A technique for evaluating species richness maps generated from collections data

M.P. Robertson[a*] and N.P. Barker[b]

There is considerable pressure on conservation planners to use existing data from herbarium and museum collections for planning and monitoring, despite the weaknesses of such data. It is thus important to be able to assess the quality of this information. One application of these data is the production of species richness maps. However, sampling effort is generally not consistent throughout a region for maps generated from collections data, and it is thus desirable to identify geographic grid cells (such as quarter degree squares: QDS) for which there has been low sampling effort. We describe a technique that can be used to identify QDS that are likely to have low species richness that is due to insufficient sampling effort rather than to low actual species richness. The technique exploits relationships between climate and species richness to detect QDS that are poorly sampled. This approach offers advantages over the current practice of applying a single threshold across the entire map region to detect QDS that are poorly sampled. Here we report on the application of our technique to plant species richness data in the PRECIS database. Results reveal that the majority of QDS in the Flora of Southern Africa region can be considered to be poorly sampled, even when using conservative thresholds for richness values. The advantages and weaknesses of the technique are discussed and issues requiring further investigation are highlighted.

Introduction

For many groups of organisms, the biodiversity of southern Africa is poorly documented, both in terms of taxonomic composition (alpha diversity) as well as taxon distribution (beta diversity). The plants are among the better-understood groups, and the documentation and management of southern Africa's plant diversity is currently under way.[1] However, this exercise is hampered by a taxonomic impediment caused by the lack of young taxonomists to replace older specialists as they retire, the freezing of research posts in museums and herbaria, and a general lack of interest in, and funding for, fundamental taxonomic studies that form the basis of biodiversity science. In southern Africa, this impediment is being addressed through a recently completed SABONET project,[2–4] as well as by the establishment of the South African Biosystematics Initiative.[5]

However, even assuming that the taxonomic impediment is addressed, taxonomists still rely on specimen and collections data, and for at least some groups our understanding of the diversity in terms of 'what is found where' may also be limited by a 'collection impediment': there is simply not enough preserved material upon which decent taxonomic studies can be conducted. Despite substantial efforts, large areas of southern Africa remain poorly collected and under-represented by holdings in the main herbaria and museums. The animal and fungal kingdoms contain major groups about which almost nothing is known.[6–8] Similarly, there is limited computerized distribution data for determining the diversity of specific areas.

Databases that contain plant or animal biodiversity data are extremely valuable and are increasingly being used for setting conservation priorities.[9–12] Capturing collections data into databases is a funding and staffing priority in museums and herbaria, and is currently being actively promoted through the recently launched South African Biodiversity Information Facility.[13] The information contained in these databases has been referred to as collections data[14] and these are increasingly being used in biodiversity studies, largely due to easier access through the internet.[9–12]

Collections data are used to predict species distributions using correlative models,[10,14,15] and these are often used as the basis for

[a]Centre for Environmental Studies, Department of Zoology and Entomology, University of Pretoria, Pretoria 0002, South Africa.
[b]Department of Botany, Rhodes University, P.O. Box 94, Grahamstown 6140, South Africa.
*Author for correspondence. E-mail: mrobertson@zoology.up.ac.za

conservation planning.[10,16–19] Collections data are also used to generate species richness maps, either directly from collection records[16] or by combining predictions of species ranges.[10,20] These maps indicate the number of species that have been collected per unit area or grid cell in a given map region.[16,21] Species richness maps are used in systematic conservation planning,[22] to detect hotspots[23] and for identifying areas that appear to be poorly collected, so that collection efforts can be directed towards these areas in the future.[21]

Although specimen databases are extremely valuable, there are quality problems associated with collections data[11,14,24,25] that may limit the usefulness of products generated from these data. These include geographical bias (more accessible areas are sampled preferentially), taxonomic bias (greater representation of easy-to-study species) and temporal bias (records are based on collection during one season only).[14] Methods have been proposed to overcome some of these data quality problems.[20,25]

Conservation decisions based on species richness maps will only be as good as the data used to generate these maps. This indicates the need to evaluate these maps and the data sources used to generate them.[24] A major problem with collections data is that sampling effort may not be consistent throughout the map region.[14,24,25] In other words, some grid cells — such as quarter degree squares (QDS), a standard grid cell used in plant studies in southern Africa[26] — are visited more frequently by collectors than others, leading to differences in recorded species richness, which is a form of bias.[24] The accuracy of the species richness map depends on the extent to which the recorded species richness reflects actual species richness on the ground.

QDS in a regional map can have a variety of recorded species richness values ranging from 'low' to 'high', depending on the thresholds chosen to define these descriptors. A QDS with a low recorded species richness value may reflect a true state of affairs, or it may warrant a much higher value, the low assigned value resulting from insufficient collecting effort. In contrast, a QDS with a high recorded species richness may be so described for two (not mutually exclusive) reasons. First, the QDS may naturally contain a large number of species, or second, the area may be well collected (in comparison with other QDS) because it is close to a major centre, along a main road, or is in a well-collected national or provincial nature reserve. For example, Rebelo[16] reported that there were two QDS that contained more than 2000 species, values that were high compared with other QDS in the southern African region. These two squares happened to be associated with centres with major herbaria. Nonetheless, data from such QDS may be considered as a more accurate record of species richness, and this richness value is a better reflection of the diversity that other QDS can contain.

In summary, two cases can be defined. In the first, a QDS has a low recorded species richness when it is actually high. In the second, a QDS is given a high value for species richness owing to a relatively good collection effort. Of these, the first is likely to have the most serious consequences for conservation planning, as QDS that are actually rich in species (and which are under-recorded) may be overlooked.

Identifying QDS for which there has been low sampling effort is thus important. This can be done by listing those cells with species richness values below a particular threshold, for example, 200 species per grid cell. Deciding on a sensible threshold value is difficult, however, and assumes that all QDS with species richness values above the threshold have been adequately sampled. This approach does not acknowledge that species richness may vary according to environmental conditions. Because environmental factors affect species richness,[27,28] a more sophisticated approach is required to define these thresholds.

In this paper, we propose a technique that can be used to identify QDS that are likely to have low species richness due to insufficient sampling effort rather than resulting from low actual species richness. We describe this technique using the PRECIS database of plant collection records as an example.

