The smooth is better than the rough: An exploitation of reporting rate information in southern African bird atlas data

Francesca Little

Supervisor: Professor Les Underhill

Thesis presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Statistical Sciences

UNIVERSITY OF CAPE TOWN

November 2003

Acknowledgements

I would like to thank Professor Les Underhill for his excellent supervision, his encouragement and his enthusiastic guidance into the world of research.

Thank you to the Avian Demography Unit for providing the data and especially to Rene Navarro for his assistance with computing issues.

A word of thanks to the Department of Statistical Sciences, where I have received all my statistical training, for giving me the time this year to finish this project.

Thank you to David, Shaun and Tanya for your patient, loving support.

This research was partially funded by the National Research Foundation.

Table of Contents

The smooth is better than the rough: An exploitation of reporting rate information in southern African bird atlas data

Francesca Little

Abstract

The Southern African Bird Atlas Project (Harrison *et al.* 1997a, b) and the Mozambique Bird Atlas Project (Parker 1999) generated data on reporting rates for birds that takes into account the likelihood of species detection in a given area. Our main objective in this thesis is to explore methods for analysing and summarising reporting rate data.

The observed reporting rate data are subject to bias due to differential sampling effort and observer errors. We use a logistic regression model suitable for binomial type data to replace the observed reporting rates with smoothed probabilities of detection. To base our prediction on data from the surrounding neighbourhood, we choose as explanatory variables the north-south and west-east coordinates relative to the target grid cell for which the prediction is being made.

We explore some variants of our general smoothing approach that relates to the presentation of the smoothed distributions. The smoothed distributions of detection probabilities are presented as multicoloured maps. We consider two alternative ways of subdividing the range of detection probabilities into sub-intervals. One approach is species-specific, while the other imposes an absolute subdivision on all species distributions. For species with highly fragmented distributions, we introduce the possibility of using a weighted average between observed reporting rates and smoothed detection probabilities as the final value to be plotted. The weights are based on the extent of coverage and the underlying degree of fragmentation.

The identification of patterns of distributions for species is an important part of biogeography and plays a major role in the identification of areas where conservation efforts should be targeted. Interest centres around identifying areas of species richness, centres for narrow endemism and zones of transition in species composition. We explore the benefits of using a range of detection probabilities in comparison to the use of presence-absence data to identify areas rich in species and rich in narrow-endemic species. We transform existing measures for species richness and species endemism by replacing presence-absence data with detection probability deciles that reflect the relative likelihood of detecting a species in a given grid cell. The resulting measures give more weight to the areas where species have the core of their distributions and down-weight the peripheral edges of the species distributions, where detection probabilities may be too small to guarantee continued survival.

The use of a mathematical model to generate smoothed distributions of detection probabilities enables us to calculate gradients for the detection probability surfaces for species. We can define the concept of individual species gradients that reflect the relative degree of change among detection probabilities within the overall range of occurrence for the species. We combine the gradients for all species in several different ways. Large values for the overall sums of the gradients indicate areas of large fluctuation in species composition. On

the other hand, small values for the overall gradient sums indicate areas of relative stability. We also sum the gradients in one of 16 directions. These directional gradient sums distinguish between areas where the changes in species detection probability distributions occur in isolated directions, thus indicating ecological transition zones, and areas of random fluctuation, indicative of species fragmentation.

In this thesis we do not derive new statistical methods. We adapt existing techniques to deal with the abundance component of the data generated by the bird atlas projects. We show how the measures based on reporting rate data, rather than presence-absence data, add substantial insight into patterns of distribution of bird species in southern Africa.

References

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997a. The Atlas of Southern African Birds. Vol 1: Non-passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997b. The Atlas of Southern African Birds. Vol 2: Passerines. BirdLife South Africa, Johannesburg.

Parker V 1999. The Atlas of the Birds of Sul do Save, Southern Mozambique. Avian Demography Unit and Endangered Wildlife Trust, Cape Town and Johannesburg.

Chapter 1

Introduction

F. LITTLE

Avian Demography Unit, Department of Statistical Sciences, University of Cape Town, Rondebosch, 7701 South Africa

Background

The Atlas of Southern African Birds (Harrison *et al.* 1997) contains accounts of 932 species and presents distribution maps for 779 of these species on observed reporting rates in 4537 quarter-degree grid cells. It has been hailed by several authors (Lawes & Piper 1998; Van Jaarsveld *et al.* 1998) as providing a new benchmark for the mapping of ecological distributions. The database underlying these distribution maps provides invaluable information regarding avian biodiversity in southern Africa. The data collected for the atlas project differ from most previous mapping projects in that they provide information on reporting rates rather than just indicators of presence or absence. The database has provided a source of new data for ecologists which have been the basis for many papers on biogeographical distributions and conservation selection (see Gaston *et al.* 2001 and in press; Bonn *et al.* (2002); Rodrigues *et al.* 2001, 2002a, b and c). However, almost all these applications have essentially converted the reporting rate information into presence-absence information. Our main objective in this thesis is to explore methods for analysing and summarising these reporting rate data. These methods should be able to highlight the additional biogeographical information contained in these reporting rates as compared to presence-absence data.

The starting point of this thesis is that the concept of reporting rates, as measured by the Southern African Bird Atlas Project, is a valid one. Harrison & Underhill (1997) provided a comprehensive analysis of the strengths and weaknesses of the concept, and the arguments there will not be repeated here. Their conclusion, however, is quoted verbatim: "There have been studies which have used reporting rates from this atlas database and related them to independent quantitative measures of species' densities, and found there to be a consistent positive correlation; the fact that reporting rates increase in a monotonic manner with increasing density has been demonstrated convincingly (e.g. Du Plessis 1989; Bruderer & Bruderer 1993; Allan 1994; Robertson *et al.* 1995). However, the relationship is not such that a doubling of reporting rate indicates a doubling of bird density. Further investigation of how reporting rate varies with density is needed, but preliminary results suggest that it is sensitive to small changes in density when density is low, and insensitive when density is high. In other words, the relationship between reporting rate and density may be a logarithmic one; analyses by Robertson *et al.* (1995) support this." Griffioen (2001) analyzed an extensive data set of bird density data collected in Australia. He performed a series of analyses into the relationship between density and reporting rate, and concluded "there is indeed a robust functional relationship between reporting rate and species density" (Griffioen 2001, p70). He demonstrated that, for reporting rates below 50%, density and reporting rate have an "almost linear" relationship.

Objectives

The reporting rates are calculated from the number of checklists for a given grid cell for which a species was observed as present. As such they are subject to bias due to differential sampling effort and observer errors. These data can be substantially improved by a smoothing process that will replace the observed reporting rates with smoothed probabilities of detection which take into consideration information from neighbouring grid cells. In this thesis, we implement a method derived and validated by Erni (1998) for generating such smoothed probability distributions. We then investigate the use of these smoothed probabilities in the analysis of patterns of species distributions. We calculate measures of species richness and narrow endemism based on detection probabilities. We determine a measure of the rate of change among detection probabilities that leads to the identification of areas with a large degree of fluctuation in species distributions. As statisticians, we do not focus on biogeographical results. Instead we focus on developing methods to generate these smoothed detection probabilities, methods to summarise these probabilities so as to highlight distributional patterns, and methods for calculating rates of change. Though we do report some biogeographical results, the interpretation and use of these are largely left to the ecologists.

Layout of Thesis and Contribution of Co-authors

Chapters 2 to 5 of this thesis are presented as independent papers with cross references to each other. Each chapter thus has its own specification of authors, references, tables, illustrations and appendices. There is inevitably some repetition between the chapters.

In Chapter 2 we discuss the smoothing of bird atlas distributions. Erni (1998), in her M.Sc thesis, developed several mathematical methods for replacing the distributions based on observed rates with smoothed probability distributions that integrate the information from the neighbouring grid cells and thus provide a more accurate representation of the true probability of detection for bird species. She used the reporting rate data from the atlas project south of 27°S for a selected number of species. She evaluated the methods developed and recommended one method as the most promising. We use this method to generate smoothed distributions for all 932 species in southern Africa. There is more discussion of the choice of method in Chapter 6. Our study area includes the regions south of the Kunene and Zambezi Rivers, as well as including the Tete Province of Mozambique, and combines data from the Southern African Bird Atlas Project (Harrison *et al.* 1997a, b) and the Mozambique Bird Atlas Project (Parker 1999). Though the mathematical model was derived and validated by Erni (1998), we re-programmed the algorithm and generated smoothed detection probability distributions from scratch, obtaining identical results. Her contribution is acknowledged by making Dr Birgit Erni the senior author of this chapter. In addition, we acknowledge useful comments received from Professor Bruno Bruderer, Swiss Ornithological Institute, Switzerland. The smoothed detection probability distributions generated by this smoothing algorithm form the basis for the analyses presented in the rest of the thesis. For this reason, we feel it is important to include it in the thesis in detail.

Chapter 3 looks at some extensions to the smoothing algorithm. We examine different ways of mapping the smoothed detection probability distributions. Then we discuss special cases for which the smoothing model may be less appropriate and we present simple modifications to deal with these cases. We once again acknowledge the contribution of Dr Birgit Erni and comments from Professor Bruno Bruderer.

Chapters 2 and 3 are introductory and mainly provide the smoothed data on which the rest of the thesis is based. The major new research is presented in Chapters 4 and 5. This research was done by Francesca Little under the supervision of Professor Les Underhill. His contribution is acknowledged by making him co-author of these chapters. Chapter 4 looks at the impact of using reporting rate data, as opposed to presence-absence data, on the analysis of patterns of distributions for bird species. We develop measures of species richness and species endemism based on reporting rate data. We use the smoothed detection probabilities generated by the method discussed in Chapter 2 as input to these measures. Many other methods of smoothing the data could have been devised, and most of these would be amenable as input to Chapters 4 and 5. In Chapter 5 we focus on estimating rates of change in species distributions. This chapter has two thrusts. Firstly, we derive a method for calculating these rates of change (or gradients) directly from the model that generates the smoothed probabilities. Secondly, we experiment with methods which combine the gradients for sets of species. Although we believe that this section of the paper represents substantial conceptual progress, we think that the challenge of devising methods to visualize the combined multi-species gradients over large areas is a direction in which this research can be extended. In Chapter 6 we bring together these different research topics to form an integrated methodology for the analysis of patterns of distribution and of change.

Computing

The computing component of this research has been intensive. The combined databases from the Southern African Bird Atlas Project (Harrison *et al.* 1997a, b) and the Mozambique Bird Atlas Project (Parker 1999) consisted of the binomial type data from which the reporting rates were computed for 932 species in 5057 grid cells. To replace the observed reporting rates with smoothed detection probabilities required 4.7 million (932×5057) regressions. This was repeated three times for different forms of the mathematical model. These estimating algorithms simultaneously generated the information on rates of change required in Chapter 5. Then we summarised these species-specific probabilities and rates of change in a number of different ways to generate the various summaries of distributional patterns and patterns of change. This computer programming was done using Microsoft Fortran (1994-95). Microsoft Excel (1985-1997) was used to generate summary and analysis tables in Chapter 4. Statistica (2000) was used to graph the surfaces for selected grid cells in Chapter 5. The distributions were mapped using Arcview (1992-1999). The thesis was typeset in TEX(Knuth 1986).

References

Allan DG 1994. The abundance and movements of Ludwig's Bustard. Ostrich 65: 95-105.

Arcview GIS 3.2 1992-1999. Environmental Systems Research Institute.

Bonn A, Rodrigues ASL, Gaston KJ 2002. Threatened and endemic species: are they good indicators of patterns of biodiversity? Ecology Letters 5: 733-741.

