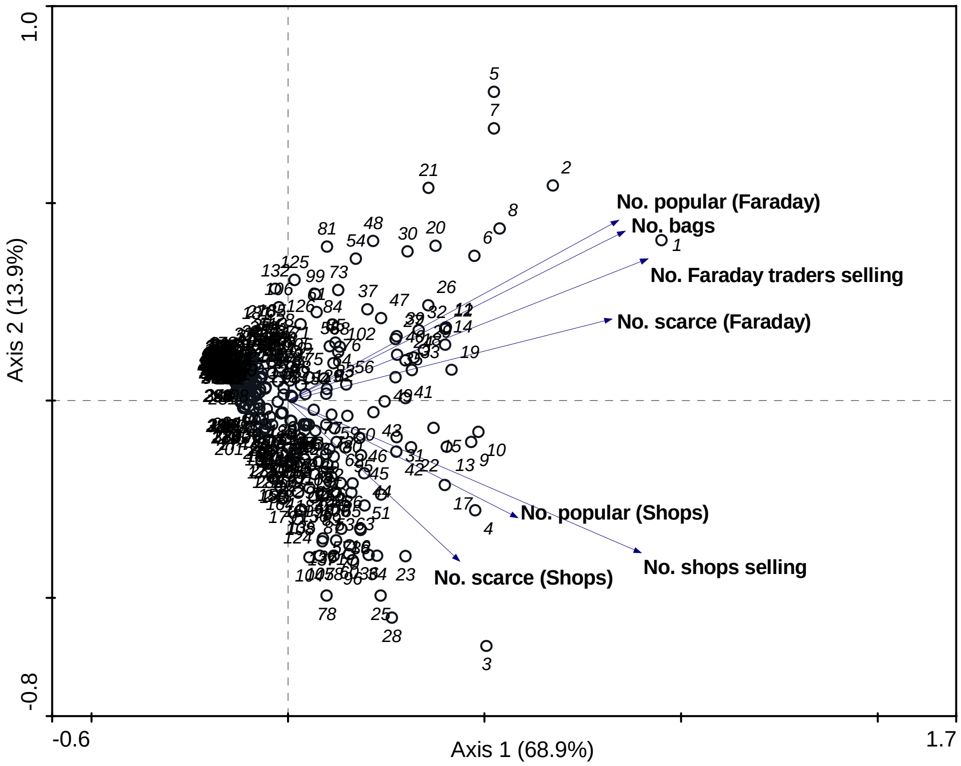


Figure 13: PCA ordination biplots for 392 ethnospecies sold by 150 traders. Figure a) is ordinated using all the variables, while b) excludes *plant part* as a variable. Axes 1 and 2 cumulatively account for 76.3% and 82.8% of the variance in a) and b) respectively. Species numbers are listed in Appendix 6.

Throughout the K-means analyses, *scarcity* was shown to have had the least influence in assigning species to clusters, especially in the very large data sets of >300 ethnospecies. In the ordination results above, *scarcity* accounts for less of the variance than some of the other variables (indicated by a shorter arrow), and is therefore not a principle component in this risk analysis process. To establish whether *scarcity* was a redundant variable in risk analyses that only incorporated trade factors, the ordinations were repeated without it. The results showed that the percentage variance of the first axis only changed by 1.4% (Table 20iii), and the positions of the species altered very little (Figure 14).

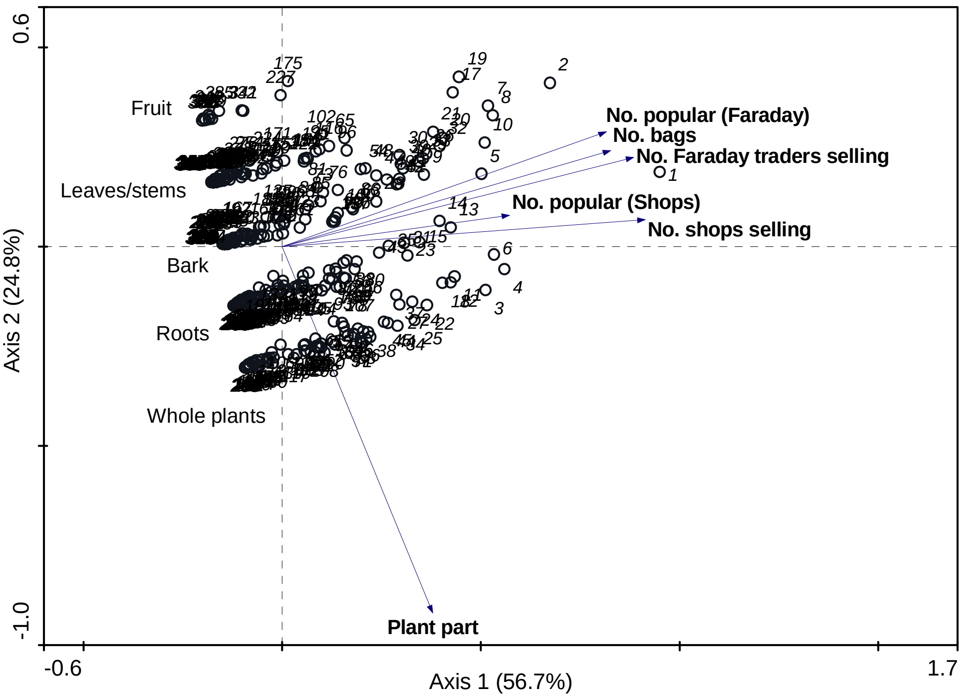


Figure 14: PCA ordination biplot for 392 species sold by 150 traders (in 50 *muti* shops and by 100 Faraday traders). This ordination excludes *scarcity* as a variable. Axes 1 and 2 cumulatively account for 81.6% of the variance. Species numbers listed in Appendix 6.

That *scarcity* is a variable weakly correlated with risk is also known from the earlier regressions in 10.5.1 (Figures 2a & 2b) and 10.5.2 that concerned the value of a citation and trader perceptions. However, it is the only variable in the initial stages of the analysis that indicates potentially over-harvested species. Some species like *S. aethiopicus, W. salutaris* and *O. bullata*, for example, would not feature in the higher-risk categories of the cluster analyses were it not for the value this variable added to the scores for species. Hence, excluding *Scarcity* as a variable would generate a risk hierarchy based only on species that are very popular and available to the traders.

The distance of a species from the centre of the biplot is indicative of the level of risk in Figure 14. Increasing distance from the origin implies increased risks of quasi-extinction. In descending order, the species furthest away from the origin of Figure 13a and 14 are: *Drimia* spp. (1), *E. transvaalense* (2), *A. adianthifolia* (7), *A. xanthophloea* (8), *A. gummifera* (19), *Helichrysum* spp. (17), *R. melanophloeos* (10), *M. plumbea* (4), *H. africana* (5), *Hypoxis* spp. (6) and *E. autumnalis* (3).

b) Ordination of 87 species

The next stage of the analysis was to ordinate the 87 short-listed species with the 10 variables that were selected. The variables, a combination of trade and biological factors, are inclusive of the factors that indicate both levels of threat and conservation priority. The same variables were used to cluster the species previously in Section 10.5.7.

Ordination of the data for 87 ethnospecies using PCA and all 10 variables explained 51.4% of the variance with the first two axes (Table 21 i). The biplot of the first and second axes shows there were 3 groups of variables that are positively correlated with the species. First, were the biological variables (*habitat specificity*, *RL, phylogenetic distinctness,* etc.). Species that are highly positively correlated with these variables include *S. aethiopicus*, *S. eriopus* and *E. woodii* (no's 3, 6 and 7 respectively, Figure 15a). Second, were three of the trade variables (*no. of bags*, *popularity* and the *no. of traders* selling the species). Species that are strongly correlated with these variables include *Drimia* spp., *E. transvaalense*, *A. adianthifolia*, *A. xanthophloea* and *Hypoxis* spp (no's 1, 13, 24, 14 and 10 respectively, Figure 15a). Third was the *Scarcity* variable that lay midway between the other two variable groups. Species strongly correlated with this variable include *H. africana*, *E. autumnalis* and *Clivia* spp. (no's 2, 4 and 5 respectively, Figure 15a). Thus in smaller data sets, *scarcity* appears to have a more significant role in assigning species to clusters.

Table 21: Axis eigenvalues for the PCA ordination of 87 short-listed ethnospecies sold by 150 traders (shop and Faraday data combined).

	(i)		(ii)		(iii)	
87 ethnospecies	• Corresponds with biplot Figure 15a • 10 variables		• Biplot not shown • 9 variables (excluding *endemism*)		• Corresponds with biplot Figure 15b • 8 variables (excluding *endemism* and *phylogenetic distinctness*)	
Axes	Eigenvalue	Cumulative variance	Eigenvalue	Cumulative variance	Eigenvalue	Cumulative variance
1	0.324	32.4%	0.331	33.1%	0.390	39.0%
2	0.190	51.4%	0.193	52.4%	0.224	61.3%
3	0.151	66.4%	0.154	67.9%	0.152	76.5%
4	0.123	78.7%	0.126	80.4%	0.104	86.9%

The ordinations also showed that not all the variables used to score the species were equally important. For this data set (i), *endemism, QGrids* and *phylogenetic distinctness* had less weight in determining the conservation priorities of species. This corresponded with the results of the K-means (K_3 and K_4) analyses, where these variables were flagged as least important for assigning species to clusters.

The data set was reanalysed without the variables that were not strongly correlated in the initial 10 variable ordination. First, *endemism* was omitted because it showed the least variance in the ordination and the weakest correlation with the species (Table 21 ii). Results showed there to be very little change in the cumulative percentage variance accounted for by the axes compared with the full 10 variable ordination (Table 21ii). Second, *endemism* and *QGrids* were omitted (the latter also showing a weak correlation in Figure 15a). The optimal linear combination of variables appeared to be where *endemism* and *phylogenetic distinctness* were omitted (Table 21 iii, Figure 15b). The remaining variables in this biplot were thus the principle components for assessing the threats and/or conservation priorities of species for *this data set*.