Sampling geographical space and climatic space

Funk and Richardson[14] recently investigated the use of collections data in biodiversity studies to determine how well data from localities held in a GIS database sampled actual geographical and environmental space. Their technique identifies sampling gaps in collection records using a two-dimensional climate (environmental) space defined by annual mean rainfall and annual mean temperature. More sophisticated approaches have subsequently been developed to determine the optimal design of surveys that take into account environmental diversity of proposed sites and how these complement sites that have already been surveyed.[29]

Some studies have described relationships between plant species richness and climatic or environmental (e.g. nutrients or soil) variables (reviewed by Pausas and Austin[27]). Acevedo and Currie[30] have shown that climate (precipitation and temperature) and vegetation influence bird species richness. O'Brien[31] found that water–energy dynamics (determined by climate) accounted for more than 78% of the variation in woody plant species numbers, and Francis and Currie[28] found a globally consistent correlation between angiosperm richness and climate. It is thus clear that at certain scales species richness is correlated with climate. This relationship has been exploited to predict species richness at regional scales using various environmental variables, including those that account for energy and water.[32]

The technique that we describe is based on the approach described by Funk and Richardson,[14] with some important differences. In Funk and Richardson's approach, sampling localities are plotted on a map to indicate their geographical position, using coordinates. Thereafter, they are plotted in a climate space defined by annual mean rainfall and annual mean temperature. Any gaps or regions of interest within this climate space can then be digitized. Regions in geographical space that correspond with the regions in climate space can then be identified using the combinations of values of the two climatic variables. This technique allows the identification of specific combinations of climatic variables that have not been sampled and to indicate where these regions are in geographical space.

In contrast, in the technique that we describe the geographical space comprises a grid, with each grid cell (QDS) having an associated species richness value (Fig. 1), instead of having points in the geographical space that represent collection localities. Each of the QDS in the geographical space can be plotted in a two-dimensional climate space (Fig. 2). In this paper, we use temperature and rainfall variables to define the climate space. Other combinations of variables such as average temperature and potential evapo-transpiration could also be investigated. It must be noted, however, that the ability of the variables to explain species richness is related to scale.[33]

The climate space can be divided up into a grid of climate cells (Fig. 2). It is possible for a climate cell to contain several QDS, depending on its size. The QDS that occur within a particular climate cell can be identified and plotted in the geographical space (Fig. 3). If species richness depends on moisture and temperature, then areas with similar moisture and temperature should contain similar numbers of species. Thus QDS that are close to one another in the climate space should have similar

species richness values. A climate cell that contains several QDS, some with high and some with low species richness values, is interesting because the higher values can be used to indicate the potential species richness of that cell. The QDS with low species richness values can be flagged as having low values that are likely to be due to poor sampling effort, and QDS that have species richness values below a given threshold may be plotted in geographical space. Most importantly, this threshold is specific to each climate cell, and its value is determined by the maximum recorded species richness values in the climate cell.

Heterogeneity within QDS

In addition to climate, environmental heterogeneity within QDS or other areas may be an important determinant of species richness.[27,34–36] Squares with highly heterogeneous abiotic conditions are assumed to provide more niches than those that are relatively homogeneous, thereby allowing more species to coexist. Vivian-Smith[37] found a positive relationship between species richness and heterogeneity at the local scale, using pot experiments. At a regional scale, Rey Benayas and Scheiner[36] found a positive relationship between species richness and altitudinal range (which is a measure of heterogeneity). Harner and Harper[34] noted a positive relationship between species richness and an index of environmental heterogeneity at local scales. Recently, Pausas *et al.*[38] investigated the extent to which heterogeneity of the terrain and of the environmental conditions could explain species richness in a 10 km × 10 km grid system in the Iberian Peninsula. They found that species richness was a function of both within-grid heterogeneity (which accounted for approximately two-thirds of the explained deviance) and environmental conditions. In our study, environmental heterogeneity is accounted for by using an index that incorporates both spatial and temporal components.[36]

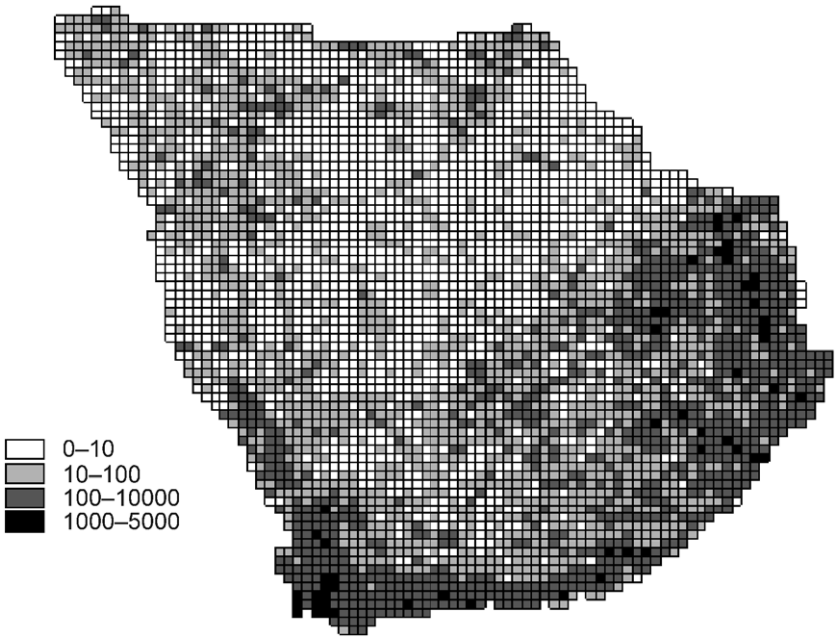

Fig. 1. Species richness values per quarter degree square (QDS) recorded in the PRECIS database for the FSA region (including South Africa, Lesotho, Swaziland, Namibia and Botswana) as of March 2003.