Bruderer B, Bruderer H 1993. Distribution and habitat preference of Redbacked Shrikes *Lanius collurio* in southern Africa. Ostrich 64: 141-147.

Du Plessis MA 1989. The influence of roost-cavity availability on flock size in the Redbilled Woodhoopoe *Phoeniculus purpureus*. Ostrich Supplement 14: 97-104.

Erni B 1998. Analysis of Distribution Maps from the Bird Atlas Data. Unpublished M.Sc thesis, University of Cape Town.

Gaston KJ, Rodrigues ASL (in press). Reserve selection in regions with poor biological data. Conservation Biology.

Gaston KJ, Rodrigues ASL, Van Rensburg BJ, Koleff P, Chown SL 2001. Complementary representation and zones of ecological transition. Ecology Letters 4: 4-9.

Griffioen P 2001. Temporal Changes in the Distributions of Bird Species in Eastern Australia. Unpublished Ph.D thesis, La Trobe University, Bundoora, Victoria, Australia.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997a. The Atlas of Southern African Birds. Vol 1: Non-passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997b. The Atlas of Southern African Birds. Vol 2: Passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Underhill LG 1997. Introduction and methods. In: Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds). The Atlas of Southern African Birds. Vol 1: Non-passerines, xliii-lxiv. BirdLife South Africa, Johannesburg.

Knuth DE 1986.The TEXbook. Computers & Typesetting. Addison Wesley, Reading Massachusetts.

Lawes MJ, Piper SE 1998. There is less to binary maps than meets the eye: The use of species distribution data in the southern African sub-region. South African Journal of Science 94: 207-210.

Microsoft Developer Studio Fortran Powerstation 4.0 1994-95. Microsoft Corporation.

Microsoft Excel 97 SR-2 1985-1997. Microsoft Corporation.

Parker V 1999. The Atlas of the Birds of Sul do Save, Southern Mozambique. Avian Demography Unit and Endangered Wildlife Trust, Cape Town and Johannesburg.

Robertson A, Simmons RE, Jarvis AM, Brown CJ 1995. Can bird atlas data be used to estimate population size? A case study using Namibian endemics. Biological Conservation 71: 87-95.

Rodrigues ASL, Gaston KJ 2002a. Optimisation in reserve selection procedures - why not? Biological Conservation 107: 123-129.

Rodrigues ASL, Gaston KJ 2002b. Rarity and conservation planning across geopolitical units. Conservation Biology 16: 674-682.

Rodrigues ASL, Gaston KJ 2002c. Maximising phylogenetic diversity in the selection of networks of conservation areas. Biological Conservation 105: 103-111.

Rodrigues ASL, Gaston KJ 2001. How large do reserve networks need to be? Ecology Letters 4: 602-609.

STATISTICA for Windows Release 5.5 2000 Statsoft.

Van Jaarsveld AS, Gaston KJ, Chown SL, Freitag S 1998. Throwing biodiversity out with the binary data? South African Journal of Science 94: 210-214.

Chapter 2

Smoothing of Bird Atlas Distribution Maps using Logistic Regression on a Local Scale

B. ERNI[1], F. LITTLE[1], L.G. UNDERHILL[1] & B. BRUDERER[2]

1 Avian Demography Unit, Department of Statistical Sciences, University of Cape Town, Rondebosch, 7701 South Africa

2 Swiss Ornithological Institute, CH-6204 Sempach, Switzerland

Introduction

The Atlas of Southern African Birds (Harrison *et al.* 1997a, b) contains accounts of 932 species and shows distribution maps for 779 of these based on the observed reporting rates in 4537 quarter-degree grid cells. The observed reporting rate for a species in a given grid cell corresponds to the proportion of checklists on which this species was recorded. The observed reporting rates may be considered as a measure of the average detection probability, the likelihood of encountering the species in a grid cell. They provide estimates of the true probabilities of species occurrence. The reporting rates compound three factors: relative abundance, conspicuousness and sampling error. If, within a given species conspicuousness is assumed to be uniform and sampling error to be zero, variation in reporting rates between grid cells is a measure of the relative abundance of the species. However, since relative abundance is just one of the factors that influences reporting rates, reporting rates at best provide an index which fluctuates with changes in density. For a discussion of reporting rates and their limitations see Underhill *et al.* (1992), Harrison & Underhill (1997) and Griffioen (2001). They refer to the following facts. It has been shown how reporting rates fluctuate in accordance with known migratory patterns. The variation of reporting rates over geographical space conforms to the expected or known changes in densities. The distribution of reporting rates based on vegetation biomes also usually agree with the known habitat preferences of species. Several studies have tried to relate reporting rates from the southern African atlas to independent quantitative measures of density. All of them have found a consistent positive correlation and many were able to illustrate a strong monotonic relationship (see for example Du Plessis 1989; Bruderer & Bruderer 1993; Allan 1994; Robertson *et al.* 1995). In addition Griffioen (2001) used data from the Australian Bird Count database to investigate the relationship between reporting rate data and species densities. He was able to show a strong monotonic relationship between reporting rates and abundance for a large number of species. As an even stronger result, he demonstrated that, for reporting rates less than 50%, this relationship is almost linear.

In the published atlas, the observed reporting rates were transformed into four categories, with a different intensity of shading for each. For each species, the accuracy of the reporting rate statistic in a grid cell as an estimate of its "true" reporting rate depends primarily on the number of checklists available for a grid cell. This varied between zero and 2297 (Fig. 1). Within this data set, there was therefore a large degree of variability in the accuracy of the observed reporting rates which results in distribution maps which have a "rough" texture, with neighbouring grid cells in areas of uniform habitat showing different reporting rate categories, especially in areas where the numbers of checklists are small. On the other hand, "rough" texture can arise in areas with large numbers of checklists, even with uniform habitats, because the large numbers of checklists in these grid cells were frequently contributed by a single observer. Smoothing would reduce this sampling variation by rendering the single-cell reporting rates more consistent with the surrounding information. In addition, it would provide interpolated values for the grid cells for which no checklists was collected and improved values for the cells for which few checklists were available. Measurement errors due to

observer effects would be reduced. Visual perceptibility of the maps would be improved, and would become available for practical application elsewhere: e.g., the maps should enable biogeographers and ecologists to understand patterns of bird distribution more precisely; used in field guides, they should provide birdwatchers with immediate information on the detection probability of a species in a given area.

The underpinning assumption of smoothing is the continuity of distributions in neighbouring grid cells. The aim of this study was to design a statistical tool to produce smoothed maps which integrate the information from the neighbouring grid cells, but avoid as far as possible blurring real patchiness. The degree of adjustment had to take into account the accuracy of the information in the surrounding grid cells, giving more weight to those with large numbers of checklists. The smoothing algorithm was required to be rapid, because of the large area covered and large number of species.

Material and Methods

The Atlas Data

The data used in this paper come from two sources: the database of the Southern African Bird Atlas Project, upon which Harrison *et al.* (1997a, b) was based, provided the data for Botswana, Lesotho, Namibia, South Africa, Swaziland and Zimbabwe; the database for the Mozambique Bird Atlas Project provided the data for southern Mozambique (south of the Save River) (Parker 1999) and for central Mozambique (Sofala, Manica and Tete Provinces) (V. Parker unpubl. data). Southern Africa, as defined for this thesis, is thus the region south of the Kunene and Zambezi Rivers, but including the Tete Province of Mozambique (Fig. 2). The taxonomy used in this chapter (and elsewhere in this thesis) is that used in Harrison *et al.* (1997a, b).

For most of the study area, data were collected for quarter-degree grid cells, which were approximately 27 km × 25 km in the north and 27 km × 23 km in the south of the area with shrinkage being longitudinal. The exception was in Botswana, where half-degree grid cells were used. For each of the grid cells of southern Africa, checklists of species were collected. On each of these checklists, each species was either recorded as present or absent. The observed response is the proportion of total checklists on which the species was recorded as present — the reporting rate. Harrison & Underhill (1997) provided further details of the database. They pointed out that the variation in area of grid cells has no impact on the reporting rate.

To include the Botswana data in our modelling, we subdivided each of the half-degree grid cells into four quarter degree cells and assigned the half-degree rates to each of the four quarter-degree grid cells. This is a satisfactory approach for dealing with scale variation in this case, because Botswana is characterised by relatively uniform environmental gradients (Harrison & Underhill 1997). The total number of quarter-degree grid cells generated in this way was 5057. Checklists were available for 4918 (97.3%) of them. Grid cells with no checklists were left blank in Fig. 1.

In this Chapter we shall use the following notation:

- n_i = the total number of checklists that have been collected for grid cell i
- S_i = the number of checklists on which a species was recorded
- r_i = the reporting rate, the proportion of checklists recording the species, S_i/n_i
- π_i = the true, but unknown, detection probability of the species in grid cell i.

General Smoothing Approach

We develop a smoothing approach based on regression concepts. This aims to "predict" the reporting rate for a target grid cell i from the observed reporting rate at the target cell and a small set of surrounding

cells, by fitting a smoothed "surface" of reporting rates to a local area around the target cell. Environmental factors that influence the suitability of a habitat for many bird species change more or less gradually so that nearby grid cells can be expected to have more similar reporting rates than cells that are far apart. Hence the reporting rates are not independent but spatially autocorrelated. Continuity in occurrence between neighbouring grid cells was, in fact, demonstrated by Erni (1998) for a large suite of species occurring in southern Africa.

This continuity can be modelled if it is incorporated into the explanatory variables of the regression equation. Regression techniques aim to model the response on a set of explanatory variables x_1, x_2,...,x_p. We assume that the reporting rates follow approximately the binomial distribution and with error variances dependent on π_i and n_i (Underhill *et al.* 1992). For such data we can use the linear logistic model for each target grid cell i (Cox & Snell 1989) which relates the logistic transform of π_i linearly to the explanatory variables, viz.,

$$\ln \frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p. \quad (1)$$

Each grid cell within the atlas area becomes the target cell and we use the predicted reporting rate for the target grid as smoothed value. In the case of southern Africa, this means that the algorithm needs to be applied 5057 times for each species.

However, because we know that the process that generated the observed reporting rates is subject to various biases, we are reluctant to use this generalized linear model approach at any deaper level than as a descriptive tool that meets the need to take the varying number of checklists per grid cell into account in an appropriate way. We therefore do not use the inferential statistical theory that goes along with the GLM, because we feel that the caveats associated with reporting rates do not warrant it.

The Explanatory Variables and Regression Model

To base our prediction on data from surrounding grid cells we choose as explanatory variables the north-south and west-east coordinates relative to the target grid cell for which the prediction is being made, and which will have coordinates $(x, y) = (0, 0)$. The y-coordinate will represent the north-south axis, increasing from north to south. The x-coordinate represents the west-east axis and increases from west to east. For a block of 25 and nine grid cells these coordinates are shown in Table 1. The block of nine grid cells extends the range by one grid cell and the block of 25 grid cells extends the range by two grid cells. Adding an interaction term, xy, enables the model to fit a curved surface of reporting rates, and not merely a plane, to the data. For a prediction based on nine cells we have chosen the model

$$\ln \frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy. \quad (2)$$

The four regression parameters β_0, β_1, β_2 and β_3 need to be estimated from nine observed rates and the addition of further terms to the model would therefore be ill-advised.

The reporting surface for the block of nine grid cells then has the following possibilities:

- It can increase towards the east if β_1 is positive, and vice versa.
- It can increase towards the south if β_2 is positive, and vice versa.
- It can increase from the north-west to the south-east if both β_1 and β_2 are positive (with other combinations of parameters possible).
- If, in addition, β_3 has a non-zero value, curvature is added to this plane.