Increased risk and conservation priority is implied for species furthest away from the biplot origin (Figure 15b). In descending order, the species furthest away from the origin of Figure 15b are: *Drimia* spp (1), *O. bullata* (24), *E. transvaalense* (13), *S. aethiopicus* (3), *E. autumnalis* (4), *H. africana* (2), *A. xanthophloea* (14) and *E. woodii* (7). Some of these species already have a high national conservation priority in terms of their known threats and vulnerabilities.

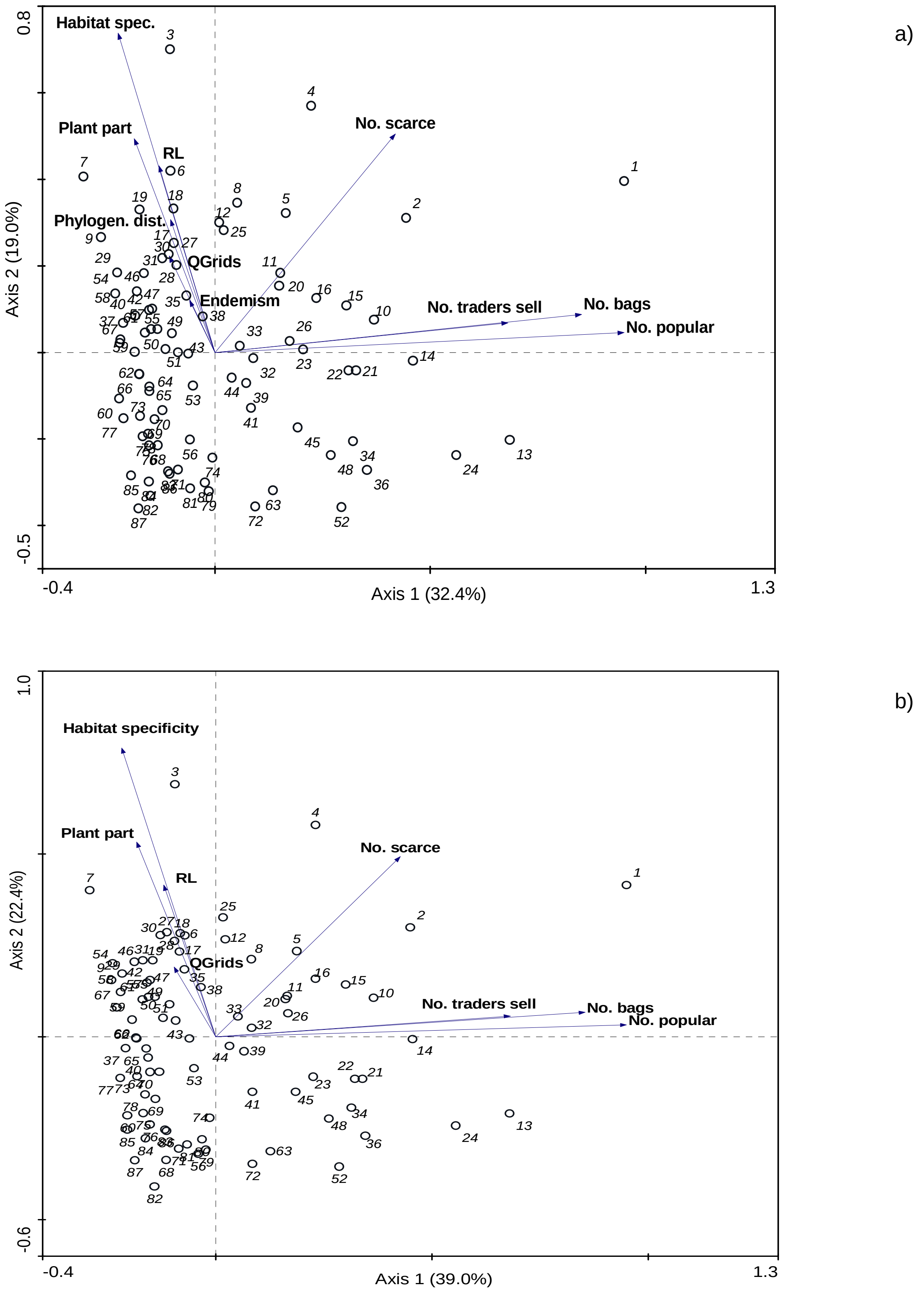


Figure 15: PCA ordination biplot for 87 short-listed species. Figure a) includes all variables, whereas b) excludes variables *endemism* and *phylogenetic distinctness*. Axes 1 and 2 cumulatively account for 51.4% and 61.3% of the variance in a) and b) respectively. The species numbers are listed in Table 22.

Table 22: Reference to species numbers in the PCA biplot for 87 species (Figures 15a & 15b). The species are listed in descending numerical order of their total scores for the variables used in the analyses.

Number	Species	Number	Species
1	Drimia spp. (altissima, elata, sanguinea)	45	Maytenus undata
2	Hydnora africana	46	Senecio coronatus
3	Siphonochilus aethiopicus	47	Secamone gerrardii
4	Eucomis autumnalis	48	Helichrysum spp. (8spp.)
5	Clivia spp. (caulescens, gardenii, miniata, nobilis)	49	Dioscorea dregeana
6	Stangeria eriopus	50	Agathosma ovata
7	Euphorbia woodii	51	Tulbaghia spp. (ludwigiana, natalensis)
8	Gunnera perpensa	52	Trichilia spp. (dregeana, emetica)
9	Schlechterina mitostemmatoides	53	Elephantorrhiza elephantina
10	Hypoxis spp. (colchicifolia, hemerocallidea)	54	Silene bellidioides
11	Warburgia salutaris	55	Gnidia kraussiana
12	Callilepis laureola	56	Hippobromus pauciflorus
13	Elaeodendron transvaalense	57	Aster bakerianus
14	Acacia xanthophloea	58	Anemone spp. (caffra, fanninii)
15	Merwilla plumbea	59	Crinum spp. (delagoense, macowanii, moorei)
16	Urginea spp. (delagoensis, epigea)	60	Ansellia africana
17	Eucomis bicolor	61	Pelargonium luridum
18	Bowiea volubilis	62	Haworthia limifolia/Aloe aristata
19	Eriospermum mackenii	63	Balanites maughamii
20	Ekebergia capensis	64	Pentanisia prunelloides
21	Schotia brachypetala	65	Scabiosa columbaria
22	Sclerocarya birrea spp	66	Polygala spp. (confusa, serpentaria)
23	Curtisia dentata	67	Eriosema sp. (mackenii, salignum)
24	Albizia adianthifolia	68	Pappea capensis
25	Ocotea bullata	69	Pleurostylia capensis
26	Dioscorea sylvatica	70	Gnidia spp. (cuneata, kraussiana)
27	Stapelia gigantea	71	Bridelia cathartica
28	Dianthus mooiensis	72	Croton spp. (gratissimus, moorei)
29	Dietes spp (iridoides, butcheriana)	73	Aristea spp.(ecklonii, woodii)
30	Eulophia speciosa	74	Acalypha villicaulis
31	Cassipourea spp. (flanaganii, gerrardii, malosana)	75	Cryptocarya spp.(latifolia, myrtifolia, woodii)
32	Ornithogalum longibracteatum	76	Diospyros galpinii
33	Talinum caffrum	77	Andrachne ovalis
34	Adenia gummifera var. gummifera	78	Vitex spp. (rehmanii, wilmsii)
35	Alepidea amatymbica var. amatymbica	79	Calodendrum capense
36	Rapanea melanophloeos	80	Capparis spp. (brassii, tomentosa)
37	Synaptolepis kirkii	81	Zanthoxylum spp. (capense, davyi)
38	Raphionacme hirsuta	82	Bersama spp. (lucens, staynerii, swinnyi, tysoniana)
39	Dombeya rotundifolia	83	Plectranthus hadiensis
40	Myrothamnus flabellifolia	84	Rubia cordifolia
41	Rhoicissus tridentata	85	Agapanthus spp. (africanus, campanulatus, caulescens, praecox)
42	Cephalaria humilis	86	Vernonia tigna
43	Boophone disticha	87	Asparagus spp. (asparagoides, laricinus, plumosus, ramosissimus, virgatus)
44	Thesium pallidum		

As shown throughout the previous cluster analyses for the larger data sets, *plant part* is an important variable for establishing the vulnerability of a species to over-harvesting. As the data sets get smaller, however, *plant part* plays less of an important role in assigning species to clusters. Conversely, *scarcity* was less important in large data sets and more important for the short-listed species. *Endemism* is also a largely redundant variable, possibly because there is not a high level of endemism amongst species primarily found in southern African grasslands and savanna. Were this data set to have been ordinated for species originating from the Cape Floral Kingdom where the level of endemism is higher, for example, the variable might have been less redundant in the analyses.

10.5.9: Plant risk analysis: 22 Test species

Time and resource constraints will limit the number of quantitative data that can be collected. Hence, extensive surveys like that of the *muti* shops and Faraday market in this study might not always be feasible. A question that therefore arises is: how different are the risk hierarchy and conservation priority cluster results for large data sets from the clusters constructed with smaller numbers of species and analysed with the same data? Also, what combination of variables is most strongly correlated with risk? Ideally, the clusters should be analogous, i.e. the same species should appear in similar clusters, thereby validating the use of cluster analysis to predict extinction risks and conservation priorities.

In this section of the results, the 22 test species are analysed using the same variables that were used to cluster the list of 392 ethnospecies (trade variables) and the 87 short-listed species (trade and biological variables). In the last part of the analysis (section 10.5.9 c), additional trade variables that can usually only be collected for smaller data sets, were investigated to see if they added value to the species assessments (see Appendix 7 for the species-attribute matrix). The 22 test species had been previously selected to reflect a range of risk profiles from high to low.

a) Comparison with the 392 species risk hierarchy

The results presented here were compared with the results of 'All' (data set III), i.e. the amalgamation of shop and street trader data (Section 10.5.6; Appendix 3), using the same combination of eight variables. From the preliminary results of the K-means analysis, however, it was evident that the *plant part* variable was not significant in allocating species to clusters (p=0.79). This corroborates earlier evidence that plant part plays an increasingly insignificant role in assigning species to clusters as the number of species in a data set decreases. *Plant part* was subsequently omitted as a variable from the remainder of the analyses, except where specified.