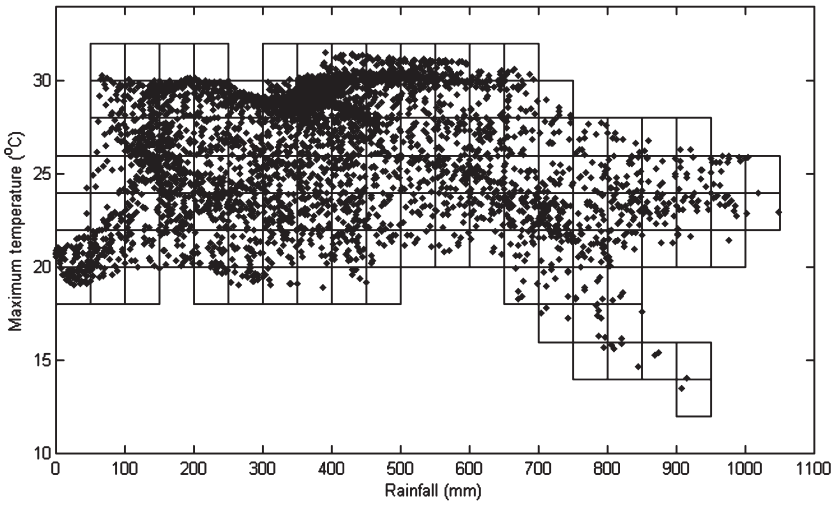

Fig. 2. QDS in the FSA map region plotted as points in a two-dimensional climate space that is defined by the annual mean of daily maximum temperature (*y*-axis) and annual rainfall (*x*-axis). The climate space is divided into a number of climate cells defined by 2°C temperature increments and 50-mm rainfall increments.

Methods

The PRECIS database

The National Herbarium (PRE) is a leader in collection databasing. PRE houses over 700 000 plant specimens from the Flora of Southern Africa (FSA) region, an area that includes South Africa, Lesotho, Swaziland, Botswana and Namibia. Data on all of these plant collections are stored in the PRECIS (PREtoria Computerised Information System) database.[39–43] Databases such as PRECIS not only aid curation and management of the holdings, but also (increasingly) provide data to other scientific users, environmental consultancies and lobby groups. Much of the information required is essentially geographic in nature; species distributions and ecological notes that can be incorporated into GIS databases and mapping routines. An additional advantage is that species diversity at a landscape scale can be readily assessed, providing invaluable data to landscape ecologists and conservationists. In terms of biodiversity, PRECIS has been used to investigate alpha, beta and gamma diversity at the regional and biome levels,[44–46] floristic analyses,[47] and reserve selection[16] and has formed the basis of some major taxonomic treatments.[48]

The data

A list of all QDS in the FSA region, together with the number of species recorded in each, was obtained from the National Herbarium in 2003. These data were used to produce an initial species richness map. It must be noted that not all the specimens in the PRECIS database have QDS reference data, so we are working with a subset of all PRECIS specimen data.

The climate data were obtained from interpolated grids (maps) for Africa from Hutchinson *et al.*[49] The information included annual mean daily maximum temperature, means of daily maximum temperature for each month of the year, mean annual rainfall and mean monthly rainfall for each month. The variables were converted from a grid with a spatial resolution of three

minutes to one with QDS resolution by averaging the values within a QDS. We obtained a high-resolution digital elevation model (DEM) with a grid cell size of 0.5′ × 0.5′ from WORLDCLIM.[50]

Analyses

Programs were written in MatLab[51] to process the data and to produce the figures. Points that represent QDS in the FSA region were plotted in a two-dimensional climate space, defined by the annual mean daily maximum temperature (*y*-axis) and annual rainfall (*x*-axis). The position of each point (representing a QDS) in the climate space was determined by obtaining values associated with each QDS from the climate variable maps. The climate space was then divided into a number of climate cells defined by 2°C temperature increments and 50-mm rainfall steps; summary statistics were calculated for the QDS that fell within each of the climate cells. These statistics included the maximum species richness value, the range of species richness values, and the number of QDS within each climate cell. A map of the QDS that fell within a selected climate cell was then produced. Finally, maps were drawn that show those QDS identified as having low species richness due to poor sampling effort. These squares were identified as follows: for each climate cell the maximum species richness was calculated from the richness values for the QDS that occurred within that climate cell. A species richness threshold was then defined using a richness value that was a selected percentage of the maximum value. Conservative values of 10% and 20% were used.

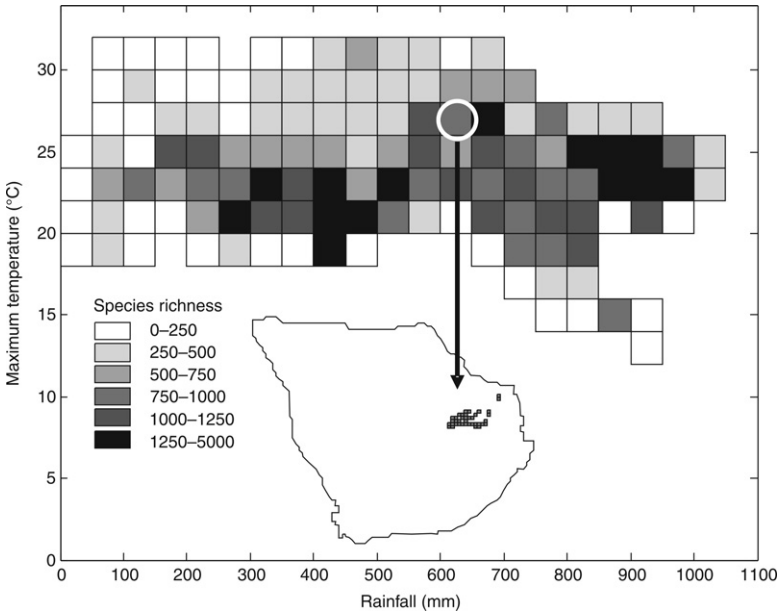

Fig. 3. The range of the species richness values (recorded in QDS) occurring within each climate cell. The inset indicates the QDS that occur within the selected climate cell.