For 25-cell blocks, terms x^2 and y^2 can be added to the model,

$$\ln \frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy + \beta_4 x^2 + \beta_5 y^2, \quad (3)$$

making the surface more general. In this case, six parameters are estimated from 25 observed reporting rates. The algorithm for estimating the regression parameters is described fully in the Appendix. In view of the large number of regressions that had to be carried out (5057 grid cells and 932 species), we needed a fast computational tool which would be easy to manipulate and which required minimal user interaction. The smoothing algorithm was implemented in Microsoft Fortran (1994-95). This allowed for programs that were able to run the entire analysis in one continuous process, from finding a block of nine or 25 grid cells, to predicting and saving values and carrying on to the next block of grid cells. Based on extensive simulation and experimentation, the number of iterations of the algorithm was fixed at five; see Erni (1998) for details. This was done for two reasons. Firstly, it eliminates the programming overheads needed to compute termination criteria. Secondly, the smoothed detection probabilities are going to be categorised into a small number of classes for mapping purposes, so do not need to be calculated with accuracy. All four terms were included in the nine grid cell models, regardless of whether the regression parameters were statistically significant or not.

Exemplar Species

Smoothed maps are presented for two species [1]. Namaqua Dove *Oena capensis* and Whiterumped Swift *Apus caffer* were selected because both species are widely distributed across southern Africa (Fig. 2 shows areas mentioned in the text) and because their distributions include areas for which the number of available checklists per grid cell was small. In the standard field guides and handbooks for southern Africa, the generalized distribution of the Namaqua Dove is shown to cover southern Africa almost entirely, and that of the Whiterumped Swift covers this region apart from Botswana (e.g. Maclean *et al.* 1993). For mapping purposes, we divide the range of reporting rates into five categories and use five different colours to represent these sub-intervals. All rates in the bottom sub-interval of the distribution that are less than 2% are illustrated with a cross.

Results

The distribution map of the Namaqua Dove based on observed reporting rates (Fig. 3) shows a large degree of spatial autocorrelation in the areas of intensive atlas fieldwork (Fig. 1) in the southern and eastern regions of South Africa; the autocorrelation is displayed for this region because adjoining grid cells in this region mostly have the same shading or an adjacent shading (Fig. 3). The species is shown to be largely absent, except as a vagrant, from Lesotho, much of the Eastern Cape, especially the former Transkei, southern KwaZulu-Natal, most of Mozambique and the Namib Desert region of south-western Namibia. Between this area of absence and the core of the distribution in the interior of southern Africa is a broad area where the species occurs at low reporting rates, between 2% and 18.2% (the green area in Fig. 3). In contrast, in the Northern Cape, Namibia, Botswana and Zimbabwe, where the atlas fieldwork was less intensive (Fig. 1), the distribution map shows many neighbouring grid cells with contrasting reporting rates, with darker shadings frequently abutting the lighter. In the extreme case, those grid cells for which only one checklist is available, will either be represented by the darkest shade (the species was recorded present on this checklist) or as blank (the species was not recorded). This gives the impression of a random chess board, and leads to the conclusion that the species has a fragmented distribution. However, because the distribution in areas with many checklists shows strong autocorrelation of reporting rates, it seems more likely that the random chessboard effect in the north and west of the distribution of the Namaqua Dove is due to sampling variation rather than to genuine fragmentation of distribution.

[1] Further examples of smoothed distributions may be found in Chapters 3 and 5 of this thesis

The smoothed distribution for this species using the nine grid cell model confirms and enhances the overall observed patterns described above for the areas of intense fieldwork in the south and east of the region, but conveys a totally different impression in the remainder of southern Africa (Fig. 4). The smoothed distribution map shows a series of core areas having high reporting rates (> 61%); superficially, most of these do not appear to conform to geographical features. However, the "hole" in the distribution in southern Botswana is more clear cut; this is an area where there is no standing water, even from man-made sources, and the area of high reporting rates in Zimbabwe closely follow the Central Highlands, where most of agriculture in the country is located. These two observations probably provide the clue for understanding the complex distribution pattern; the patches of high reporting rates coincide with relatively arid areas where water is permanently available, either from natural sources such as rivers or springs, or from artificial sources, mostly associated with agriculture, such as bore holes and small reservoirs.

The smoothed distribution using the 25 grid cell model with four parameters further smoothes and simplifies the distribution (Fig. 5). This distribution map needs to be compared and contrasted with the map produced by the 25 grid cell model with six parameters, in which the terms x^2 and y^2 are included (see equation (3)) (Fig. 6). The inclusion of these two additional terms in the model generates a considerable amount of artefact into the distribution, and fails to achieve a smoothed map. The reason for this is that the additional terms enable the model to fit a surface with a complex shape to the observed reporting rates, leading to unrealistic estimates of the reporting rate for the target grid cell. This is, for example, evident in Mozambique, where a cluster of four grid cells immediately south-east of 18°S 34°E reports the presence of Namaqua Dove (Fig. 3), including two grid cells with a reporting rate between 18.2% and 33.3%; each of these four grid cells has at least three checklists (Fig. 1). In Fig. 6, these four grid cells generate a modelled distribution containing a grid cell in which the species is reported as occurring with a reporting rate exceeding 45.2%. The six-parameter model (Fig. 6) therefore fails to achieve the goal of smoothing for this species, and we revert to the four-parameter model (Fig. 5).

The core of the distribution, shown by the two darkest shades in the four-parameter model (reporting rates exceeding 44.5% in Fig. 5), is simplified to a ring around the waterless Kalahari sand sea of southern Botswana, the Central Highlands of Zimbabwe, and an area from the Etosha Pan westwards in northern Namibia. The immediate impression conveyed by both smoothed maps (Figs 4 and 5) is different from that of the observed data (Fig. 3). However, with hindsight, the basis for the patterning revealed in the smoothed maps is present in the original map. For example, the band of slightly greater reporting rates that stretches east-west across the central part of the region at about 22°S which is a conspicuous feature of Fig. 5 is present, though not obvious, in Fig. 3.

The Whiterumped Swift is an intra-African migrant which breeds in southern Africa during the austral summer. It is a widespread forager on aerial invertebrates (Maclean *et al.* 1993). Its distribution is probably less closely related to any particular natural habitat than most species, and therefore a fairly strong spatial autocorrelation can be predicted. One constraint to the breeding distribution is that Whiterumped Swifts are nest "predators", mainly taking over the nests of those species of swallow which build retort-shaped nests (Brooke 1997).

The distribution map based on observed reporting rates shows that the core of its southern African distribution lies in the south and east of the region, with scattered records across the Northern Cape, Namibia, Zimbabwe and central Mozambique, and an almost total absence from Botswana (Fig. 7). The smoothed distribution maps based on the nine grid cell model (Fig. 8) and the 25 grid cell model with four parameters (Fig. 9) are both strikingly different from the distribution map based on the observations. The patchy distribution in central Namibia is transformed, especially in Fig. 8, to a large area where Whiterumped Swifts may rarely occur, and a small area, close to Windhoek, with high reporting rates; this overall Namibian

distribution coincides closely with the composite of the distributions of four species of swallows which build retort-shaped nests usurped by Whiterumped Swifts; Redbreasted Swallow *Hirundo semirufa*, South African Cliff Swallow *H. spilodera*, Greater Striped Swallow *H. cucullata* and Lesser Striped Swallow *H. abyssinica* (Earle *et al.* 1997).

The 25 grid cell model with six parameters for the Whiterumped Swift produces a distribution map (Fig. 10) which is "rougher" in texture than that for the four-parameter model (Fig. 9). As was the case with the Namaqua Dove, a comparison of this map with the observed distribution (Fig. 7) shows that the additional flexibility of the six-parameter model creates artefacts which are not supported by the observed data. This is particularly evident in Namibia, where the thin scatter of observed records (Fig. 7) is more consistent with the smoothed distribution shown in Figs 8 and 9 than with the "bumpy" distribution of Fig. 10. Use of the six-parameter model is therefore not recommended.

The distribution map based on observed reporting rates shows the Whiterumped Swift as being largely absent from much of Lesotho (Fig. 7), but the intensity of fieldwork in this mountainous country where many grid cells could only be accessed with difficulty was low (Fig. 1). The smoothed distribution based on the nine grid cell model retains an area of absence in Lesotho (Fig. 8), but the distribution based on the 25 grid cell model shows low reporting rates throughout Lesotho (Fig. 9). Without extensive fieldwork, it is impossible to decide from the maps which of the representations of the distribution is closest to the truth. However, given the generally low intensity of fieldwork and the fact that Whiterumped Swifts were recorded from several grid cells in the interior of Lesotho, it seems logical to conclude that Whiterumped Swifts may be encountered almost anywhere in Lesotho, and that one of the smoothed maps is preferable to the observed distribution map. Biologically, this conclusion is strengthened by the knowledge that the Greater Striped Swallow, whose nests are regularly usurped by Whiterumped Swifts, occurs throughout Lesotho, at reporting rates that appear continuous with those in the neighbouring regions of South Africa (Earle *et al.* 1997).

Discussion

We have presented a statistical model for generating smoothed distribution maps for the reporting rates of bird species in southern Africa. The model makes use of information in surrounding grid cells and thus capitalises on the spatial autocorrelation inherent in the observed data from neighbouring grid cells. It removes the variation in observed distributions due to sampling error and reduces the random checkboard appearance of many of the distribution maps presented in Harrison *et al.* (1997a, b), especially in areas where the average number of checklists per grid cell was small. For example, we have illustrated how it provides improved maps in areas with sparse data collection. The smooth maps are visually simpler and show trends and characteristic features more clearly. The nine grid cell model is the best when local detail is important, while the 25 grid square model is appropriate when the broad brush picture is important.

The validation of the chosen smoothing model involves assessing the validity of the binomial assumption, evaluating the choice of the size of the blocks of grid cells used for each local regression, deciding on the number of parameters to include in the models and assessing the appropriate degree of smoothing. These checks were done by Erni (1998) in the initial derivation of the smoothing model and we summarize her conclusions. In general for the binomial distribution to be valid, observations should be independent with a constant probability of success. For the atlas data the first requirement is that for a given grid cell the observations from the different checkists and hence from different observers should be independent. This would generally be true except in the presence of informative communication between observers that would be more likely to have occurred for rare species. Secondly, the probability of detection of a species in a given grid cell should have remained constant over the period of data collection. Seasonal changes in conspicuousness and differential observer skill and effort will, however, have affected the probability of detection (Harrison & Underhill 1997). As a result the assumption of a constant success probability may not be valid. The resulting

effect on the logistic models is that of overdispersion which inflates the variances of the parameter estimates. Erni (1998) found that 50% of the overdispersion estimates were less than two. We are not interested in the significance of the individual regression parameters, but in the estimates of the probabilities of detection which are unaffected by the possible presence of overdispersion.

The choice of the size of the block of grid cells on which to base the local regressions depend on considerations involving spatial autocorrelation, overdispersion and what degree of interpolation is acceptable to biologists. In a nine grid cell block, all surrounding cells are equidistant from the central cell for which the prediction is being made. Thus no extra spatial autocorrelation function of distance has to be considered and the spatial autocorrelation is sufficiently captured by the spatial explanatory variables chosen. As we increase the blocks to 25 cells and beyond, this may no longer be the case. The extra 16 cells in the 25-cell blocks also add more variation to the estimates so that it may become necessary to include the extra explanatory variables. However, in general, these models lead to worse predicted probabilities of detection. These two problems of overdispersion and spatial autocorrelation becomes worse as the blocks on which the regressions are performed, are increased. Furthermore, the nine grid cell model will base predictions on a neighbourhood of c. 6 075 km^2 while the neighbourhood for the 25-grid cell model is as large as c. 16 875 km^2. Habitats can change considerably within 50 km so that the use of larger blocks may necessitate the inclusion of other variables in addition to the simple spatial relationships. Such modelling should be investigated. Finally, biologists were extremely reluctant to accept the use of larger areas than the 25-grid cell model on which to base smoothing, mainly because they were unwilling to contemplate the possibility of interpolating the occurrence of a species at larger distances. The concept of smoothing using a 7x7 grid cell block was biologically unacceptable to them.