Using Ward's method, three clusters were identified at ≈72% relative distance, namely A, B and C (Figure 16). When compared with hierarchical clustering of species in the similarly analysed dendrograms of Figure 9 (Section 10.5.6) and Appendix 3, the species are seen to be similarly clustered. Species in clusters A and C (Figure 16) correspond with species in Y and X sub-clusters respectively in Figure 9. Species in B (Figure 16) mostly correspond with species in B (Figure 9), although *Diospyros galpinii* was originally clustered in D (Figure 9). The risk characterisations of the species are similar to those of Figure 9, however given the size of the 22 species data set, clusters C, A and B can be characterised as higher, intermediate and lower risk respectively.

K-means clustering partitioned the species similarly to Figure 16. For 3 clusters, **$K_3$3** is the same as **C**, **$K_3$2** and **$K_3$1** are similar to **A** and **B** respectively – the only difference being that *S. aethiopicus* is positioned in the lower rather than intermediate risk group in $K_3$1. The Faraday variables were primarily responsible for assigning the species to the clusters; and the popularity of species in the shops was also significant. Least significant were the scarcity variables. The means for the variables in the three clusters, and the species in each cluster, are shown in Table 23. Forming four clusters did not change the cluster compositions except for *S. aethiopicus* and *E. autumnalis* being assigned their own cluster. A three cluster partition was therefore accepted for this analysis.

Overall, the 22 test species were positioned similarly in the risk hierarchies using the same 7-8 variables as they were in the matrix of 392 species.

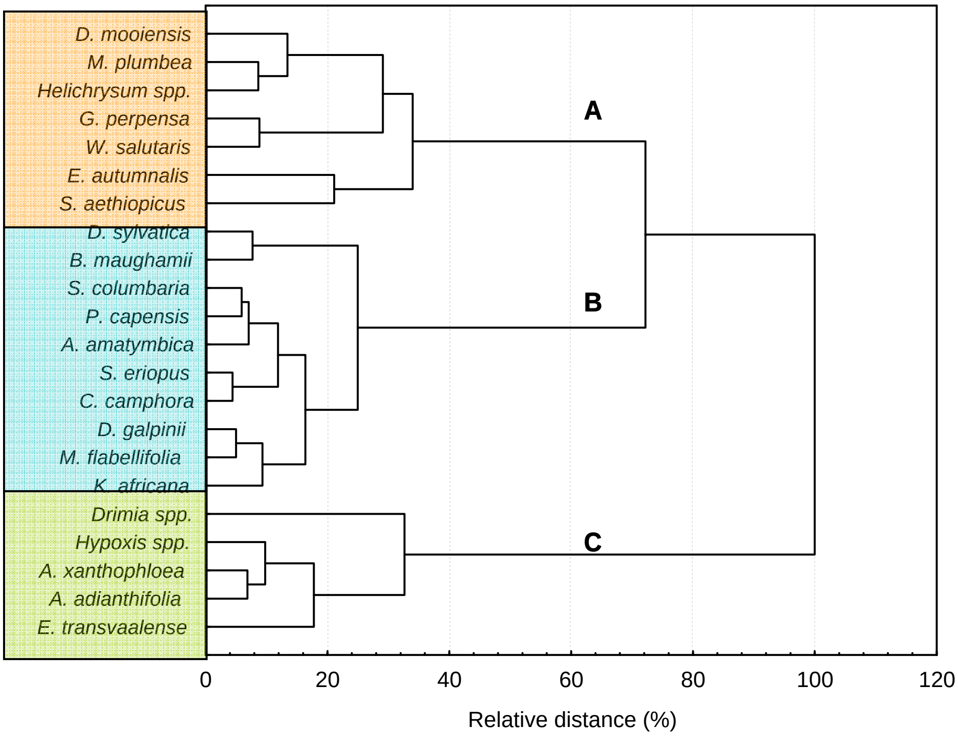


Figure 16: Dendrogram based on Euclidean distances and Ward's agglomerative clustering method for 22 test species using the same variables tested with 392 ethnospecies (compare with Figure 9 and Appendix 3). [df = 3,18]

Table 23: Mean values per cluster corresponding with 3 K-means clusters for the 22 test species.

	$K_3 1$	$K_3 2$	$K_3 3$	F	P
Mean no. bags	2.3	6.9	17.7	20.3	<0.001
Mean no. Faraday traders sell	10	19	46	21.1	<0.001
No. scarce (Faraday)	2	5	7	7.5	0.002
Mean no. popular (Faraday)	4	7	23	22.6	<0.001
Mean no. shop traders sell	26	35	35	1.8	0.175
Mean no. scarce (shops)	5	4	3	9.7	<0.001
Mean no. popular (shops)	1	6	2	9.0	<0.001
<u>Species</u>	*A. amatymbica* *B. maughamii* *C. camphora* *D. galpinii* *D. sylvatica* *K. africana* *M. flabellifolia* *P. capensis* *S. aethiopicus* *S. columbaria* *S. eriopus*	*D. mooiensis* *E. autumnalis* *G. perpensa* *Helichrysum* spp. *M. plumbea* *W. salutaris*	*A. xanthophloea* *A. adianthifolia* *Drimia* spp. *E. transvaalense* *Hypoxis* spp.		

b) Comparison with the 87 ethnospecies conservation priority clusters

The results presented here were compared with the results of cluster analyses for the 87 short-listed species (Section 10.5.7; Figure 11), using the same combination of 10 variables. In this sequence, however, *endemism* was omitted because all 22 species had the same score for the variable. The analyses were also run with and without *plant part* as a variable.

With *Plant part* as a variable, 5 clusters were identified at >40% relative distance (graph not shown). Without *Plant part*, four clusters were identified at >40% relative distance (Figure 17). The only difference between the two dendrograms that included and excluded *plant part* was that *Drimia* was in its own cluster in the former. When the *plant part* was excluded, however, the species were clustered in a similar way to the 87 ethnospecies in the dendrogram of Figure 11. Furthermore, *Drimia* was assigned to the same group of species as cluster A of Figure 11. Hence, the 22 ethnospecies clustered in a similar manner to Figure 11 when eight of the ten variables were used (Figure 17). The only other difference was that the cluster linkage distances were shorter when fewer species were analysed.

Kigelia africana and *Cinnamomum camphora* (an exotic) were originally not present in the group of 87 species because they were not short-listed as higher risk; however, they were included in the 22 species data set and Ward's agglomerative clustering placed them in a group of similarly low conservation priority species.

The conservation priorities of the 22 species in Figure 17 can be described in the following way:

- **Cluster A:** analogous with clusters B1b and B2 of Figure 11, therefore species have a higher conservation priority
- **Cluster B:** analogous with cluster Cb of Figure 11, therefore species have a lower conservation priority
- **Cluster C:** cluster consists only of *S. aethiopicus* and is analogous with B1a of Figure 11, therefore species has a higher conservation priority
- **Cluster D:** analogous with cluster A of Figure 11, therefore species have a higher conservation priority.

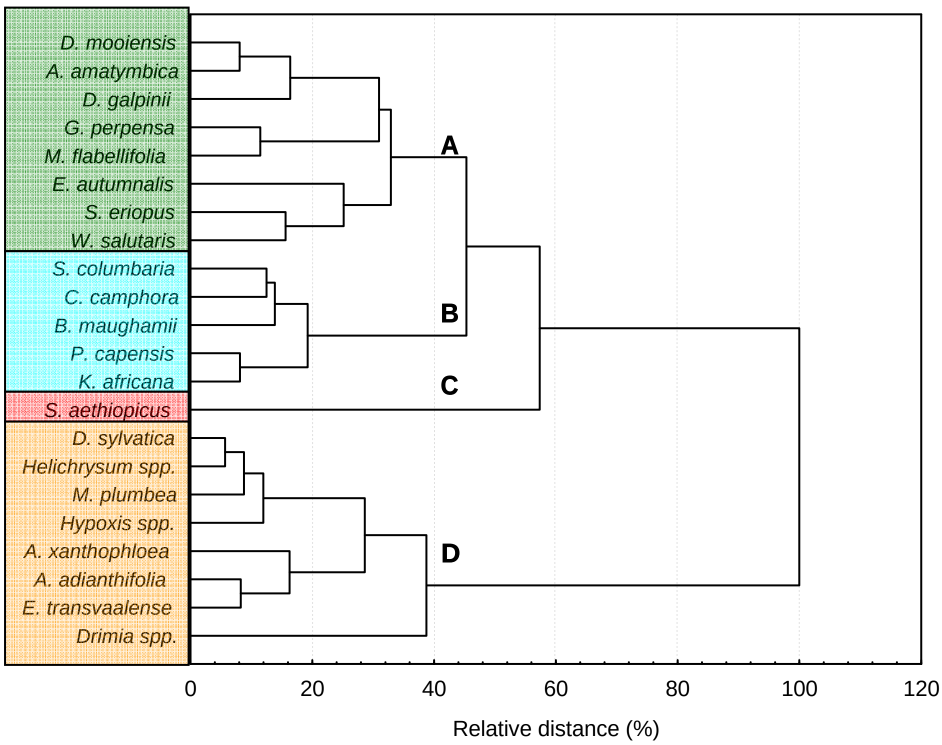


Figure 17: Dendrogram based on Euclidean distances and Ward's agglomerative clustering method for 22 test species using the same variables (but excluding *plant part*) tested with 87 ethnospecies (compare with Figure 11).