For each climate cell, those QDS within the climate cell with species richness values below the species richness threshold could be explained by a combination of sampling effort and environmental heterogeneity. Assuming that there is a positive relationship between environmental heterogeneity and species richness, as several studies have suggested,[27,34,36–38,52] QDS with high environmental heterogeneity could have higher species richness than those with low environmental heterogeneity. We consider the possible contribution of environmental heterogeneity to species richness by defining an environmental heterogeneity threshold in addition to the species richness threshold described above. The environmental heterogeneity threshold is defined using the value of an environmental heterogeneity index recorded for the QDS with the maximum recorded species richness in each climate cell.

The index considers both spatial and temporal components and comprises three factors. For the first factor (spatial), we calculated the standard deviation of altitude values, from a high-resolution digital elevation model, for each QDS. Each square contained 900 grid cells from the high-resolution DEM. The two other factors consider the temporal component of environmental heterogeneity. For these, we calculated the coefficients of variation in rainfall and temperature across 12 months, respectively, using grids at QDS resolution. All three factors were scaled between zero and one by dividing each by the maximum value recorded for that factor. The scaled factors were then added to produce a single index of environmental heterogeneity and could theoretically take any value between zero and three.

Our intention was to identify those QDS within a climate cell that had low recorded species richness due to poor sampling effort rather than environmental heterogeneity. To do this, we assumed that those QDS within a selected climate cell that a) have recorded species richness values below the species richness threshold and b) have higher environmental heterogeneity values than the QDS with the maximum recorded species richness, have not been sampled adequately. We can be less certain about QDS that have environmental heterogeneity values that were lower than the QDS with the maximum recorded species richness, as low recorded species richness could be attributed to low environmental heterogeneity.

We identified QDS with low species richness that had environmental heterogeneity values above and below the environmental heterogeneity threshold, and then plotted these on a map of the FSA region.

There are two broad rainfall regimes in southern Africa: the winter rainfall regime (WRR) and summer rainfall regime (SRR). Because of this, we considered it appropriate to explore the application of this method to the WRR separately, since rainfall seasonality may have a marked effect on species diversity, and it could be inappropriate to use climate cells which comprised a mix of WRR and SRR QDS to determine and assess diversity and (more importantly) collection intensity. We thus undertook a second analysis, selecting QDS from the WRR for further investigation and comparison with the analysis of the region as a whole.

Results and discussion

Overall species richness

The 'traditional' species richness map based on the PRECIS data is shown in Fig. 1. As was the case when such a map was first published in 1984,[21] and again in 1994,[16] it is clear that recorded species richness varies considerably across the FSA region. A large number of QDS have recorded species richness values between zero and 10; fewer QDS have species richness values between 1000 and 5000. As reported by Rebelo,[16] some QDS with high species richness are associated with major cities, where collecting is likely to have been most intense. In other regions, recorded species richness is very low. We are not implying that collection efforts have ceased. Between Rebelo's 1994 paper[16]

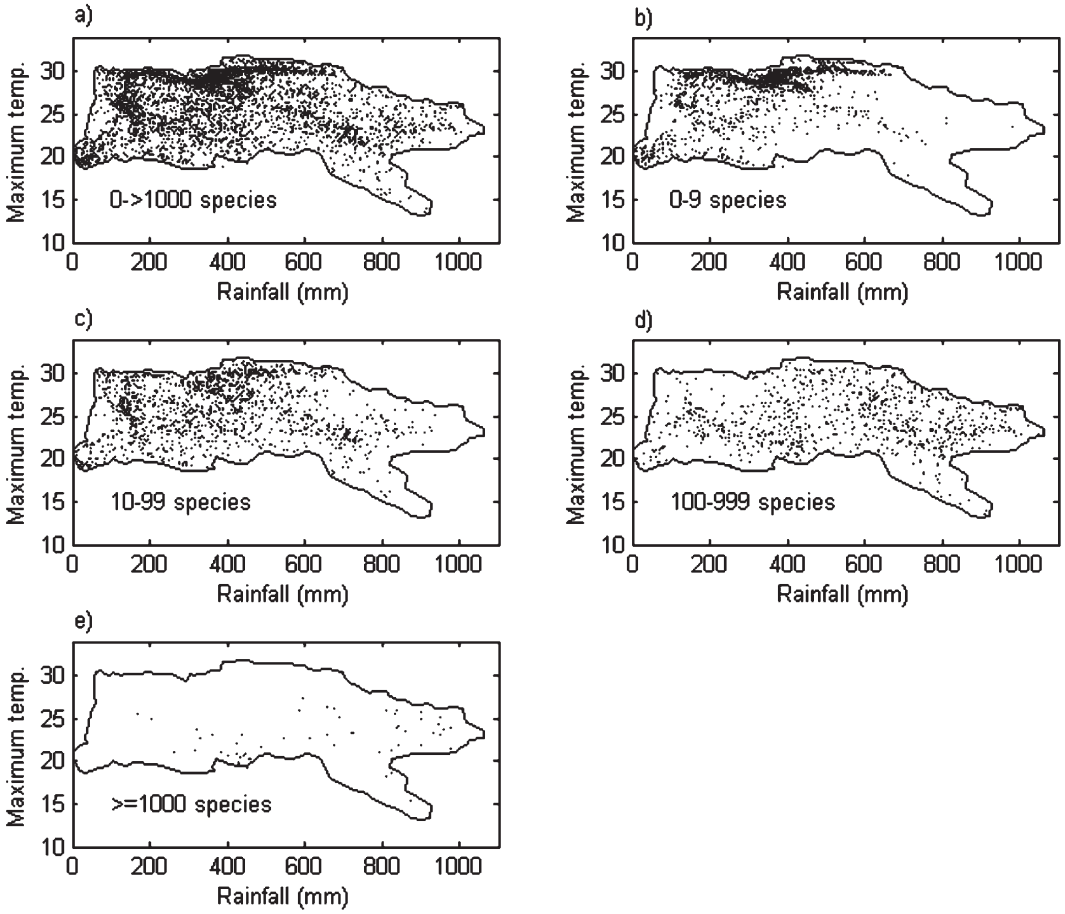

Fig. 4. QDS in the FSA map region plotted as points in a two-dimensional climate space that is defined by the monthly mean daily maximum temperature (*y*-axis) and annual rainfall (*x*-axis). **a**, All data; **b**, QDS containing between zero and nine species; **c**, QDS containing 10 to 99 species; **d**, QDS containing 100 to 999 species; and **e**, QDS containing 1000 or more species.

and 2003, there has been an increase in recorded richness for many QDS, as summarized in Table 1.