The number of terms in the model is of less importance than the usual modelling process where the final model describes the relationships between the response and the explanatory variables in addition to generating a predicted response. In the case of smoothing we are not concerned with the model terms themselves but only in a single predicted value for the central grid cell from a model that fits the data well. The number of parameters is restricted by the limited amount of data available, a maximum of nine or 25 data points for each local regression. We can thus not add the x^2 and y^2 terms to the regressions based on nine grid cells. For the 25 grid cell models, the addition of these terms provides a better fit to the observed data but at the same time decreases the degree of smoothing. In fact we saw that these terms sometimes introduce artefacts into the modeled reporting rates (Fig. 7). The actual fit of the models to the observed data is related to the degree of smoothing that one wishes to achieve and hence traditional goodness-of-fit statistics are not appropriate. The appropriate degree of smoothing is dependent on the degree of fragmentation in the species distributions and the data quality. We address this issue in more detail in Chapter 3.

The way in which we choose to map the observed and smoothed distributions of reporting rates is an important factor in that it imposes a further degree of smoothing. We discussed the different distributions generated by the different smoothing models and showed how these smoothed distributions differ from the distribution of observed rates. This meant that the division of these distributions into five sub-ranges differed between the observed and smoothed distributions and differed from model to model. In addition, the Namaqua Dove is a more abundant species than the Whiterumped Swift with generally higher overall reporting rates. Hence the subdivision of the reporting rate ranges of these two species differed substantially from one another. These differences in the mapping scales between species and within species between models are illustrated in the legends of the different maps. As in Harrison *et al.* (1997a,b), the colour shades thus show relative rather than absolute reporting rates. We address this issue further in the next chapter.

We defer to Chapter 6 a discussion regarding the choice of our specific method of smoothing over alternative approaches. One of the considerations was to find a method that could cope with the scale of the problem, i.e., that of generating smoothed detection probabilities for 932 species in 5057 grid cells. The smoothing

algorithm is easy to compute and fast to implement. It can generate maps with accurate information for species with an acceptable degree of continuity in their distributions. These maps are visually simple and can form the basis for methods for reserve selection that take into account the probability of species occurrence in individual grid cells. The method can easily be applied to ecological distributions for species other than birds.

References

Allan DG 1994. The abundance and movements of Ludwig's Bustard. Ostrich 65: 95-105.

Brooke RK 1997. Family Apodiformes: swifts. In: The Atlas of Southern African Birds. Vol 1: Non-passerines. Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds), pp. 116-150, 764-765. BirdLife South Africa, Johannesburg.

Bruderer B, Bruderer H 1993. Distribution and habitat preference of Redbacked Shrikes *Lanius collurio* in southern Africa. Ostrich 64: 141-147.

Cox DR, Snell EJ 1989. Analysis of Binary Data, 2nd edn. Chapman & Hall, London.

Du Plessis MA 1989. The influence of roost-cavity availability on flock size in the Redbilled Woodhoopoe *Phoeniculus purpureus*. Ostrich Supplement 14: 97-104.

Earle RA, Spottiswoode CN, Allan DG, Herremans M, Brooke RK 1997. Family Hirundinidae: swallows and martins. In: The Atlas of Southern African Birds. Vol 1: Non-passerines. Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds), pp 116-150, 764-765. BirdLife South Africa, Johannesburg.

Erni B 1998. Analysis of Distribution Maps from the Bird Atlas Data. Unpublished M.Sc thesis, University of Cape Town.

Griffioen P 2001. Temporal Changes in the Distributions of Bird Species in Eastern Australia. Unpublished Ph.D thesis, La Trobe University, Bundoora, Victoria, Australia.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997a. The Atlas of Southern African Birds. Vol 1: Non-passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997b. The Atlas of Southern African Birds. Vol 2: Passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Underhill LG 1997. Introduction and methods. In: Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds). The Atlas of Southern African Birds. Vol 1: Non-passerines, xliii-lxiv. BirdLife South Africa, Johannesburg.

Maclean GL 1993. Robert's Birds of Southern Africa, 6th edn. The Trustees of the John Voelcker Bird Book Fund, Cape Town.

McCullagh P, Nelder JA 1989. Generalized Linear Models, 2nd edn. Chapman & Hall, London.

Microsoft Developer Studio Fortran Powerstation 4.0 1994-95. Microsoft Corporation.

Parker V 1999. The Atlas of the Birds of Sul do Save, Southern Mozambique. Avian Demography Unit and Endangered Wildlife Trust, Cape Town and Johannesburg.

Robertson A, Simmons RE, Jarvis AM, Brown CJ 1995. Can bird atlas data be used to estimate population size? A case study using Namibian endemics. Biological Conservation 71: 87-95.

Underhill LG, Prys-Jones RF, Harrison JA, Martinez P 1992. Seasonal patterns of occurrence of palaearctic migrants in southern Africa using atlas data. Ibis. 134, supplement 1: 99-108.

Appendix

The regression parameters can be estimated using an iterative process, called Iteratively Reweighted Least Squares Regression (IRWLS), (McCullagh & Nelder 1989). For the logistic model, rewriting the regression equations (2 and 3) in vector notation in the standard way, maximum likelihood theory gives estimates b for the regression parameters as :

$$b = (X^T X)^{-1} X^T W z \tag{4}$$

where

- X is the design matrix with elements the explanatory variables; in the case of the nine cell block

$$X = \begin{pmatrix} 1 & -1 & -1 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

- W is a diagonal matrix of weights, with diagonal elements,

$$w_{ii} = n_i \pi_i (1 - \pi_i) \tag{5}$$

- the z values are modified dependent variables,

$$z_i = \ln \frac{\pi_i}{1 - \pi_i} + \frac{S_i - n_i \pi_i}{n_i \pi_i (1 - \pi_i)} \tag{6}$$

Equations (4) to (6) form an iterative procedure. The w_{ii} and z_i in equations (5) and (6) must be recalculated after each iteration using the regression coefficient estimates from the previous iteration, obtained from equation (4). At each iteration the π_i values are updated using the inverse of either equation (2) or (3).

To start the iterative process, an initial set of weights and dependent variables are needed. We could use

$$z_i = \ln \frac{r_i}{1 - r_i} \tag{7}$$

and

$$w_{ii} = n_i r_i (1 - r_i). \tag{8}$$

However, observed reporting rates of "zero" and "one" would lead to zero corresponding weights and cause the dependent variables to be undefined because of either division by zero or taking the logarithm of zero. "Zero" or "one" reporting rates are often especially observed in areas where few checklists have been collected. This problem is resolved by using the empirical logistic transform (Cox & Snell 1989) as modified dependent variables for the first iteration. These are

$$z_i = \ln \frac{S_i + 0.5}{n_i - S_i + 0.5} \tag{9}$$

with corresponding weights

$$w_{ii} = \frac{n_i (S_i + 1)(n_i - S_i + 1)}{(n_i + 1)(n_i + 2)}. \tag{10}$$

In addition to solving the problem of taking the logarithm of "zero" and "one" reporting rates at the first iteration, this approach reduces the number of iterations until convergence. This is an important advantage since the regression has to be run over all 5057 grid cells of southern Africa and this process has to be repeated for every one of the 932 species. An artefact of the use of the Cox & Snell (1989) empirical logistic transformation is that the predicted reporting rates for the grid cells are strictly positive. We therefore chose to set all predicted reporting rates less than 0.005 to zero. This means that we assume a species to be absent if the predicted rate is so small that the species is expected to be recorded only once in more than 200 checklists.

Table 1: The x- and y- co-ordinates for blocks of 9 and 25 grid cells

		X				
	(X,Y)	-2	-1	0	1	2
Y	-2	(-2,-2)	(-1,-2)	(0,-2)	(1,-2)	(2,-2)
	-1	(-2,-1)	(-1,-1)	(0,-1)	(1,-1)	(2,-1)
	0	(-2,0)	(-1,0)	(0,0)	(1,0)	(2,0)
	1	(-2,1)	(-1,1)	(0,1)	(1,1)	(2,1)
	2	(-2,2)	(-1,2)	(0,2)	(1,2)	(2,2)

Figure 1: Atlas coverage

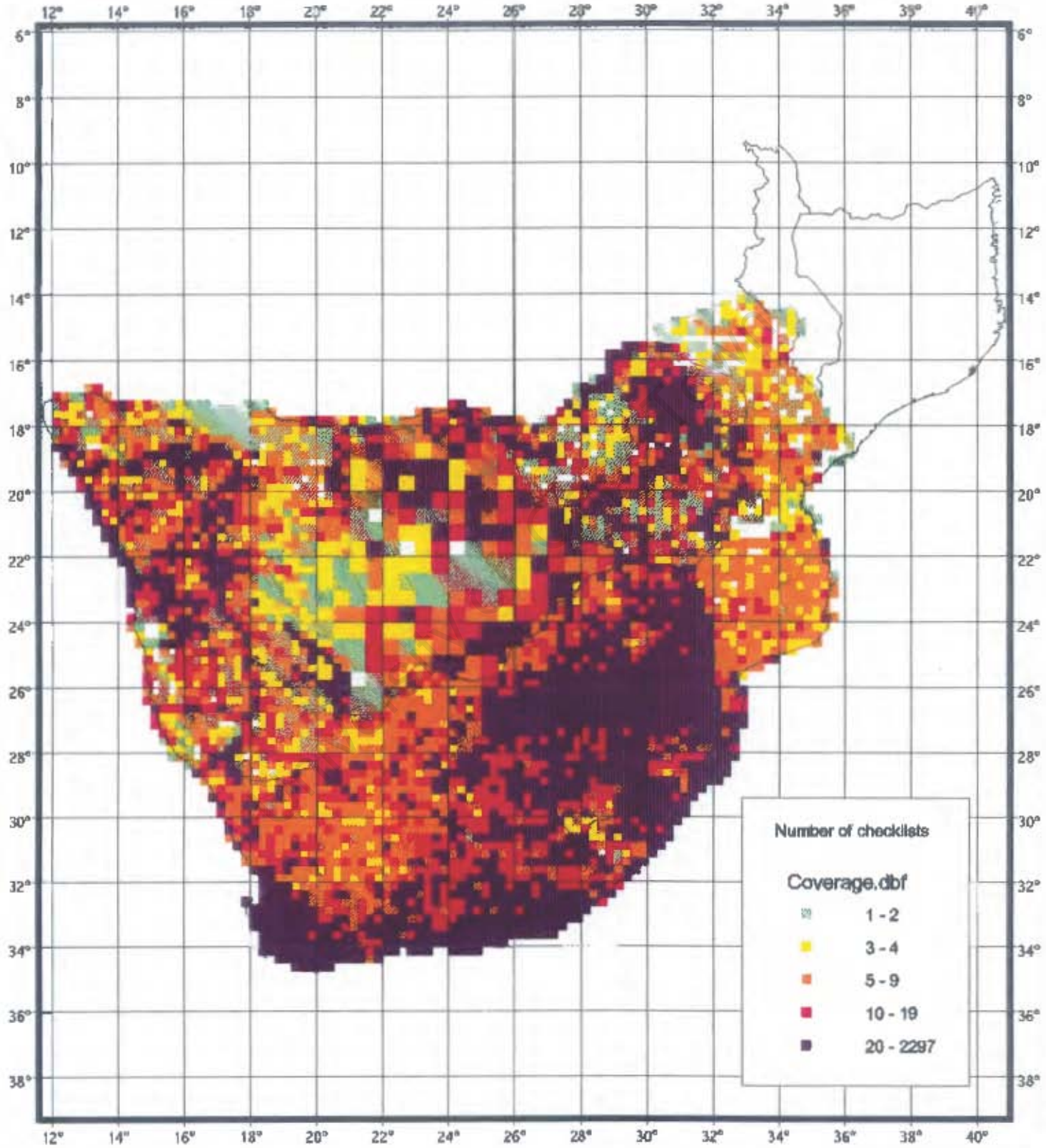


Figure 2: Map of southern Africa showing national boundaries, boundaries of South African provinces and some frequently mentioned regions