When K-means cluster analysis was run using the eight variables, the species assigned to K_4 clusters was exactly the same as the composition of the four clusters in the dendrogram of Figure 17 above. The mean values per variable for assigning species to clusters show that $K_4$1, $K_4$2 and $K_4$4 species have a higher conservation priority than species in $K_4$3 (Table 24). $K_4$1species are a priority because they are more prevalent and popular in the markets. $K_4$2 species are found in ≤2 habitats and have a greater level of phylogenetic distinctness; furthermore, the species usually have the second-highest trade variable values. *Siphonochilus aethiopicus* ($K_4$4) is high priority because the species has already been Red Listed in South Africa as Critically Endangered; it also has a limited geographic distribution and is regarded as the most scarce by traders. Species in $K_4$3, by contrast, are of lower conservation priority in this group of 22 species because the mean variable values are lower than the other three clusters. The *Red List* variable was primarily responsible for assigning species to clusters, followed by volume, popularity and prevalence (Table 24). Least important was *phylogenetic distinctness*.

Table 24: Mean values per variable per cluster for 4 K-means clusters. Shaded values are the highest in the row and/or indicated higher risk or conservation priority. [See Page 14 for biological variable codes] [df = 3, 18]

	$K_4$1	$K_4$2	$K_4$3	$K_4$4	F	P
Habitat Specificity	1.8	2.7	1.3	2.0	9.4	0.001
Phylogenetic distinctness	2.0	3.4	2.7	2.0	2.5	0.089
Red List	-	0.7	-	5.0	42.1	<0.001
QGrids	67	42	60	11	6.4	0.004
No. traders sell	72	45	34	22	12.3	0.001
No. traders scarce	8	9	4	24	3.7	0.031
No. traders popular	21	7	4	3	13.9	0.001
No. bags	14.6	3.9	2.0	0.3	14.3	0.001
Species	*D. sylvatica* *Drimia spp.* *Hypoxis spp.* *M. plumbea* *A. xanthophloea* *A. adianthifolia* *E. transvaalense* *Helichrysum spp.*	*D. mooiensis* *E. autumnalis* *S. eriopus* *A. amatymbica* *G. perpensa* *W. salutaris* *M. flabellifolia*	*S. columbaria* *P. capensis* *B. maughamii* *C. camphora* *D. galpinii* *K. africana*	*S. aethiopicus*		

A broad observation that can be drawn based on the results of the cluster analyses is: that even with fewer species, cluster analysis can assign species to clusters in an analogous or similar order (and reflecting similar conservation priorities) to larger sets of species. One feature with smaller numbers of species, however, is that the scores for some of the variables might be all the same (e.g. *endemism*) and other variables, such a *plant part* and *phylogenetic distinctness*, might be less important in assigning species to clusters – especially if those species have been selected on the basis of those variables at the outset.

c) The use of additional trade variables to assess risk, threat and/or priority

A criterion for selecting variables to assess the risk and conservation priority of species should be the ease and cost of data availability. For large data sets, the variable options are principally limited to counts of the number of traders selling plants and the number of citations of scarcity and popularity. For medium-size data sets, biological variables can be included – but the cost of obtaining distribution information (e.g. *QGrids*) can inhibit the number of species assessed.

Other data that can be considered in a risk/threat/priority assessment process are trade variables such as mean price per unit mass (e.g. *R/kg*), quantity per sale (e.g. *kg/S*) and number of sales per annum (*S/a*). However, an adequately large sample size per species is necessary to reduce the variance of the mean – and this requirement can restrict the number of species investigated. Furthermore, as results from Chapter 9 demonstrated (Williams *et al.* in press), *R/kg* and *kg/S* are better used as indicators of risk/threat for groups of species of specific plant part types (e.g. bark, bulbs or roots), thus limiting the use of these two variables in a cluster analysis for a range of species. Sales in the medicinal plant markets are volume- rather than mass-based. Hence, analysis results using these variables could artificially inflate the threats to species associated with plant parts that are smaller and/or lighter (e.g. leaves, stems and small roots), compared to similar volumes of a species associated with plant parts that are larger and/or heavier (e.g. bulbs and tubers).

Plant price can be an indicator of scarcity. Higher prices are often paid for hard-to-get resources (or, the same price is paid for a smaller quantity); hence, R/kg values are usually higher for scarcer resources. For example, the mean cost of *B. maughamii* bark in Faraday in 1996 was R6.30/kg versus R11.20/kg for *W. salutaris* bark. However, the potential scarceness of a resource using R/kg as a variable is best seen within a plant part group because the unit quantity sold is usually of the same magnitude.

To show how species are positively or negatively correlated with the additional variables in a risk or threat assessment process, the 22 test species were reanalysed using 12 variables and data derived from the 1994/5 *muti* shop survey. PCA results showed how the risks of species like *S. aethiopicus* are strongly correlated with *R/kg*, whereas bulbs like *M. plumbea* and *Drimia* spp. are positively correlated with *kg/S* and prevalence in the shops (Figure 18). The threats to *Helichrysum* spp., however, are higher due to the number of *Sales/a*. The eigenvalues for the ordination are listed in Table 25.

Table 25: Summary of the ordination for 22 species and 12 variables.

Axes	1	2	3	4
Eigenvalues	0.364	0.249	0.108	0.089
Cumulative % variance of variable data	36.4%	61.3%	72.1%	81.0%

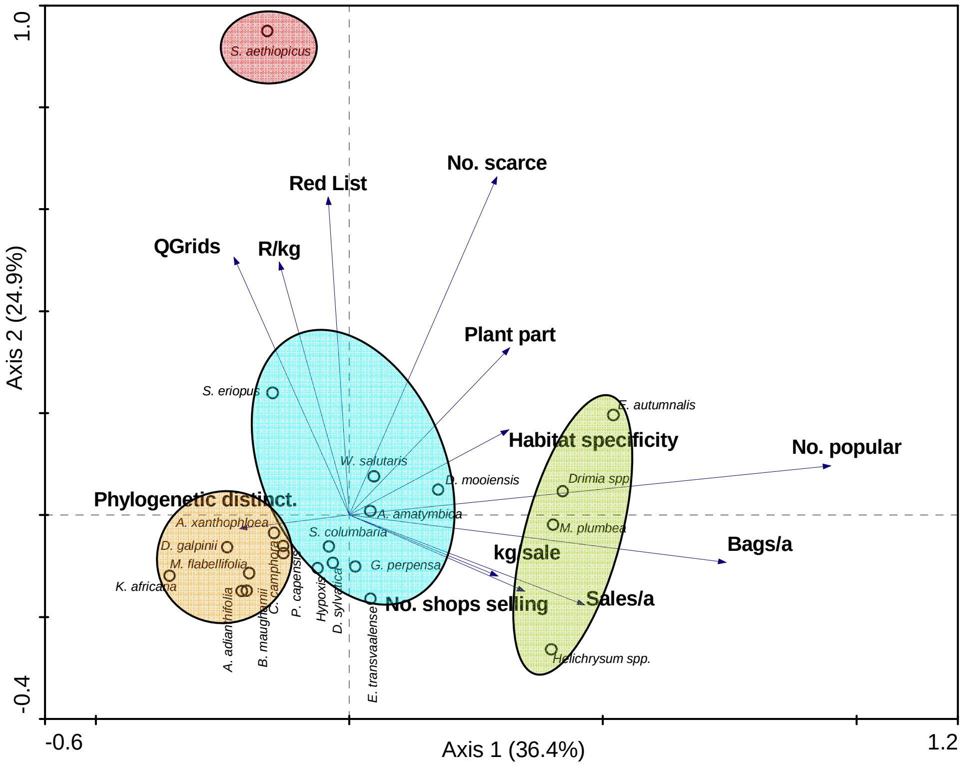


Figure 18: PCA for 22 test species using additional trade variables of R/kg, kg/sale, sales/a and bags/a obtained from the *muti* shop survey in 1994/5. Areas in coloured circles correspond with the similarly coloured clusters in Figure 19.

Ward's cluster analysis of the same data identified four clusters at >40% relative distance (Figure 19). The colours of the clusters correspond with the areas and species highlighted in the biplot of Figure 18. Cluster D (*Drimia* spp. etc.) are relatively higher risk due to trade factors, whereas cluster C (*S. aethiopicus*) was higher risk due to a combination of biological variables. Species in clusters A and B could be considered lower risk than species in C/D, especially species in cluster B (*P. capensis* etc). However, some species in A are known to be overexploited; hence, current harvesting scenarios have resulted in a decline in their prevalence and volume in the market.

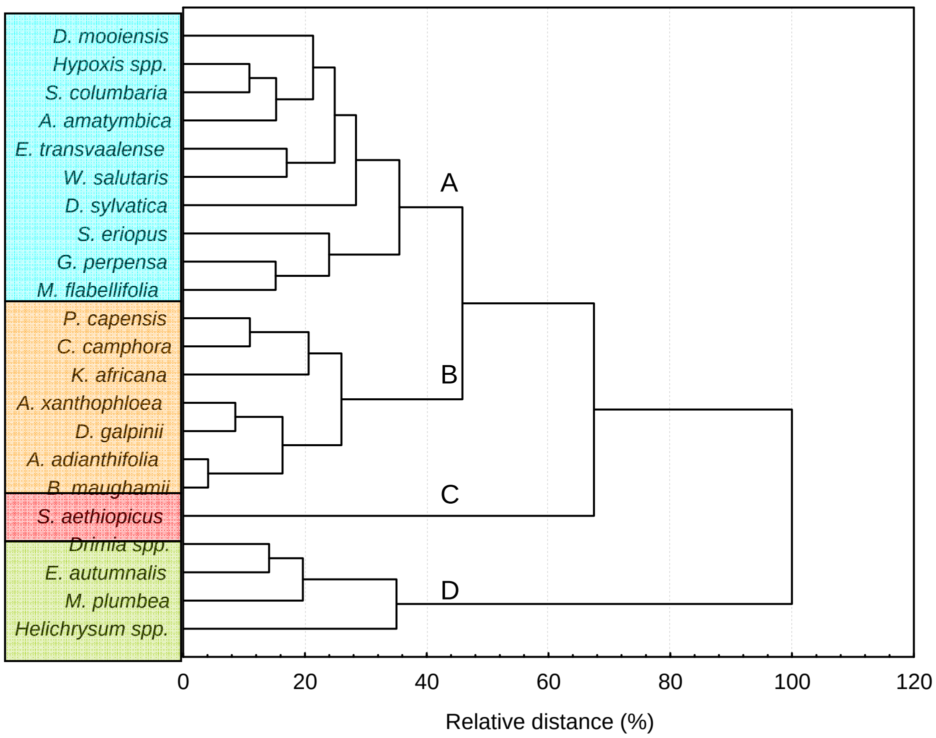


Figure 19: Dendrogram based on Euclidean distances and Ward's agglomerative clustering method for 22 test species using data from the 1994/95 shop survey, including *R/kg* and *kg/S*. Compare with Table 26.