Figure 4a shows the QDS in the map region plotted as points in a two-dimensional climate space that is defined by the annual mean daily maximum temperature and annual rainfall. QDS with between zero and nine species tend to occur more frequently towards the lower rainfall side of the climate space (Fig. 4b); those with higher recorded species richness tend to occur more frequently in the higher rainfall region (Fig. 4c–e). However, there is considerable overlap in the regions of the climate space associated with QDS with different ranges of recorded species richness. This means that within a given region of the climate space there are dots representing QDS with a range of species richness values.

Table 1. Frequency distributions of number of species per quarter degree square, using the same categories as Rebelo.[16] The final column indicates the change in the frequency for each species richness category.

Number of species per grid square	Frequency (1994)	Frequency (2003)	Change (1994–2003)
0	Not given	468	
1	316	251	–65
2–5	549	538	–11
6–10	365	335	–30
11–50	940	1035	95
51–100	376	423	47
101–500	606	736	130
501–1000	117	153	36
1001–2000	26	49	23
>2000	2	5	3

When the climate space is divided into climate cells, each cell contains one or more points that represent QDS in geographical space (Fig. 2). Climate cells containing the greatest number of QDS tend to be in the hotter, drier regions of the climate space, indicating that this climate covers proportionally more of the FSA region. The range of the species richness values (recorded in QDS) occurring within each climate cell is shown in Fig. 3. Climate cells with narrower richness ranges (0–250 species) tend to occur on the edges of the climate space, whereas those with intermediate to large ranges (>750 species) tend to occur in the cooler regions of climate space, but span a considerable moisture gradient. The QDS that occur within any particular climate cell can be identified and plotted in the geographical space (Fig. 3).

Climate cells contain different numbers of QDS (Fig. 2) and have different ranges of species richness (Fig. 3). The relationship between the number of QDS per climate cell plotted against the range in species richness (as opposed to total richness) is weak (Fig. 5). This figure suggests that species richness range is not purely a function of the number of QDS within a climate cell, and there are many climate cells in which both well-collected and poorly collected QDS are found.

Identifying under-collected areas: FSA region

QDS from the FSA region that were identified as having low species richness due to poor sampling effort using a species richness threshold are shaded according to their environmental heterogeneity in Fig. 6. The QDS shaded in grey have recorded species richness values that are less than the richness threshold

and have environmental heterogeneity values below the corresponding threshold. The black squares have recorded species richness values that are less than the species richness threshold and also environmental heterogeneity values above the corresponding threshold. The QDS that are shaded (either grey or black) are considered to have low species richness due to low sampling effort, but the certainty with which we can attribute low recorded species richness to sampling effort is influenced by environmental heterogeneity. We are most certain about the QDS that are recorded in black, since they have environmental heterogeneity values that are higher than for the squares with the highest species richness value in a given climate cell. If sampling effort were consistent throughout, then we would expect these QDS to have a higher species richness value due to higher environmental heterogeneity. In contrast, the QDS that are shaded in grey have environmental heterogeneity values that are lower than the QDS with the highest species richness value in a given climate cell. Lower environmental heterogeneity could in part explain the low recorded species richness values in these QDS. However, using conservative species richness thresholds (10% or 20% of the maximum value) results in selecting those QDS that are much lower than the maximum value recorded and the chance that these low values are due to environmental heterogeneity alone will be minimized.

Figure 6 shows those QDS with recorded species richness values of less than 20% of the maximum species richness value recorded in each of the climate cells. Seventy-seven per cent of the total number of QDS in the FSA region were identified as being associated with low species richness using this threshold. Using a more conservative threshold of 10% identified 63% of the QDS in this region as having low species richness (data not shown). Most of the QDS that are poorly sampled are in Botswana, Namibia and in the semi-arid to arid central region of South Africa. Interestingly, the areas indicated as adequately sampled include the fynbos and succulent karoo biomes, possibly reflecting the long-standing and ongoing research interest of botanists in these vegetation types.

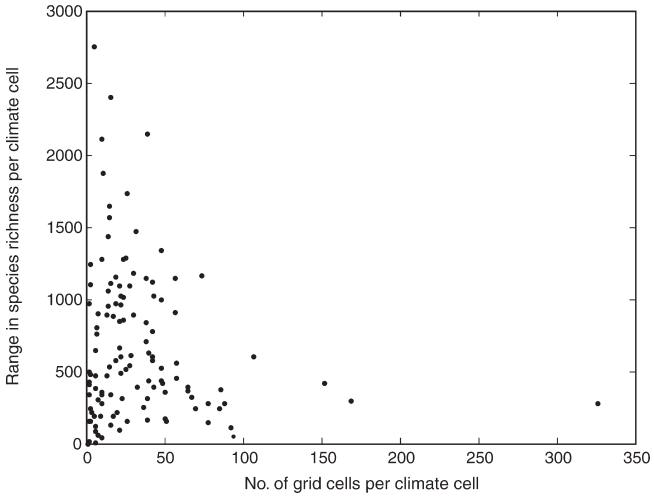

Fig. 5. The range of the species richness values per climate cell plotted against the number of QDS within each climate cell.