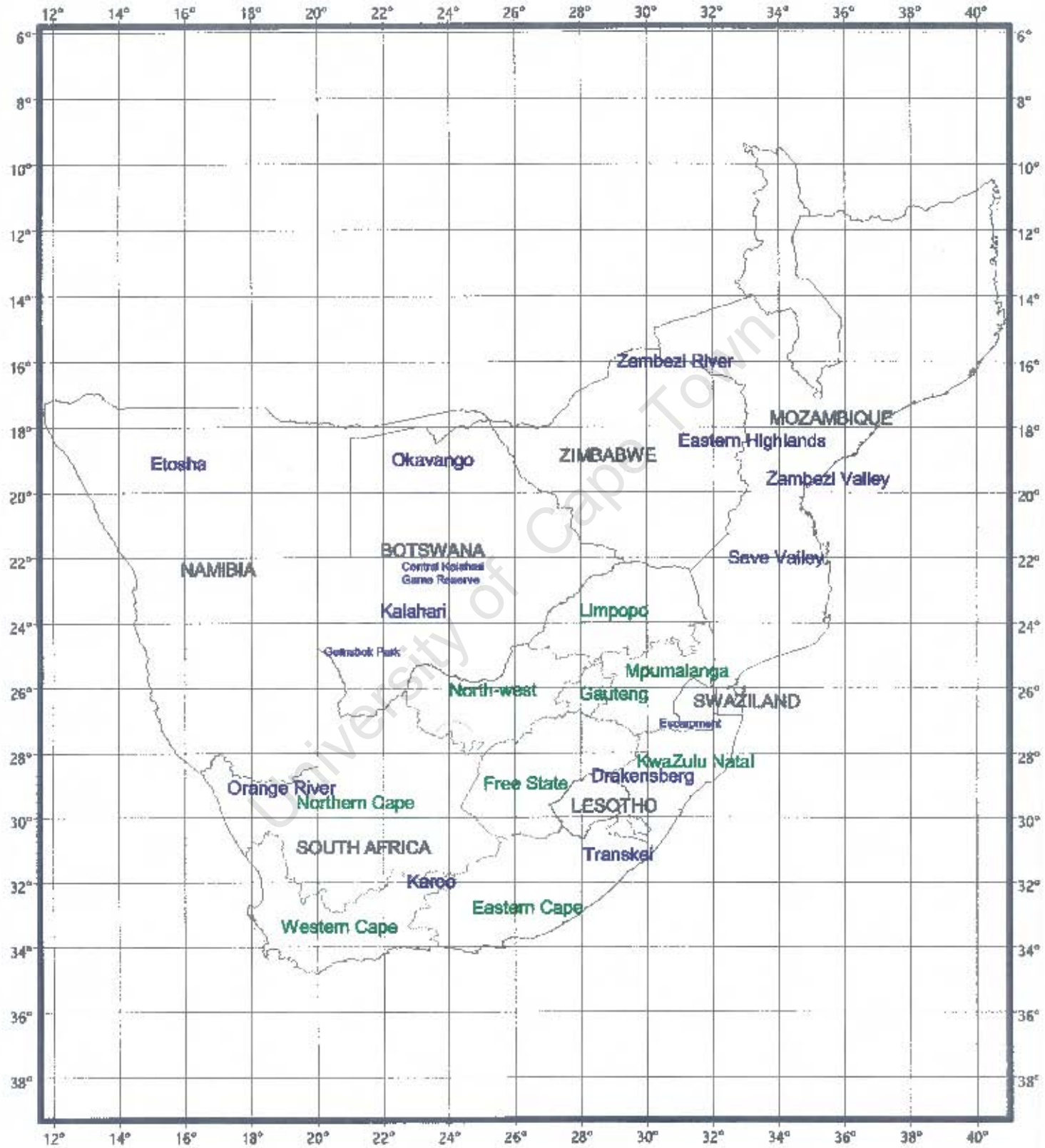


Figure 3: Namaqua Dove - Distribution of observed reporting rates

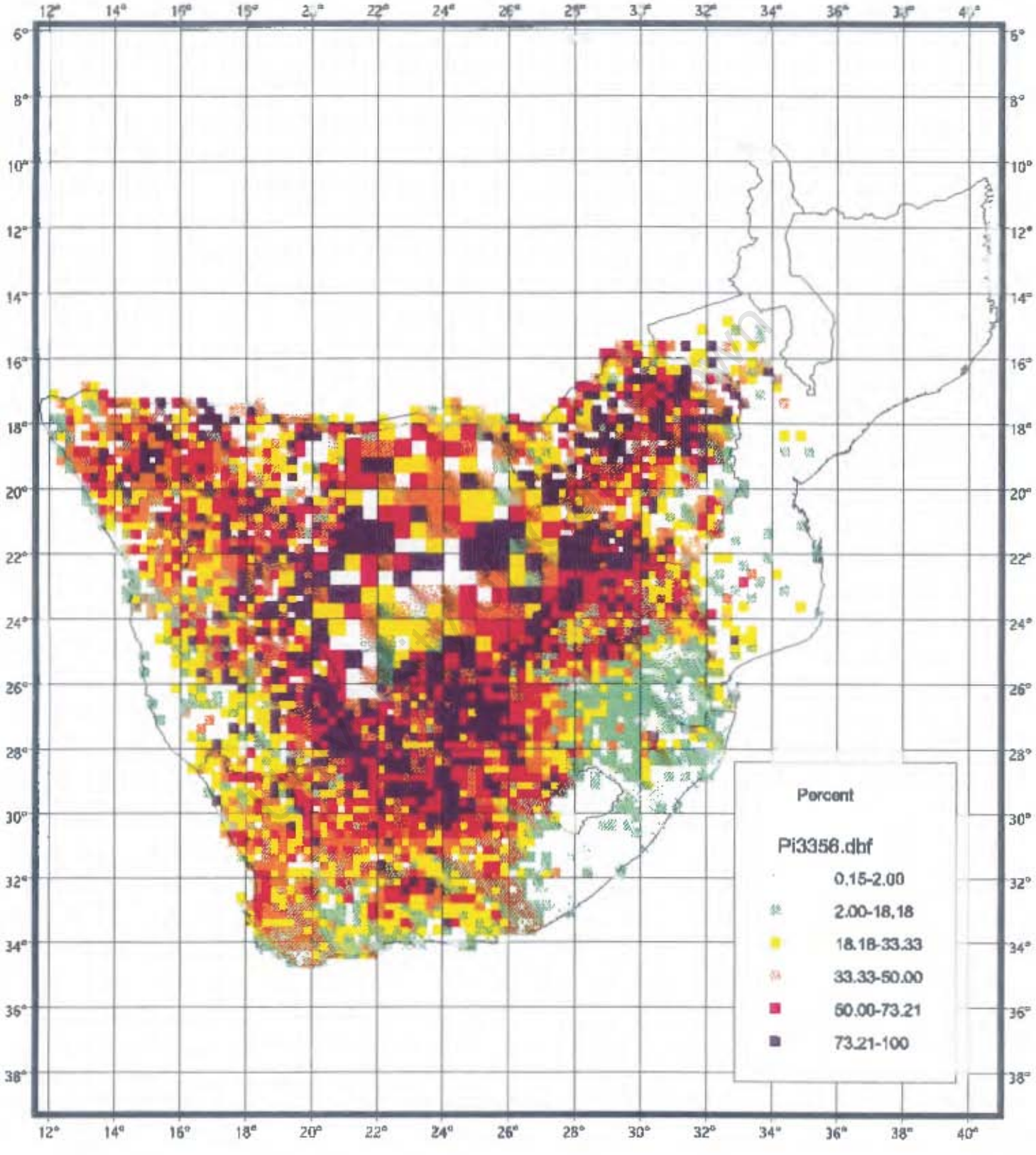


Figure 4: Namaqua Dove - Smoothed distribution using model with four parameters for nine grid cells

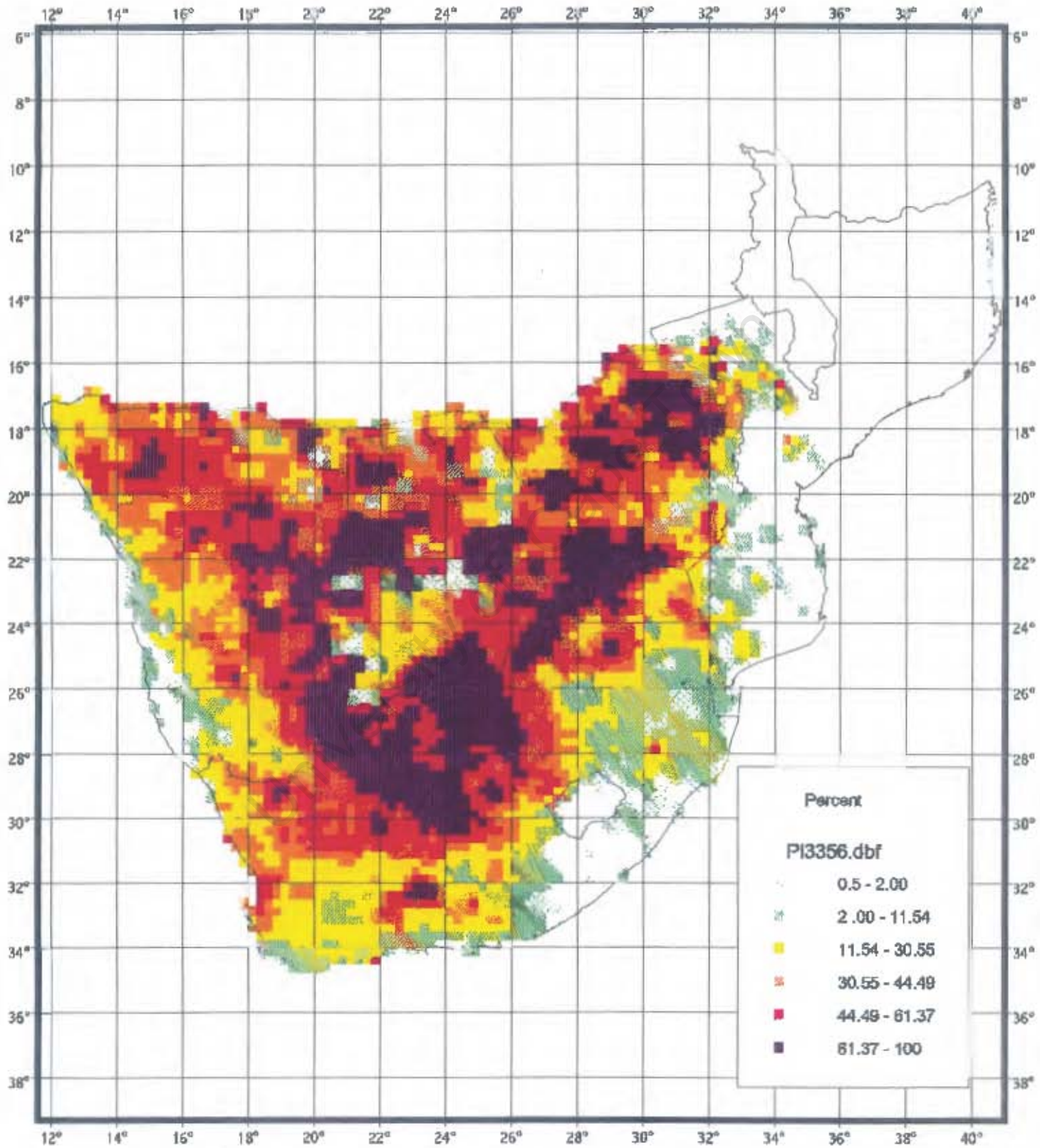


Figure 5: Namaqua Dove - Smoothed distribution using model with four parameters for 25 grid cells