K-means analyses for the same species and variables optimally clustered the species into 4 groups (Table 26). Three of the four bulb species were in cluster $K_4$1, distinctive because of the popularity and prevalence of the species in the shops, as well as the high mass per unit sale. *Siphonochilus aethiopicus* and *Helichrysum* spp. were assigned to separate clusters ($K_4$3 and $K_4$4 respectively), the former mainly because of the very high mean *R/kg* value, *Red List* status and restricted distribution (Table 26). *Helichrysum* spp., by contrast, appears threatened because of the high number of sales per annum to customers and the large volume purchased by the shops annually.

The K-means analyses showed that *R/kg* was hugely significant for assigning the species to the K-means clusters. The highest R/kg values were for *S. aethiopicus* (R760/kg), *K. africana* (R116/kg) and *A. amatymbica* (R99/kg). *Kigelia africana*, however, is not a scarce species, and the fruits are chopped into chunky pieces that dry out. The lowest R/kg values were for *Drimia* spp. (R12/kg) and *M. plumbea* (R10/kg). Following *R/kg* in descending order of significance for assigning species to clusters were the variables *bags/a*, *sales/a*, *Red List*, *no. scarce*, *no. popular*, *no. shops sell* and *QGrids* (Table 26). Least significant were the variables *kg/S*, *plant part*, *phylogenetic distinctness* and *habitat specificity*.

Table 26: Mean variable values per cluster corresponding with 4 K-means clusters for 22 test species and 12 variables. Yellow-shaded variable values are the highest in the row. The coloured species boxes correspond with the PCA clusters in Figure 18 and the dendrogram clusters of Figure 19.

Variables	$K_4 1$	$K_4 2$	$K_4 3$	$K_4 4$	F	P
Habitat specificity	2.3	1.9	2	2	0.3	0.824
Phylogenetic distinctness	2	2.8	2	2	0.6	0.598
Plant part	5	3.5	5	2	2.5	0.092
Red List	0.3	0.2	5	0	22.8	<0.001
QGrid	96	48	11	432	7.0	0.003
No. shops sell	40	29	14	36	7.4	0.002
No. scarce	12	2	18	0	16.2	<0.001
No. popular	6	1	2	6	8.0	0.004
R/kg	13.4	55.4	759.7	48.4	239.2	<0.001
kg/sale	0.36	0.15	0.02	0.09	2.6	0.084
Sales/a	977	467	0	2224	29.2	<0.001
Bags/a	1669	314	20	2205	58.0	<0.001
Species	*Drimia spp.* *E. autumnalis* *M. plumbea*	*S. eriopus* *D. sylvatica* *Hypoxis spp.* *S. columbaria* *D. mooiensis* *A. amatymbica* *G. perpensa* *M. flabellifolia* *E. transvaalense* *W. salutaris* *P. capensis* *D. galpinii* *C. camphora* *A. xanthophloea* *A. adianthifolia* *B. maughamii* *K. africana*	*S. aethiopicus*	*Helichrysum spp.*		

The pattern of species clusters for the PCA, Ward's and K-means analyses are generally similar, as shown by the similarly coloured blocks of species in Figures 18, 19 and Table 26. Each method offers a complimentary view of species with similar risks, threats and priorities, as well as the major criteria for assigning species to clusters. What is also evident from the PCA and cluster analyses that include trade variables such as *R/kg*, *kg/sale* and *bags/a*, is that some species previously flagged as high risk or conservation priority can be clustered with lower risk/priority species because of the non-linear, elevated values associated with smaller and/or lighter plant parts and/or species.

When used as clustering variables, *R/kg* and *kg/sale* clearly differentiate between plant part groups when allocating species to clusters. In the example shown in Figure 20, species sold as bulbs, roots, rhizomes and tubers were placed into two clearly defined clusters: Cluster A (red), comprising perennial herbs with rhizomatous roots or geophytes and herbs with woody rootstock; Cluster B (blue) comprising heavier bulbous geophytes, tubers or lignotubers. The variable primarily responsible for assigning species to these clusters was *kg/sale* (Table 27). K-means analyses of the bulb, root and rhizome data also clustered the species into the same two groups (Table 27). The species were assigned to two clusters: a) bulbs and tubers with a high mean kg/S, and b) smaller roots and rhizomes with high R/kg values.

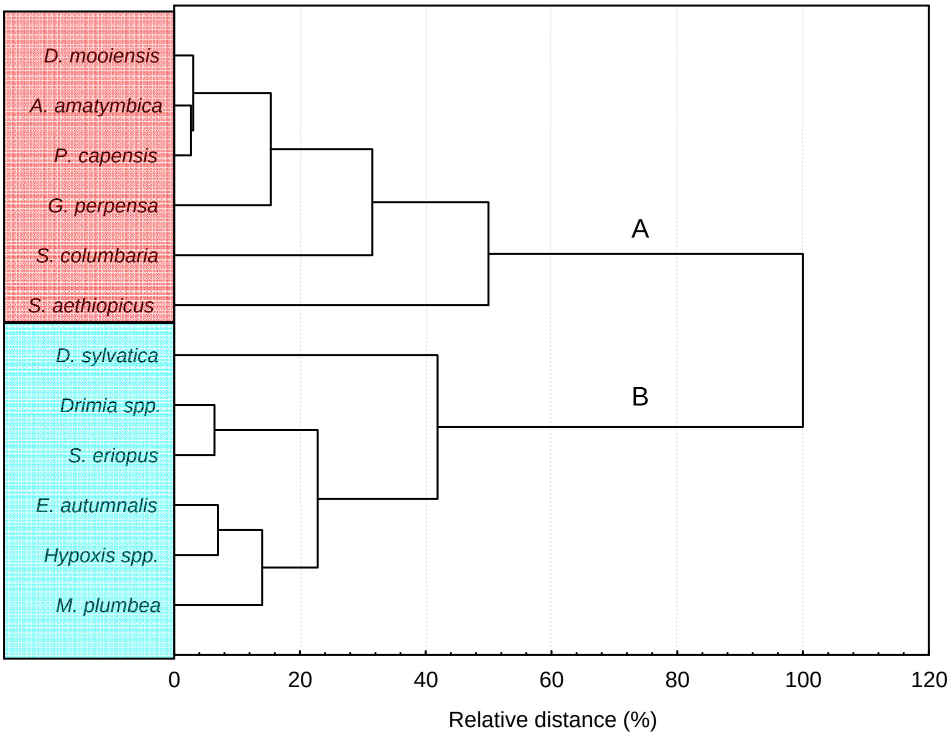


Figure 20: Dendrogram based on Euclidean distances and Ward's agglomerative clustering method for 12 species using four variables, namely R/kg and kg/sale derived from both shop and Faraday data. The blue group comprises heavier tubers and bulbs, whereas the red cluster comprises lighter rhizomes and roots.

Table 27: Mean values for variables corresponding with K-means analysis for two clusters and 12 species. Yellow-shaded values indicate the highest variable values in the row, whereas red and blue shaded species correspond with clusters in Figure 20. [df = 1,10]

	$K_2 1$	$K_2 2$	F	P
Faraday: R/kg	38.6	6.4	5.3	0.043
Shops: R/kg	181.9	22.8	1.9	0.200
Faraday: kg/sale	0.10	0.52	19.7	0.001
Shops: kg/sale	0.07	0.33	11.2	0.007
Species	*D. mooiensis* *S. columbaria* *S. aethiopicus* *G. perpensa* *A. amatymbica* *P. capensis*	*Drimia spp.* *E. autumnalis* *M. plumbea* *Hypoxis spp.* *D. sylvatica* *S. eriopus*		

10.5.10: Cluster Analysis versus Numerical Importance Value Indices

The results of the cluster analysis for 87 species were compared with the Numerical Importance Values (NIVs) for 10 variables (Table 28). Species in Table 28 were placed in descending order of their NIVs, and the species groups partitioned by 4 K-means clusters (Section 10.5.7: Figure 11 and Table 19) were identified by the colours red, orange, green and blue. The species NIV scores were as follows:

- species numbered 1 (*Drimia* spp.) scored 6.2
- species numbered 2–5 scored between 5.2 and 5.0;
- species numbered 6–21 scored between 4.6 and 4.0;
- species numbered 22–54 scored between 3.9 and 3.0;
- species numbered 55–86 scored between 2.9 and 2.2;
- and, the species numbered 87 (*Asparagus* spp.) scored 1.9.

Each NIV score group was placed in boxes enclosed with double lines in Table 28.

The comparative list in Table 28 shows how clustering groups species of similar priority, but not necessarily in the exact order of the NIV. Hence, a high NIV doesn't necessarily imply a high priority in the same numerical order as the NIV. If species in Table 28 had to be assigned priorities based on NIV scores, then the cut-off points for each threat group could have been at species numbers 1, 5, 21, 54 (see boxes with double lines in the Table). Under the NIV system, some lower conservation priority species [e.g. *Eucomis bicolor* (#17), *Synaptolepis kirkii* (#37) and *Rhoicissus tridentata* (#41)] would have had a higher conservation priority than the cluster analyses identified. Similarly, some higher conservation priority species [e.g. *Myrothamnus flabellifolia* (#40), *Adenia gummifera* (#34) and *Rapanea melanophloeos* (#36)] would have had lower priorities than the cluster analyses identified.

Correlations between the rankings of species in the various risk categories were positive, but not completely concordant (r_s = 0.78) (Figure 21). In general, species identified as belonging to the highest conservation priority categories from the cluster analyses generally ranked highest, or next highest priority when compared with NIV categories. When the correlations were carried out with Red Listed species, however, the rank correlation was low (r_s = 0.345) (graph not shown), thus indicating that very few of the 87 short-listed species have a Red List rating.