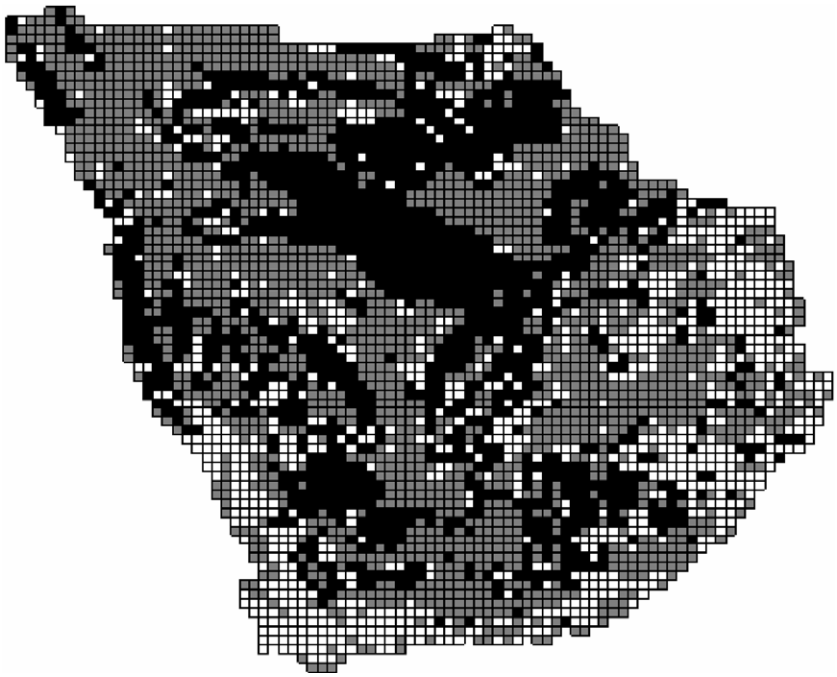

Fig. 6. QDS from the FSA region in grey and black that have been identified as having low species richness due to poor sampling effort. The QDS shaded in grey have recorded species richness values that are less than the species richness threshold and have environmental heterogeneity values below the environmental heterogeneity threshold. The QDS shaded in black have recorded species richness values that are less than the species richness threshold and have environmental heterogeneity values above the environmental heterogeneity threshold. The species richness threshold was calculated as 20% of the maximum value recorded in the climate cells. The environmental heterogeneity threshold was that environmental heterogeneity value calculated for the QDS with the maximum recorded species richness.

Identifying under-collected areas: Winter rainfall region

QDS identified as belonging to the WRR are shown in Fig. 7. These squares fall into relatively few of the climate cells for the entire FSA region, and share climate cells with QDS from the SRR. The WRR, and particularly the Cape Floristic Region (CFR) of which it is a part, is environmentally heterogeneous, which may influence species richness. Furthermore, Linder[53] found total precipitation to be the best predictor of species richness in the southwestern part of the CFR for several plant taxa. We analysed only those QDS from the winter rainfall region and identified those that were suspected to have been poorly sampled, using a 20% species richness threshold and the environmental heterogeneity threshold outlined above (Fig. 8). A similar pattern emerges from the analysis of the WRR QDS compared with the entire FSA region (Figs 6 and 8). The only difference appears to be in the environmental heterogeneity values, suggesting that the QDS from the SRR that fall into climate cells from the WRR do not change the value of species threshold, nor do they meet the criteria for being poorly sampled. Rainfall seasonality thus does not affect the application of our method across the region as a whole.

Strengths and weaknesses of the method

A major advantage of the technique that we propose is that the

threshold value chosen to indicate a lack of collection effort is specific to each climate cell. This incorporates the notion that different climates may support different numbers of species. This is more realistic than applying a single threshold (e.g. 20 species) across the entire map region in an attempt to identify poorly sampled QDS. The limitations and caveats to this approach are discussed below.

Climate cell species richness

The implicit assumption of the technique is that at least some of the QDS within the climate cell have been adequately sampled and that, as a consequence, these have higher recorded richness values than those inadequately sampled. Wide ranges in species richness within a climate cell (as indicated in Fig. 3) are of interest from a conservation viewpoint because this may indicate insufficient sampling effort, although environmental heterogeneity is also likely to play a role. We assume that at least some of the QDS have been adequately sampled, accounting for the wide range of richness values. However, this is not necessarily true as QDS with existing records of 'high' species richness could actually contain many more species than have been recorded. Similarly, small ranges in species richness do not necessarily indicate that the QDS within that climate cell have been well sampled. A small range may result because none of the squares within the climate cell has been adequately sampled.

Climate cell size

The conditions within a climate cell are considered to be sufficiently similar such that QDS within the climate cell will support similar numbers of species. However, we have not investigated just how big the climate cells should be. Clearly, different climate cell sizes will influence the range of conditions experienced within each one, and as the size of the climate cell increases, so the range of conditions within the cell widens. Larger climate cells will contain more QDS than smaller ones. For the technique to work, there should preferably be several QDS within a climate cell, so that there are several species richness values available with which to make comparisons. The implications of climate cell size require further investigation.

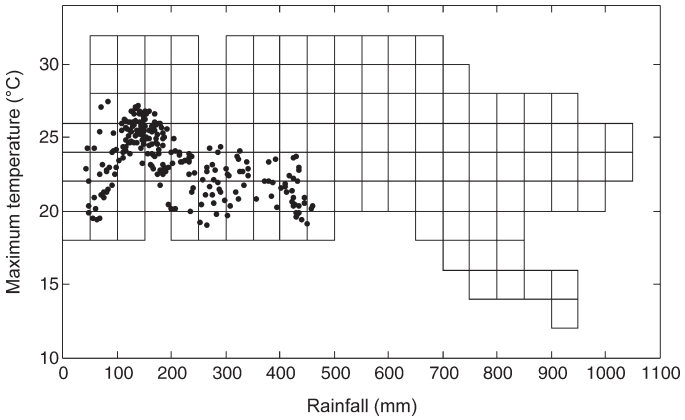

Fig. 7. QDS in the winter rainfall region plotted as points in a two-dimensional climate space that is defined by the annual mean daily maximum temperature (*y*-axis) and annual rainfall (*x*-axis). The climate space is divided into a number climate cells defined by 2°C temperature increments and 50-mm rainfall increments.