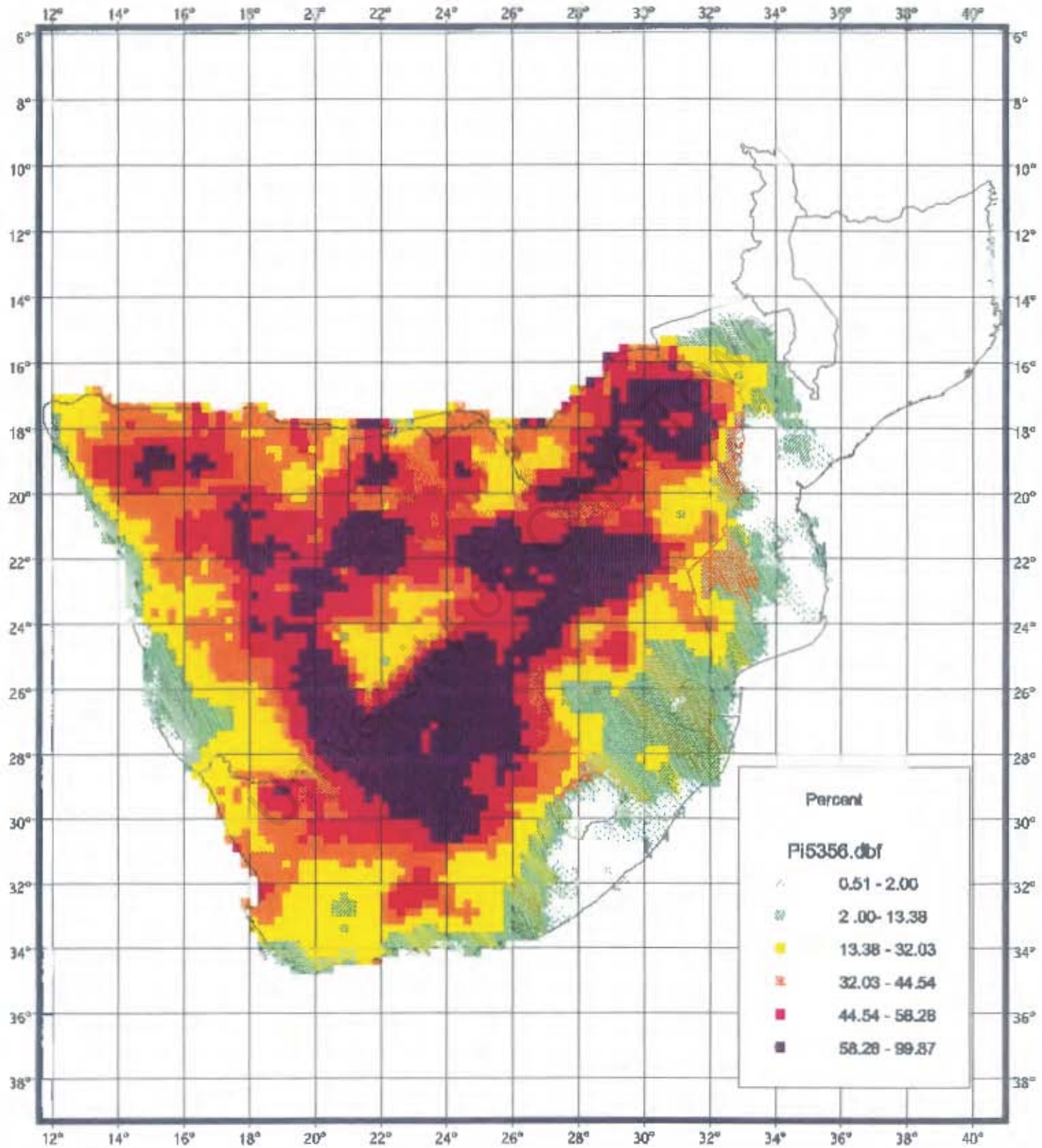


Figure 6: Namaqua Dove - Smoothed distribution using model with six parameters for 25 grid cells

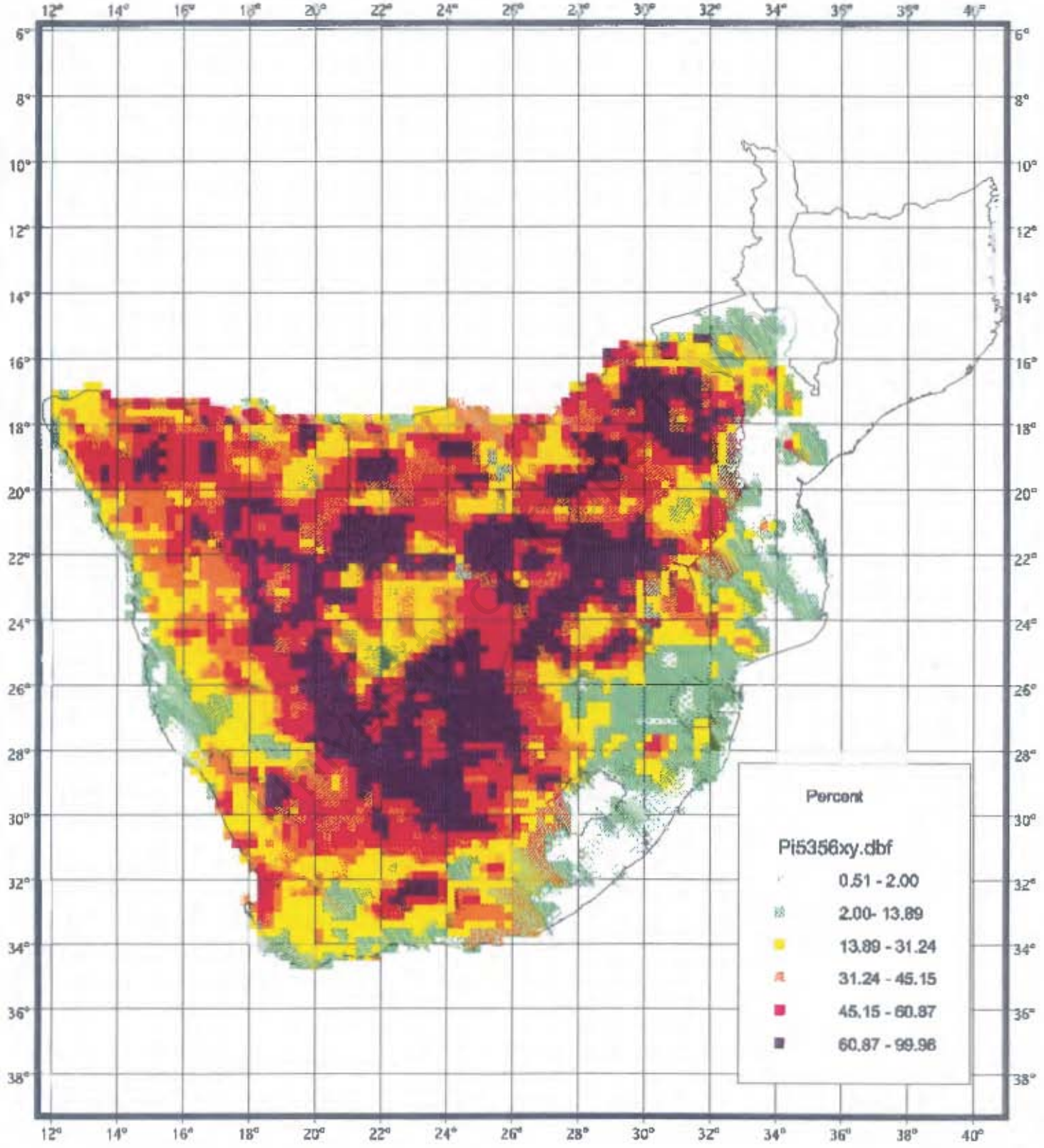


Figure 7: Whiterumped Swift - Distribution of observed reporting rates

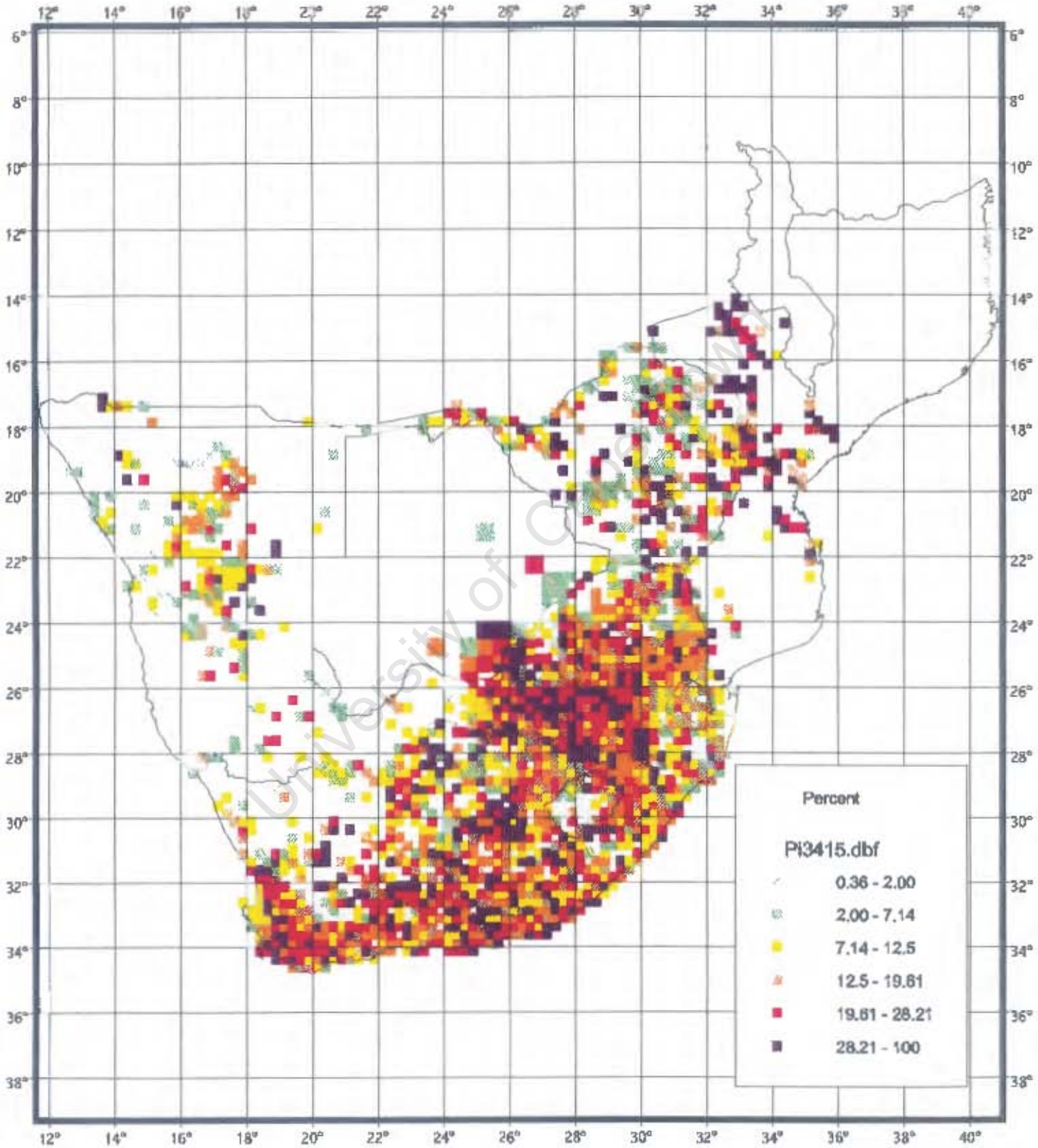


Figure 8: Whiterumped Swift - Smoothed distribution using model with four parameters for nine grid cells