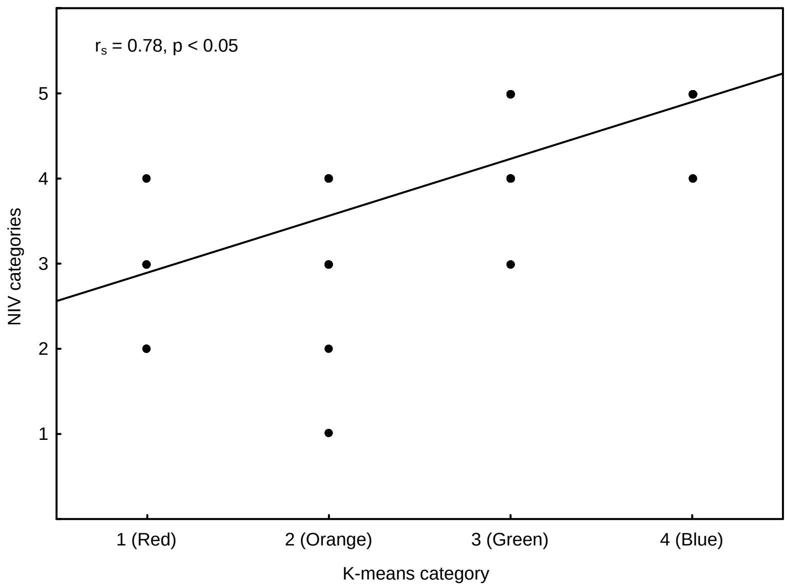


Figure 21: Relationship between the conservation priority assessments of 87 species made by K-means and the categories based on similar ranges of NIV scores. Categories correspond with those in Table 28.

Table 28: Rank order of 87 ethnospecies listed in descending value of the Numerical Importance Value (NIV) for 10 variables. The colours of red, orange, green and blue correspond with the species partitions for 4 K-means clusters (see Figures 11 and Table 19). Species in the red and orange blocks are high conservation priority; species in green have intermediate conservation priorities, and species in blue have lower priorities. Species enclosed in boxes bordered by double lines had a similar range of NIV scores, a criterion normally used for categorising priority.

Number	Species
1	Drimia spp. (altissima, elata, sanguinea)
2	Hydnora africana
3	Siphonochilus aethiopicus
4	Eucomis autumnalis
5	Clivia spp. (caulescens, gardenii, miniata, nobilis)
6	Stangeria eriopus
7	Euphorbia woodii
8	Gunnera perpensa
9	Schlechterina mitostemmatoides
10	Hypoxis spp. (colchicifolia, hemerocallidea)
11	Warburgia salutaris
12	Callilepis laureola
13	Elaeodendron transvaalense
14	Acacia xanthophloea
15	Merwilla plumbea
16	Urginea spp. (delagoensis, epigea)
17	Eucomis bicolor
18	Bowiea volubilis
19	Eriospermum mackenii
20	Ekebergia capensis
21	Schotia brachypetala
22	Sclerocarya birrea spp
23	Curtisia dentata
24	Albizia adianthifolia
25	Ocotea bullata
26	Dioscorea sylvatica
27	Stapelia gigantea
28	Dianthus mooiensis
29	Dietes spp (iridoides, butcheriana)
30	Eulophia speciosa
31	Cassipourea spp. (flanaganii, gerrardii, malosana)
32	Ornithogalum longibracteatum
33	Talinum caffrum
34	Adenia gummifera var. gummifera
35	Alepidea amatymbica var. amatymbica
36	Rapanea melanophloeos
37	Synaptolepis kirkii
38	Raphionacme hirsuta
39	Dombeya rotundifolia
40	Myrothamnus flabellifolia
41	Rhoicissus tridentata
42	Cephalaria humilis
43	Boophone disticha
44	Thesium pallidum
45	Maytenus undata
46	Senecio coronatus
47	Secamone gerrardii
48	Helichrysum spp. (8spp.)
49	Dioscorea dregeana
50	Agathosma ovata
51	Tulbaghia spp. (ludwigiana, natalensis)
52	Trichilia spp. (dregeana, emetica)
53	Elephantorrhiza elephantina
54	Silene bellidioides
55	Gnidia kraussiana
56	Hippobromus pauciflorus
57	Aster bakerianus
58	Anemone spp. (caffra, fanninii)
59	Crinum spp. (delagoense, macowanii, moorei)
60	Ansellia africana
61	Pelargonium luridum
62	Haworthia limifolia/Aloe aristata
63	Balanites maughamii
64	Pentanisia prunelloides
65	Scabiosa columbaria
66	Polygala spp. (confusa, serpentaria)
67	Eriosema sp. (mackenii, salignum)
68	Pappea capensis
69	Pleurostylia capensis
70	Gnidia spp. (cuneata, kraussiana)
71	Bridelia cathartica
72	Croton spp. (gratissimus, moorei)
73	Aristea spp.(ecklonii, woodii)
74	Acalypha villicaulis
75	Cryptocarya spp.(latifolia, myrtifolia, woodii)
76	Diospyros galpinii
77	Andrachne ovalis
78	Vitex spp. (rehmanii, wilmsii)
79	Calodendrum capense
80	Capparis spp. (brassii, tomentosa)
81	Zanthoxylum spp. (capense, davyi)
82	Bersama spp. (lucens, staynerii, swinnyi, tysoniana)
83	Plectranthus hadiensis
84	Rubia cordifolia
85	Agapanthus spp. (africanus, campanulatus, caulescens, praecox)
86	Vernonia tigna
87	Asparagus spp. (asparagoides, laricinus, plumosus, ramosissimus, virgatus)

Hence, the value of multivariate-based analysis, especially cluster analysis, is that groups of species with similar risk and conservation profiles are recognized. The clusters largely correspond with species placed in rank order of their NIV, and higher and lower priority clusters can be identified, but species are not ranked within a homogenous cluster. This makes it easier to set management guidelines objectively for species clusters.

10.6 General Discussion and Conclusions

Models are built to answer specific questions. In doing so, features and processes seen as important are incorporated, and superfluous ones are omitted (Burgman *et al*. 1993). In this chapter, I explored the development of a risk assessment method/protocol to answer questions about the relative risks of species harvested for the medicinal plant trade on the Witwatersrand, using features seen as important to the trade. The risks in question are the risks of over-utilisation, harvesting and ultimately extinction should current levels of harvesting persist. Another risk is to the abundance and persistence of specific plant populations due to unsustainable harvesting.

Previously in Chapter 6, a question regarding the selection of indicator species[3] was posed, namely: how does one objectively select criteria and categories for delimiting high risk species from those that are lower risk? Cunningham (2001) has described some of the categories for choosing priority species for monitoring as "filters" which help sift out species that are likely to be more vulnerable to over-harvesting. The simple (albeit crude) method developed in Chapter 6 was obtained from Hill's diversity numbers and used to calculate the number of species of common, intermediate and rare abundances in an ethnobotanical sample (Williams *et al.* 2005). This method can be used as a first step in choosing species for monitoring based on their relative abundances. Of the 392 ethnospecies sold between 150 traders, 1% (4 esp) were very commonly sold and 56% (220 esp) were of intermediate abundance. In this chapter, a multivariate approach was taken to assess the risks, threats and conservation priorities to species. The risk assessment method explored here incorporated cluster analysis and PCA to place species into a number of categorical risk hierarchies.

One purpose of cluster analysis is to partition objects (such as species) into groups suggested by the data, not defined *a priori*, so that objects in a given cluster tend to be similar to each other and objects in different clusters are dissimilar (Ramos 2001). Ward's hierarchical clustering has the advantage of visualising (in dendrograms) the combination of the observations to form clusters, which thus facilitates the number of clusters to be retained (Ramos 2001). The K-means divisive technique is also powerful enough to classify the species. Additionally, the analysis of variance and cluster means assists with identifying the number of clusters to be retained and the characteristics of the species in the cluster. Results from the combination of hierarchical and divisive cluster methods led to the selection of the final number of clusters, the identification of risk and conservation priority hierarchies/groups, the identification of higher risk species and, the recognition of species with greater conservation priority. Similarly, redundant variables were identified. In smaller data sets, for example, *plant part* plays a diminished role in assigning species to clusters, whereas *scarcity* has less of an influence in large data sets. As data sets decrease in size (i.e. fewer species), there is a greater chance that individual species scores for a variable will be the same, hence the variable is rendered ineffective in assigning species to clusters (e.g. *endemism*).

Groups of higher-risk species were identified from the individual surveys of the *muti* shops and Faraday market (Tables 6 and 11 respectively) using cluster analysis and 4-5 trade variables. However, it was from combining the two data sets to form a compound list of 392 ethnospecies that a short-list of 119 higher-risk ethnospecies was identified (Table 16). This list was further reduced to 87 species for further analysis to ascertain conservation priorities based on the additional inclusion of several biological variables. Species not on the short-lists are not considered to be *as* threatened by harvesting for the medicinal plant trade; however, they may be at risk due to other deterministic and stochastic factors and extra pressures from the *muti* trade may exacerbate the problem. From the list of 87 higher-risk species, ≈31 species were concluded to have a higher conservation priority (Table 19) due to greater prevalence in the markets, narrower geographic distribution, greater phylogenetic distinctness, more habitat specificity and a higher level of endemism.

The methods experimented with here are not conventionally used to differentiate the risks or threats to groups of species. In systematics, for example, clustering is used to decide whether there are groups of species and not whether they have different levels of advancement. However, clustering has been applied in at least three studies to group species of similar risk. Hall (1993) applied cluster analysis in setting conservation priorities for 20 threatened plant species from the Cape Peninsula. Given and Norton (1993) used cluster analysis (the unweighted pair-group centroid linkage rule) and

[3] Indicator species in ethnobotany are usually those species in high demand by resource users and at risk of over-exploitation and population decline.