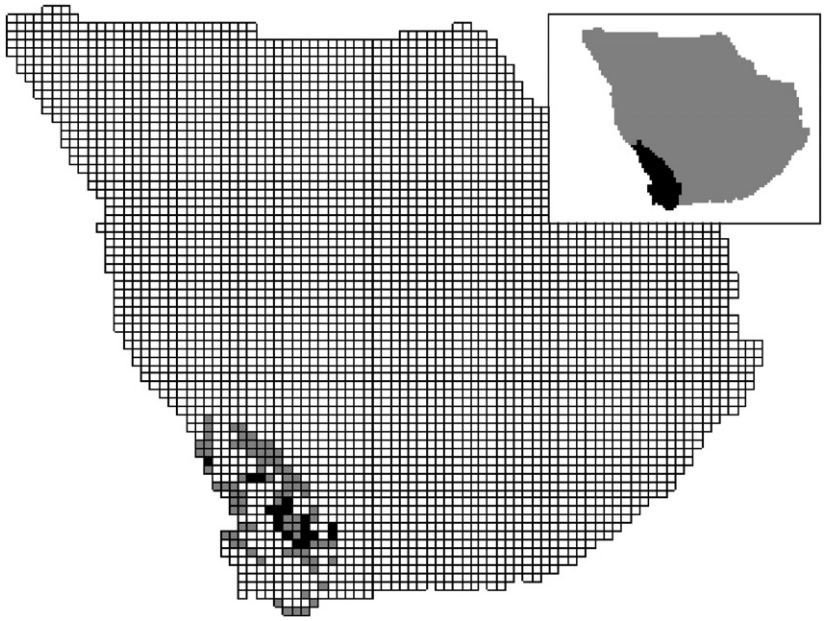

Fig. 8. QDS from the winter rainfall region in grey and black have been identified as having low species richness due to poor sampling effort. The grey QDS have recorded species richness values that are less than the species richness threshold and have environmental heterogeneity values below the environmental heterogeneity threshold. The black QDS have recorded species richness values that are less than the species richness threshold and have environmental heterogeneity values above the environmental heterogeneity threshold. The species richness threshold was calculated as 20% of the maximum value recorded in the climate cells. The environmental heterogeneity threshold was that value calculated for the QDS with the maximum recorded species richness. The inset indicates the QDS in the winter rainfall region in black.

Conclusion

We have proposed a technique for identifying quarter degree squares that are likely to record low species richness due to insufficient sampling effort rather than as a result of low actual species richness. This technique accounts for differences in species richness that are due to variation in climate across the map region and is superior to using a single threshold that is applied across the region. Although there are some theoretical limitations and technical issues that require additional exploration, this technique has considerable potential for users of collections data to assess its quality. This is particularly important, since there is an increasing need to use existing collections data, despite their limitations, for biodiversity planning.[12]

Using data on plant species richness from the PRECIS database, we found that most (>60%) QDS in the FSA region can be considered to be poorly sampled based on a conservative threshold of 10% of the maximum recorded species richness value in each climate cell. Further research is needed to make recommendations on the appropriate value for the threshold; our initial findings suggest, however, that we are arrogant to assume that southern Africa's plant diversity is adequately collected and represented in the national herbarium, and that even in a supposedly well-studied group, the land plants, there is indeed a collection impediment that accompanies the taxonomic impediment faced by the region's biodiversity scientists. This has serious implications for the effective monitoring, management and conservation of the region's biodiversity.

We thank the South African National Biodiversity Institute for the use of data from the National Herbarium, Pretoria (PRE) Computerised Information System (PRECIS), and the National Research Foundation and Rhodes University Joint Research Council for financial support. We also thank Serban Procheş and Laco Mucina for discussions and ideas. Susi Vetter and Berndt van Rensburg made helpful comments on an earlier draft of this article, and Anjanette Haller-Barker assisted in its preparation. Anonymous reviewers gave valuable comments that helped to improve the paper.

Received 10 June. Accepted 24 September 2005.