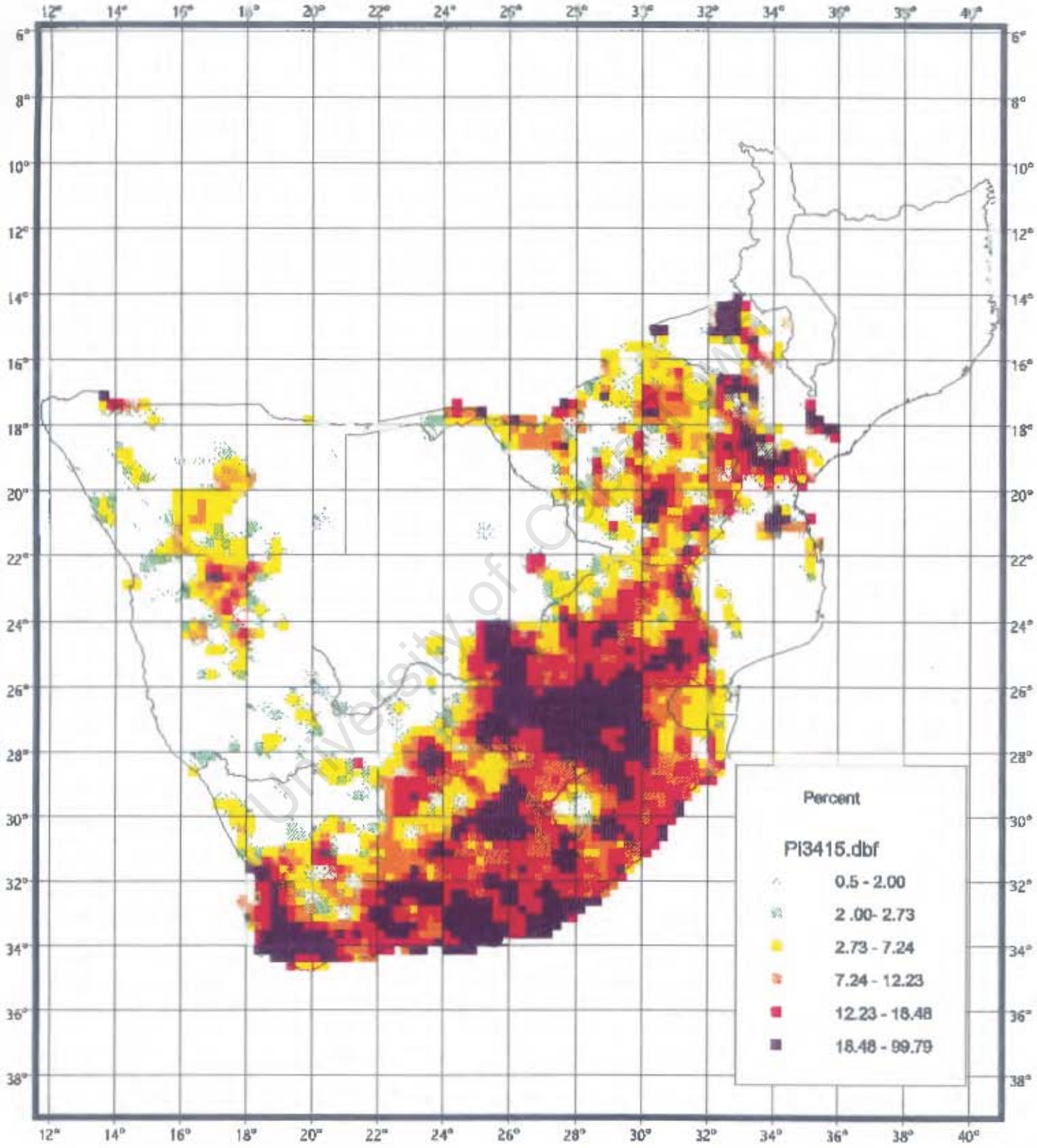


Figure 9: Whiterumped Swift - Smoothed distribution using model with four parameters for 25 grid cells

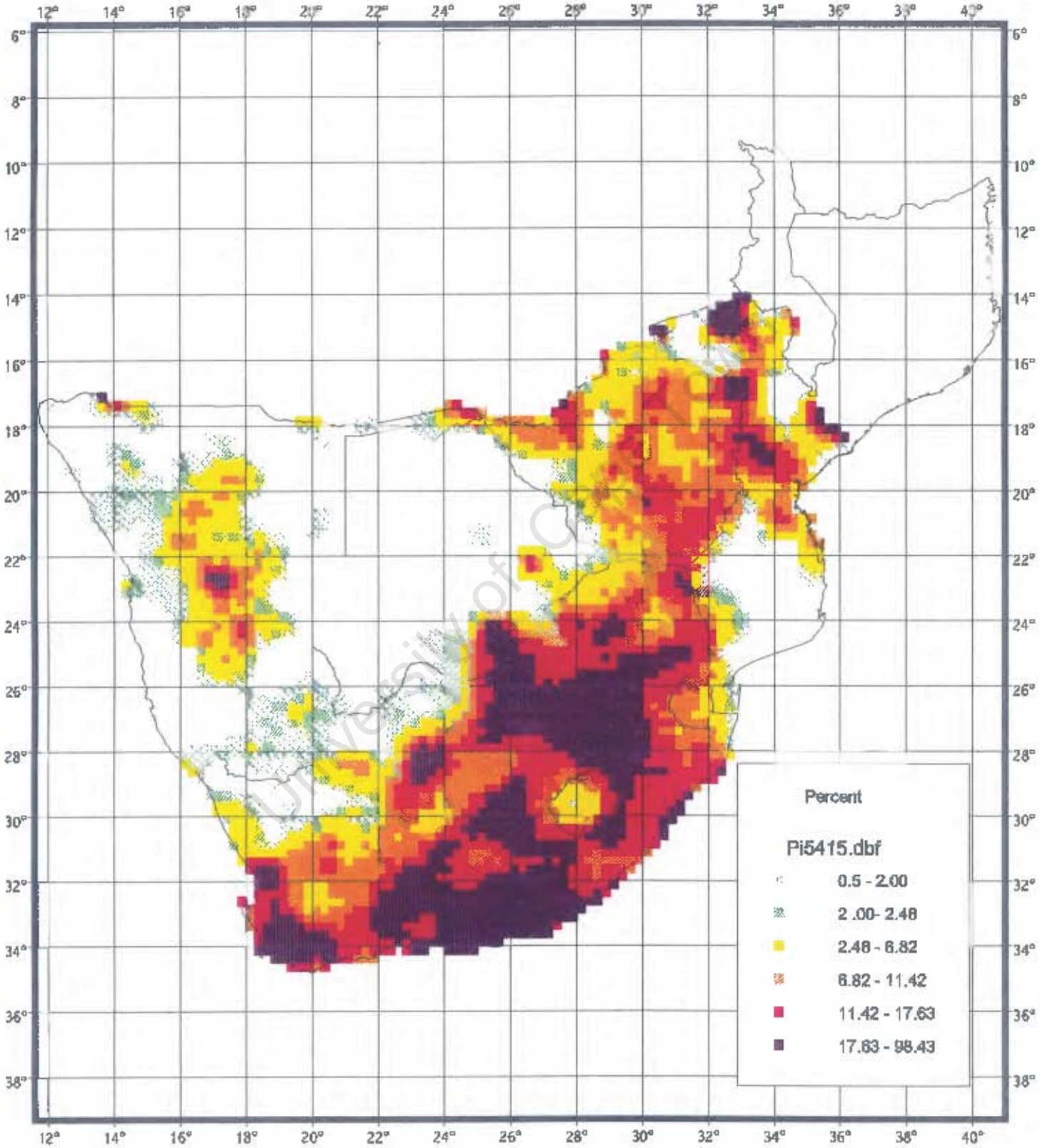
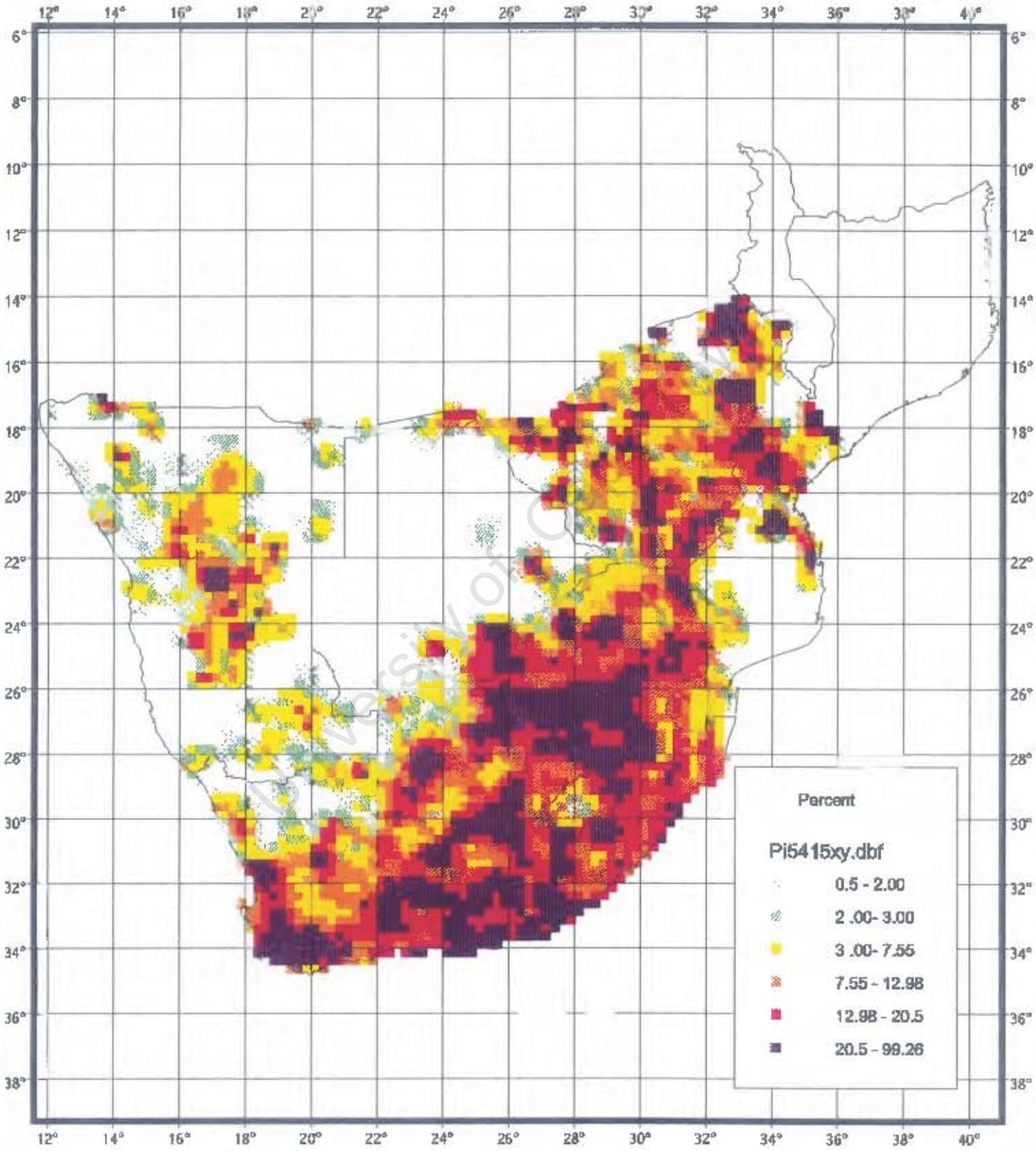