PCA to assess threats and set conservation priorities for selected plant species in New Zealand. Myers (1999) used cluster analysis (Ward's minimum variance) on a species attribute matrix for birds in New Mexico to identify groups with similar risks to contaminant exposure. The species in this study were statistically placed into clusters of analogous variable values using the cluster techniques. With the valid assumption that higher variable values are indicative of higher risk or priority, the decision to categorise a cluster of species as high/low risk or priority was based on the range or mean value of the variables. The higher the values for a cluster, the higher the threats to the species were considered to be.

Conservation priority setting and threat assessments for medicinal plants in South Africa have primarily been based on numerical rating systems and ranked scores (e.g. McKean 1993; Mander & Quinn 1997; Dzerefos & Witkowski 2001). Part of the difficulty with these Numerical Importance Value (NIV) indices, is that the criteria for choosing a numerical cut-off point between one risk/threat/conservation priority category and another is often subjective, arbitrary, unexplained or vague. According to Nel *et al.* (2004), there is no objective criterion for determining when a score is sufficient to qualify a species as high priority. In a study on conservation priorities for non-volant Chilean mammals, for example, Cofré and Marquet (1999) assumed that priority species for conservation had a conservation priority index (CPI) value of ≥12. This cut-off corresponded with the median and mean of the CPI frequency distribution, and 60% of the species were considered to be of conservation priority. However, the authors admitted that their criteria for choosing a CPI cut-off of 12 were arbitrary. Furthermore, while taxa with very high and low scores tended to be high and low risk respectively, they were uncertain of the priorities for species between the extremes (Cofré and Marquet 1999). When I reanalysed the Chilean data using Ward's and K-means clustering (using data from the species attribute matrix in the published paper), three clusters were identified that corresponded to high, intermediate and low conservation priorities respectively. However, 76% of the species were of high to intermediate priority, compared to 60% identified by Cofré and Marquet (1999). Hence, a significant benefit of using cluster analyses in risk and conservation priority studies is the ability to objectively define category cut-off points where there are uncertainties, and to be able to categorise species that lie in the middle of the 'risk continuum'.

Like numerical conservation priority indices, where increased risk and conservation priority are implied from high scores for the variables, the PCA analyses illustrated the increased or decreased levels of threat depending on how far a species was from the ordination biplot origin. Different species were correlated more strongly with different variables; however, a low variable score for a threatened species might obscure the actual threats to that species. For example, *W. salutaris* was not conspicuous as a high priority species with the PCA for 392 and 87 species. In the cluster analyses, however, *W. salutaris* was usually grouped with similarly high priority species. Hence, risk and conservation priority analyses benefit from using a range of multivariate analyses to highlight groups of species with similar levels of threat depending on the relative magnitude of the variable scores.

While the methods used and proposed are a departure from the often-used linear ranking schemes and advocate a multivariate approach, there is still some merit in using NIVs (as evidenced from Table 28). The alternate approach to risk assessment is thus to use NIVs to prioritise species based on rank, and then to use PCA to assign risk categories to species based on their distance from the origin of the biplot. In some cases it will be possible to see what factors lead to, and are more strongly correlated with, their level of risk and conservation priority.

The risk assessment process highlighted the complex nature of threats to species and of the variables used to reveal the potential risks. The risks to some species were more positively correlated with trade variables, while others were correlated with biological variables. Furthermore, some variables can only be used when wanting to assign conservation priorities, while others (such as R/kg) can only be used on certain groups of species (e.g. species used for bark). Through the species-attribute matrices and the clustering process, the information could be reduced to a simple dendrogram showing clusters that were homogenous within and heterogeneous between (Kuo *et al.* 2002). In addition, the risk assessment process did not specify the order of priority of individual species at risk by the trade, but rather clusters of species with similar risks/threats and/or conservation priorities. Hence, it took the focus off single species scores and ranks, and directed attention towards groups of similarly threatened species.

Robertson *et al.* (2003) describe 'prioritization systems' (rather than rating or ranking systems) as methods that generally involve a set of criteria and some sort of scoring system against which threats can be evaluated. By using cluster analysis, a 'prioritization system' protocol was developed that assigned risk and conservation priorities to species based on their trade and biological attributes.

A subject not addressed in this study is how species should be managed given the risk characterisations and conservation priorities. Broadly, however, the risk hierarchies indicate how immediate the conservation actions taken should be and indicate the priorities for resource management. Perhaps the multi-pronged approach recommended by Dzerefos and Witkowski (1999, 2001) for the Abe Bailey Nature Reserve should be adapted and considered for species identified as high, medium or low risk/conservation priority. *Category 1* species were identified as indicators of over-exploitation, harvesting usually results in plant mortality and current levels are not sustainable (Dzerefos and Witkowski (1999, 2001). Propagation in nurseries and alternatives to the plants were recommended and needed. *Category 2* species were identified as having the potential to be harvested according to site- and species-specific quotas. Research was recommended to set quotas and harvesting levels according to prevailing conditions (either environmental or based on the condition of the resource). *Category 3* species were identified as appropriate for high impact harvesting (Dzerefos and Witkowski (1999, 2001). The findings and recommendations by Dzerefos and Witkowski (1999, 2001) were, however, based on species found in a specific vegetation type, namely Rocky Highveld Grassland. As a result, there is a predominance of species with perennial underground storage organs and a low diversity of species with a woody component derived from woodlands and savannas in their sample nature reserve. The extrapolation of the results and management categories to include woody species from savanna and forest biomes should therefore be done on a case-specific basis. To apply this system objectively, however, would require consideration of the combined risks of species-specific harvesting, growth rates, reproductive strategies and known distributions (Williams 2003).

In general, Category 1 can be applied to species characterised by high-impact harvesting that removes or damages the entire plant (e.g. bulbs, tubers, plants with woody rootstock), or where bark removal has negatively impacted the plant populations (Williams 2003). Category 2 can be applied to species where removal of the entire plant has not yet reached levels perceived to be currently unsustainable e.g. certain trees harvested for bark or plants harvested for their aerial parts. Category 3 management can be applied to species harvested for fruits, certain leaves and stems, prolific species and/or exotics. I would also suggest that a fourth management category be added between Categories 2 & 3 that allows for low impact harvesting on species used for whole plants, roots and bark.

In conclusion, it is generally felt that taxa identified as 'higher-risk' should be afforded greater conservation priority, even if the species is not currently Red Listed. Many species within the high risk and conservation priority categories are traded at volumes and levels that are not sustainable and the populations are threatened with fragmentation and quasi-extinction, especially if they exist outside protected areas. It does not follow, however, that species not categorized as high conservation priority by the Witwatersrand medicinal plant trade are not threatened in any way. There may be other factors determinant of their national risk profiles besides harvesting for traditional medicine. Such species would benefit from monitoring.

If conservation measures are only applied to a species once it has already become threatened, then the purpose of creating priority lists and awareness of the current unsustainability of harvesting is made partially redundant. In many respects, the priority lists generated from this research meet the criteria for Orange Listing, SANBI's proposed way of assessing the recording of the conservation importance of taxa that are of special concern but are not on a Red List and, would benefit from *"anticipatory conservation planning to avoid future Red Listing"* (Victor and Keith 2004). Like the proposed Orange List (Victor and Keith 2004), the risk assessment method developed in this thesis does not rank taxa in order of importance, but rather generates a list of taxa that should be afforded a measure of protection according to the Biodiversity Act.

References

Bennert, H.W. 2000. Risk assessment in German pteridophytes – methods and results of a field survey. *Zeitschrift für Ökologie und Naturschutz* 9: 27-34.

Burgman, M.A., Ferson, M.A. and Akçakaya, H.R. 1993. *Risk Assessment in Conservation Biology*. Population and Community Biology Series 12. Chapman and Hall, London.

Cofré, H and Marquet, P.A. 1999. Conservation status, rarity, and geographic priorities for conservation of Chilean mammals: an assessment. *Biological Conservation* 88: 53-68

Cunningham, A.B. 1988. *An Investigation of the Herbal Medicine Trade in Natal/KwaZulu*. Investigational Report No. 29. Institute of Natural Resources, Pietermaritzburg.

Cunningham, A.B. 1994. Integrating local plant resources and habitat management. Biodiversity and Conservation 3: 104-115.

Cunningham, A.B. 1996a. Working towards a "Top 50" listing. *Medicinal Plant Conservation* 2: 4-6

Cunningham, A.B. 1996b. *People, park and plant use. Recommendations for multiple-use zones and development alternatives around Bwindi Impenetrable National Park, Uganda.* People and Plants Working Paper 4. UNESCO, Paris.

Cunningham, A.B. 1997. The "Top 50" listings and the Medicinal Plants Action Plan. *Medicinal Plant Conservation* 3: 5-7

Cunningham, A.B. 2001. *Applied Ethnobotany. People, Wild Plant Use and Conservation*. People and Plants Conservation Series, Earthscan, London.

De Grammont, P.C. and Cuarón, A. 2006. An evaluation of threatened species categorization systems used on the American continent. *Conservation Biology* 20(1): 14-27.

Department of Environmental Affairs and Tourism. 2004. Draft Risk Assessment Framework. Threatened/Protected Species. Department of Environmental Affairs and Tourism, South Africa.

Diederichs, N. (ed) 2005. *Commercialising Medicinal Plants. A Southern African Guide*. Sun Press, Stellenbosch.

Ding, C.G., Woo, Y.Y., Sheu, H.-J., Chien, H.-C. and Shen, S.-F. 1996. An effective statistical approach for comparative risk assessment. *Risk Analysis* 16(3): 411-419.

Dzerefos, C.M. and Witkowski, E.T.F. 1999. *Sustainable Utilisation of Plant Resources on Abe Bailey and Roodeplaat Dam Nature Reserves*. Report to the Gauteng Directorate of Nature Conservation. Plant Ecology and Conservation Series No. 11, University of the Witwatersrand, Johannesburg.

Dzerefos, C.M. and Witkowski, E.T.F. 2001. Density and potential utilization of medicinal grassland plants from Abe Bailey Nature Reserve, South Africa. *Biodiversity and Conservation* 10: 1875-1896.