1. Smith G.F. and Wolfson M. M. (2004). Mainstreaming biodiversity: the role of taxonomy in bioregional planning activities in South Africa. *Taxon* **53**, 467–468.
2. Huntley B.J., Willis C.K., Smith G.F. and Siebert S.J. (2002). SABONET: its history and success in southern Africa. In *Rebirth of Science in Africa,* eds H. Baijnath and Y. Singh, pp. 231–246. Umdaus Press, Pretoria.
3. King N. (2002). Overcoming the taxonomic impediment to sustainable development: BioNET-INTERNATIONAL, the global network for taxonomy. In *Rebirth of Science in Africa,* eds H. Baijnath and Y. Singh, pp. 222–230. Umdaus Press, Pretoria.
4. Siebert S.J. and Smith G.F. (2004). Lessons learned from the SABONET project while building capacity to document the botanical diversity of southern Africa. *Taxon* **53**, 119–126.
5. South African Biosystematics Initiative. Online: http://www.nrf.ac.za/focusareas/conserve/sabi.stm.
6. Klopper R.R., Smith G.F. and van Rooy J. (2002). The biodiversity of Africa. In *Rebirth of Science in Africa,* eds H. Baijnath and Y. Singh, pp. 60–86. Umdaus Press, Pretoria.
7. Gibbons M.J. *et al.* (1999). The taxonomic richness of South Africa's marine fauna: a crisis at hand. *S. Afr. J. Sci.* **95**, 8–12.
8. Taylor P.J. and Hamer M. (2001). A report on the South African zoological collections audit — 1998. *S. Afr. Mus. Ass. Bull.* **25**, 5–20.
9. Bisby F.A. (2000). The quiet revolution: biodiversity informatics and the internet. *Science* **289**, 2309–2312.
10. Gioia P. and Pigott J.P. (2000). Biodiversity assessment: a case study in predicting richness from the potential of plant species in the forests of south-western Australia. *J. Biogeogr.* **27**, 1065–1078.
11. Soberón J., and Peterson A.T. (2004). Biodiversity informatics: managing and applying primary biodiversity data. *Phil. Trans. R. Soc. B* **359**, 689–698.
12. Graham C.H., Ferrier S., Huettman F., Moritz C. and Peterson A.T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.* 19, 497–503.
13. South African Biodiversity Information Facility. Online: http://www.sabif.ac.za
14. Funk V.A. and Richardson K.S. (2002). Systematic data in biodiversity studies: Use it or lose it. *Syst. Biol.* 51, 303–316.
15. Zaniewski A.E., Lehmann A. and Overton J.M. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecol. Model.* **157**, 261–280.
16. Rebelo A.G. (1994). Iterative selection procedures: centres of endemism and optimal placement of reserves. *Strelitzia* **1**, 231–257.
17. Loiselle B.A., Howell C.A., Graham C.H., Goerck J.M., Brooks T., Smith K.G. and Williams P.H. (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Cons. Biol.* **17**, 1591–1600.
18. Ortega-Huerta M.A. and Peterson A.T. (2004). Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico. *Div. Distrib.* **10**, 39–54.
19. Wilson K.A., Westphal M.I., Possingham H.P. and Elith J. (2005). Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biol. Conserv.* **122**, 99–112.
20. Stockwell D. and Peterson A.T. (2003). Comparison of resolution of methods used in mapping biodiversity patterns from point-occurrence data. *Ecol. Indicators* **3**, 213–221.
21. Gibbs Russell G.E., Retief E. and Smook L. (1984). Intensity of plant collecting in southern Africa. *Bothalia* **15**, 131–138.
22. Margules C.R. and Pressey R.L. (2000). Systematic conservation planning. *Nature* **405**, 243–253.
23. Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B and Kent J. (2000). Biodiversity hotspots for conservation priorities. *Nature* **405**, 853–858.
24. Stockwell D.R.B. and Peterson A.T. (2002). Controlling bias in biodiversity data. In *Predicting Species Occurrences: Issues of Scale and Accuracy,* eds J.M. Scott, P.J. Heglund and M.L. Morrison, pp. 537–546. Island Press, Washington, D.C.
25. Reddy S. and Dávalos L.M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *J. Biogeogr.* **30**, 1719–1727.
26. Edwards D. and Leistner O.A. (1971). A degree reference system for citing biological specimens in South Africa. *Mitt. Bot. Staatssamml. München* **10**, 501–509.
27. Pausas J.G. and Austin M.P. (2001). Patterns of plant species richness in relation to different environments: an appraisal. *J. Veg. Sci.* **12**, 153–166.
28. Francis A.P. and Currie D.J. (2003). A globally consistent richness — climate relationship for angiosperms. *Am. Nat.* **161**, 523–536.
29. Funk V.A., Richardson K.S. and Ferrier S. (2005). Survey-gap analysis in expeditionary research: where do we go from here? *Biol. J. Linn. Soc.* **85**, 549–567.
30. Acevedo D.H. and Currie D.J. (2003). Does climate determine broad-scale patterns of species richness? A test of the causal link by natural experiment. *Global Ecol. Biogeogr.* **12**, 461–473.
31. O'Brien E.M. (1998). Water–energy dynamics, climate, and prediction of woody plant species richness: an interim general model. *J. Biogeogr.* **25**, 379–398.
32. Wohlgemuth T. (1998). Modelling floristic species richness on a regional scale: a case study in Switzerland. *Biodiv. Conserv.* **7**, 159–177.
33. Ricklefs R.E. (1987). Community diversity: relative roles of local and regional processes. *Science* **235**, 161–171.
34. Harner R.F. and Harper K.T. (1976). The role of area, heterogeneity, and favourability in plant species diversity of pinyon–juniper ecosystems. *Ecology* **57**, 1254–1263.
35. Ricklefs R.E. (1977). Environmental heterogeneity and plant diversity: a hypothesis. *Am. Nat.* **111**, 376–381.
36. Rey Benyas J.M. and Schneider S.M. (2002). Plant diversity, biogeography and environment in Iberia: patterns and possible causal factors. *J. Veg. Sci.* **13**, 245–258.
37. Vivian-Smith G. (1997). Microtopographic heterogeneity and floristic diversity in experimental wetland communities. *J. Ecol.* **85**, 71–82.
38. Pausas J.G. Carreras J, Ferré A. and Font X. (2003). Coarse-scale plant species richness in relation to environmental heterogeneity. *J. Veg. Sci.* **14**, 661–668.
39. Morris J.W. (1980). Encoding the National Herbarium (PRE) for computerised information retrieval. *Bothalia* **13**, 149–160.
40. Morris J.W. and Manders R. (1981). Information available within the PRECIS data banks of the National Herbarium, Pretoria, with examples of uses to which it may be put. *Bothalia* **13**, 473–485.
41. Magill R.E., Gibbs Russell G.E., Morris, J.W. and Gonsalves, P. (1983). PRECIS, the Botanical Research Institute herbarium data bank. *Bothalia* **14**, 481–495.
42. Gibbs Russell, G.E. and Gonsalves, P. (1984). PRECIS — a curatorial and biogeographic system. In *Databases in Systematics,* eds R. Alkin and F. Bisby, pp. 137–153. Academic Press, London.
43. Gibbs Russell G.E. and Arnold T.H. (1989). Fifteen years with the computer: an assessment of the 'PRECIS' taxonomic system. *Taxon* **38**, 178–195.
44. Gibbs Russell G.E. (1985). Analysis of the size and composition of the southern African flora. *Bothalia* **15**, 613–629.
45. Gibbs Russell G.E. (1987). Preliminary floristic analysis of the major biomes in southern Africa. *Bothalia* **17**, 213–227.
46. Cowling R.M., Gibbs Russell G.E., Hoffman M.T. and Hilton-Taylor, C. (1989). Patterns of plant species diversity in southern Africa. In *Biotic Diversity in Southern Africa,* ed. B. J. Huntley, pp 19–50. Oxford University Press, Cape Town.
47. Gibbs Russell G.E. (1988). Distribution of subfamilies and tribes of Poaceae in southern Africa. *Monogr. Syst. Bot. Miss. Gard.* **25**, 555–566.
48. Gibbs Russell G.E., Watson L., Smook L., Koekemoer M., Barker N.P., Anderson H.M. and Dallwitz M.J. (1990). The Grasses of Southern Africa. *Mem. Bot. Surv. S. Afr.* No. 58.
49. Hutchinson M.F., Nix H.A., McMahon J.P. and Ord K.D. (1995). A topographic and climatic database. Centre for Resource and Environmental Studies, The Australian National University, Canberra.
50. WORLDCLIM. Online: http://biogeo.berkeley.edu/worldclim/worldclim.htm
51. The MathWorks (2000). MatLab version 7. The Mathworks, Natick.
52. Richerson P.J. and Lum K. (1980). Patterns of plant species diversity in California: relation to weather and topography. *Am. Nat.* **116**, 504–536.
53. Linder H.P. (1991). Environmental correlates of patterns of species richness in the south-western Cape Province of South Africa. *J. Biogeogr.* **18**, 509–518.