Evans, T., Jong Sam, R. and Duckworth, W. 2004. Management priorities amongst the harvested medicinal plants of Myohyang Mountains Protected Area, DPR Korea. *Medicinal Plant Conservation* 9/10: 28-38.

Farjon, A. 1995. Reorganisation and the action plan. Conifer Specialist Group Report. *Species* 24, Newsletter of the Species Survival Commission, IUCN, pp 64-65.

Frankham, R., Ballou, J.D., Briscoe, D.A. and McInnes, K.H. 2004. *A Primer of Conservation Genetics.* Cambridge University Press, Cambridge.

Garibaldi, A. and Turner, N. 2004. Cultural keystone species: implications for ecological conservation and restoration. *Ecology and Society* 9(3): 1. http://www.ecologyandsociety.org/vol9/iss3/art1.

Given, D.R. and Norton, D.A. 1993. A multivariate approach to assessing threat and for priority setting to threatened species conservation. *Biological Conservation* 64: 57-66.

Hall, A.V. 1993. Setting conservation priorities for threatened species: a joint grouping and sequencing method. *South African Journal of Botany* 59(6): 581-591.

Höft, M., Barik, S.K. and Lykke, A.M. 1999. *Quantitative Ethnobotany. Applications of multivariate and statistical* *analyses in ethnobotany.* People and Plants Working Paper 6. UNESCO, Paris.

IUCN, 2006. IUCN Red List of Threatened Species. www.iucnredlist.org

James, F.C. and McCulloch, C.E. 1990. Multivariate analysis in ecology and systematics: panacea or pandora's box? *Annual Review of Ecology and Systematics* 21: 129-166.

Kala, C.P., Farooquee, N.A. and Dhar, U. 2004. Prioritization of medicinal plants on the basis of available knowledge, existing practices and use value status in Uttaranchal, India. *Biodiversity and Conservation* 13: 453-469.

Kent, M. and Coker, P. 1992. *Vegetation Description and Analysis. A Practical Approach*. John Wiley & Sons, London.

Kuo, R.J., Ho, M.L. and Hu, C.M. 2002. Integration of self-organizing feature map and K-means algorithm for market segmentation. *Computers & Operations Research*, 29: 1475-1493.

Lange, D. 1998. *Europe's Medicinal and Aromatic Plants: their Use, Trade and Conservation*. TRAFFIC International UK.

Lepš, J. and Šmilauer, P. 2003. *Multivariate Analysis of Ecological Data using CANOCO*. Cambridge University Press, Cambridge.

Loxton, Venn and Associates. 1994. Driekoppies Dam Valuation and Resettlement Programme. Traditional and Medicinal Plants Mitigation Options. Report to Department of Home Affairs, Kangwane.

Ludwig, J.A. and Reynolds, J.F. 1988. *Statistical Ecology: a Primer on Methods and Computing*. John Wiley & Sons, New York.

Lykke, A.M., Kristensen, M.K. and Ganaba, S. 2004. Valuation of local use and dynamics of 56 woody species in the Sahel. *Biodiversity and Conservation* 13: 1961-1990.

Mander, J. and Quinn, N. 1997. *Trade in Wildlife Medicinals in South Africa*. Report prepared for TRAFFIC East/Southern Africa. Unpublished.

Mander, M. 1998. *Marketing of Indigenous Medicinal Plants in South Africa: A Case Study in Kwazulu-Natal*. FAO, Rome.

Marshall, N.T. 1998. Searching for a Cure: *Conservation of Medicinal Wildlife Resources in East and Southern Africa*. TRAFFIC International, U.K.

McKean, S.G. 1993. *Important Species Document: Utilised Plants*. Internal Report. Natal Parks Board, Pietermaritzburg.

Millsap, B.A., Gore, J.A., Runde, D.E. and Cerulean, S.I. 1990. Setting priorities for the conservation of fish and wildlife species in Florida. *Wildlife Monographs* 111: 1-57.

Myers, O.B. 1999. On aggregating species for risk assessment. *Human and ecological risk assessment* 5(3): 559-574.

Nel, J.L., Richardson, D.M., Rouget, M., Mgidi, T.N., Mdzeke, N., Le Maitre, D.C., van Wilgen, B.W., Schonegeval, L., Henderson, L. and Nesser, S. 2004. A proposed classification of invasive alien plant species in South Africa: towards prioritizing species and areas for management action. *South African Journal of Science* 100: 53-64.

Newton, D.J. and Chan, J. 1998. *South Africa's Trade in Southern African Succulent Plants*. TRAFFIC East/Southern Africa, Johannesburg, South Africa.

O'Grady, J.J., Burgman, M.A., Keith, D.A., Master, L.L., Andelman, S.J., Brook, B.W., Hammerson, G.A., Regan, T. and Frankham, R. 2004. Correlations among extinction risks assessed by different systems of threatened species categorization. *Conservation Biology* 18(6): 1624-1635.

Pisces Conservation Ltd. 2002. Community Analysis Package. Version 2.15. www.pisces-conservation.com.

Pfab, M. F. and Witkowski, E.T.F. 2000. A simple population viability analysis of the Critically Endangered *Euphorbia clivicola* R.A. Dyer under four management scenarios. *Biological Conservation* 96: 263-270.

Punj, G. and Stewart, D.W. 1983. Cluster analysis in marketing research: review and suggestions for application. *Journal of Marketing Research* 20: 134-148.

Raal, P.A., Burns, M.E.R. and Davids, H. unpublished. Beyond GIS: decision support for coastal development, a South African example. CSIR.

Ramos, M.C. 2001. Divisive ad hierarchical clustering techniques to analyse variability of rainfall distribution in the Mediterranean region. *Atmospheric Research* 57: 123-138.

Robertson, M.P., Villet, M.H., Fairbanks, D.H.K., Henderson, L., Higgins, S.I., Hoffman, J.H., Le Maitre, D.C., Palmer, A.R., Riggs, I., Shackleton, C.M. and Zimmermann, H.G. 2003. A proposed prioritization system for the management of invasive alien plants in South Africa. *South African Journal of Science* 99: 37-43.

Scott-Shaw, C. R. 1992. Conservation importance values of vascular plants in Natal. Natal Parks Board Internal Report.

Scott-Shaw, C. R. 1999. *Rare and Threatened Plants of KwaZulu-Natal and Neighbouring Regions*. KwaZulu-Natal Nature Conservation Service, Pietermaritzburg, South Africa.

Simeonov, V. 2003. Environmetric strategies to classify, interpret and model risk assessment and quality of environmental systems. *Clean Technologies and Environmental Policy* 5: 190-199.

StatSoft, Inc. 2001. STATISTICA (data analysis software system), Version 6. Electronic Manual. www.statsoft.com.

ter Braak, C.J.F. 1987. Ordination. In: Jongman, R.H.G., ter Braak, C.J.F. and van Tongeren, O.F.R. (eds) *Data Analysis in Community Landscape Ecology*. Pudoc, Wageningen, The Netherlands. Pp 91-173.

ter Braak. C.J.F. and Prentice, I.C. 1988. A theory of gradient analysis. *Advances in Ecological Research* 18: 271- 317.

Tietenberg, T. 1992. *Environmental and Natural Resource Economics*. Harper Collins, USA.

Victor, J. 2002. South Africa. In: Golding, J.S. (ed.) *Southern African Plant Red Data Lists.* Southern African Botanical Diversity Network Report No. 14. SABONET, Pretoria.

Victor, J.E. and Keith, M. (2004) The Orange List: a safety net for biodiversity in South Africa. *South African Journal of Science* 100: 139-141.

Von Ahleveldt, D., Crouch, N.R., Nichols, G., Symmonds, R., McKean, S., Sibiya, H. and Cele, M.P. 2003. Medicinal Plants Traded on South Africa's Eastern Seaboard. Porcupine Press, Brixton.

Wild, R.G. and Mutebi, J. 1996. *Conservation through community use of plant resources. Establishing collaborative management at Bwindi Impenetrable and Mgahinga Gorilla National Parks, Uganda.* People and Plants Working Paper 5. UNESCO, Paris.

Williams, V.L. 2003. *Hawkers of Health: an Investigation of the Faraday Street Traditional Medicine Market in Johannesburg, Gauteng*. Report to the Gauteng Directorate for Nature Conservation, DACEL. Plant Ecology and Conservation Series No. 15, University of the Witwatersrand, Johannesburg.

Williams, V.L. 2004. Trade and socio-economic value of forest and woodland resources within the medicinal plant market in Johannesburg. In: Lawes, M.J., Eeley, H.A.C., Shackleton, C.M. and Geach, B.G.S. (eds), *Indigenous Forests and Woodlands in South Africa: Policy, People and Practice*. University of Natal Press, Pietermaritzburg. pp. 439-472.

Williams, V.L, Balkwill, K. and Witkowski, E.T.F. 1997. *Muthi* traders on the Witwatersrand, South Africa - an urban mosaic. *South African Journal of Botany* 63(6): 378-381.

Williams, V.L., Balkwill, K. and Witkowski, E.T.F. 2000. Unravelling the commercial market for medicinal plants and plant parts on the Witwatersrand, South Africa. *Economic Botany* 54(3): 310-327.

Williams, V.L., Balkwill, K. and Witkowski, E.T.F. 2001. A lexicon of plants trade in the Witwatersrand *muti* shops, South Africa. *Bothalia* 31(1): 71-98

Williams, V.L., Witkowski, E.T.F. and Balkwill, K. 2005. Application of diversity indices to appraise plant availability in the traditional medicinal markets in Johannesburg, South Africa. *Biodiversity and Conservation* 14: 2971-3001.

Williams, V.L., Witkowski, E.T.F. and Balkwill, K. (in press) Volume and financial value of species traded in the medicinal plant markets of Gauteng, South Africa. *International Journal for Sustainable Development and World Ecology*.

Wong, J.L.G., Thornber, K. and Baker, N. 2001. *Resource assessment of non-wood forest products. Experience and biometric principles.* Non-Wood Forest Products 13. FAO