Figure 10: Whiterumped Swift - Smoothed distribution using model with six parameters for 25 grid cells



Chapter 3

Presentation of Maps Based on Observed and Smoothed Bird Atlas Reporting Rates

F. LITTLE[1], L.G. UNDERHILL[1], B. ERNI[1] & B. BRUDERER[2]

1 Avian Demography Unit, Department of Statistical Sciences, University of Cape Town, Rondebosch, 7701 South Africa

2 Swiss Ornithological Institute, Ch-6204 Sempach, Switzerland

Introduction

In Chapter 2 we introduced a method for generating smoothed distribution maps showing detection probabilities for bird species in southern Africa. We showed how these smoothed distributions provide clearer patterns for the distributions of species than the maps based on observed reporting rates. In this chapter we focus on map presentation issues and consider variations of the smoothing method to cope with special distributional characteristics of species. Firstly, we focus attention on generating maps with uniform cutoffs in reporting rate colour categories, that can be used for example in field guides for bird identification. Secondly, we look at the appropriateness of the smoothing method for species with fragmented distributions and suggest a compromise between the observed reporting rates and smoothed detection probabilities. Finally, we revisit our approach to small detection probabilities for which we set values less than 0.5% to zero (Chapter 2, Appendix).

Mapping the Smoothed Distributions

We confine our discussion here to maps based on the four-parameter smoothing model for nine grid cells. The smoothed distributions that are generated by the algorithm presented in Chapter 2 are presented as multicoloured maps. For this purpose, the reporting rates, whether observed or smoothed, need to be converted to a set of colours based on intervals of reporting rates. The choice of the number of colours to use, i.e., the number of intervals into which to subdivide the overall range, and the scale to use for the subdivision impose a further smoothing effect. The maps of observed reporting rates presented in *The Atlas of Southern African Birds* (Harrison *et al.* 1997a, b) as well as the smoothed maps presented in Chapter 2 make use of a "relative shading" method which is species-specific in that it bases the choice of intervals on the range of reporting rate values for each species. The maps in Harrison *et al.* (1997a, b) divide the total range for a species into these groups in such a way that each colour represents a third of the distribution, i.e., the cutpoints are chosen so that the number of grid cells in each colour is as close to equal as possible. All grid cells with a positive reporting rate rate less than 2% are shown with an $\times$. In Chapter 2 we also use the relative shading approach but divide the range into five quintiles. We illustrate all grid cells in the lowest quintile with a positive reporting rate less than 2% with an $\times$. In our smoothed maps, we set to zero estimated detection probabilities smaller than 0.5%.

Figs 1 and 2 show the observed reporting rates and smoothed detection probabilities, respectively, for the Cape Weaver *Ploceus capensis* using this species-specific coding. Differences between the two maps are not only due to differences between the observed rates and smoothed probabilities, but also due to different

subdivision of the ranges to define the colours. We note that for the map of smoothed detection probabilities, the darkest shade represents all probabilities larger than 42.2% (Fig. 2), while the darkest shade for the map of observed rates (Fig. 1) represents all rates larger than 54.9%.

Figs 4 and 5 show the observed and smoothed distributions for the Blackheaded Canary *Serinus alario*. Once again the scales used to define the colour coding differ between the two maps and furthermore these scales differ even more from the scales used for the maps of the Cape Weaver. The Blackheaded Canary has a sparser distribution than the Cape Weaver and thus occurs in smaller numbers, even in the areas where it is relatively abundant. We notice from the keys to these maps that the darkest shade represents rates above 40% and probabilities above 28% for the observed (Fig. 4) and smoothed (Fig. 5) distributions, respectively. However, this is not immediately clear from the shading of these distribution maps.

The species-specific shading thus complicates the comparison of species distributions and significantly dilutes the information regarding the probability of sighting an individual species in a given geographical area. To overcome this problem, we derive absolute scales to define the colour shading. We divide the range from 2% to 100% into six sub-ranges. To take into consideration the skew distribution of species' abundance, we choose greater than 50% as the top category and use a geometric scale for the subdivision of rates between 2% and 50% into five sub-ranges. This defines each cut-off value as $a_i = 2r^i$, for $i = 1$ to 5, where $a_0 = 2$ and $a_5 = 50$ (giving $r = 25^{1/5} = 1.904$). The values are 2.0%, 3.7%, 7.1%, 13.6%, 26.0%, 50.0%. The sub-ranges thus get wider with larger reporting rates.

Figs 3 and 6 use this absolute scale to illustrate the smoothed distributions of the Cape Weaver and the Blackheaded Canary, respectively. These maps give the same impressions of relative abundance within species as did Figs 2 and 5. In addition, we can now easily compare the detection probability distributions for the two species because the two maps use the same scale to define the colour shading.

The original impressions of relative density shown in the Figs 1, 2, 4 and 5 have been preserved. In addition, the maps clearly show that the Blackheaded Canary occurs in smaller numbers than the Cape Weaver. Using a fixed scale allows us to generate maps that can be used as a tool for reading the likely probability of species detection in a given geographical area. For example, from Fig. 3 we can determine that the probability of detecting a Cape Weaver in the Western Cape is in excess of 50%. Fig. 6 shows us that, for example, the Blackheaded Canary can be detected in Central Namibia with probabilities below 13.6%.

The Cloud Cisticola *Cisticola textrix* is a species that occurs in relatively small numbers in South Africa, south of 26°S between 25°E and 33°E, narrowing to between 25°E and 28°E south of 28°S, with an even sparser distribution in the Western Cape. This is clearly illustrated on the map that uses an absolute scale for the colour shading (Fig. 8). We note how none of the grid cells has detection probabilities that fall in the top interval and very few grid cells have probabilities that fall in the second interval from the top. When using a relative scale that subdivides the range of detection probabilities for the Cloud Cisticola into five quintile intervals, we obtain the distribution illustrated in Fig. 7. The dark shades now give the wrong impression of high detection probabilities. Only careful scrutiny of the key to the map shows that the darkest shade reflects probabilities that merely lie between 12.4% and 31.3% and that all of the bottom quintile and most of the second quintile from the bottom reflect probabilities less than 2% that are indicated with an × on the map. For this species, the more accurate reflection of its detection probability distribution is probably given by the map based on the absolute scale (Fig. 8).

By contrast, the Blue Waxbill is a species with a fairly uniform distribution of high detection probabilities. Using the absolute mapping scale generates a distribution map of predominantly one colour (Fig. 10) that fails to display variations among these high detection probabilities. The map based on the relative scale (Fig. 9) divides the range between 45.7% and 100% that was mostly shaded in purple in Fig. 10 into three quintile intervals. It clearly shows the variations among the top 60% of the reporting rate distribution for the Blue Waxbill.

There are pros and cons to both methods for dividing the reporting rate range into sub-intervals for mapping. The main advantage of employing an absolute scale for the subdivision of the range across all species is that the categories for all species are the same, making it easier to interpret the distribution maps. However, reporting rates should ideally not be compared between species because of variability in conspicuousness, and the use of absolute rates encourages this and can therefore be misleading. The absolute scale does give an accurate reflection of the reporting rates for sparsely distributed species, especially for those whose reporting rates fall predominantly in the lower ranges of our absolute range. However, the reverse is true for species with uniform distributions of high reporting rates where the absolute scale will mask variations among these high reporting rates.

There may be instances though, where the reporting rates are truly more or less constant across the region, in which case the relative reporting rates impose a false sense of the degree of variability. The class intervals between the different colours will then be very narrow. This is more likely to be a problem for small areas rather than large ones. The final choice of subdivision will depend on the intended use of the species distribution maps.

Deciding between Observed and Smoothed Data

We may not wish to smooth reporting rates in areas where there are large numbers of checklists and wish only to smooth in areas where data collection was sparse. Likewise, smoothing might be deemed inappropriate for species with highly fragmented distributions. Examples of such species would be forest species, which are limited to grid cells in which forests occur, and waterbirds which are limited to grid cells in which suitable habitat is present. To achieve this, Erni (1998) suggested using as the final value to be mapped, a weighted average between the observed reporting rates and the smoothed detection probabilities from the regression model,

$$R_{sm} = f(n_i) x R_{mod} + (1 - f(n_i)) x R_{obs}, \quad (1)$$

where

- R_{sm} is the value to be mapped
- R_{mod} is the smoothed detection probability found by the regression model
- R_{obs} is the original observed reporting rate
- $f(n_i) = \exp(-\alpha n_i)$ with
  - n_i = the number of checklists for grid cell i, and
  - α is some value greater than 0 which determines how strongly the number of checklists influence the weights.

For distributions with a great degree of continuity, the value of α should be close to zero, giving more weight to the modelled rates. For highly fragmented distributions, the regression approach is less valid and here α

should be close to one. In geostatistics the variogram has been developed to describe spatial autocorrelation (Cressie 1993). Erni (1998) estimated the variogram for binomial data. She suggests using the first value of the estimated variogram as a measure of continuity and that α should depend on this value. A value of $\alpha = 0.05$ could be used for distributions where significant spatial autocorrelation is present and a value of between 0.1 and 0.5 if the estimated variogram value is larger than approximately 0.8.

We focus attention once again on the distribution maps for the Cape Weaver. Figs 1 and 2 give the maps based on the observed reporting rates and the smoothed detection probabilities, respectively. The features of the original data are well preserved in the smoothed map, especially in the block of grid cells in the area defined by 30°S to 33°S and 20°E to 24°E. One area where too much smoothing possibly occurs, is in the area 31°S and 28°E. In the original data, zero reporting rates occur in what seems not a random pattern. One could argue that these zero rates should have been preserved in the modelling. However, in Fig. 2 these zero cells have disappeared due to smoothing. Figs 11a-11d give the distribution maps for the Cape Weaver based on a weighted average between observed rates and smoothed detection probabilities for different values of α. These maps clearly illustrate how the smoothness of the distribution decreases with increasing alpha. The distribution map based on the model with $\alpha = 0.50$ (Fig. 11d) resembles the distribution maps based on the observed rates (Fig. 1), while setting $\alpha = 0.05$ (Fig. 11a) generates a map not very different from the map based solely on detection probabilities from the regression model (Fig.2). We can see how the map based on $\alpha = 0.05$ starts re-inserting the blank cells in the above mentioned area and how more of these cells turn blank as α increases and more weight is given to the observed rates. The smoothing algorithm deletes isolated single values, for example, the data of cell 2820CB (see Fig. 2 versus Fig. 1), where the observed reporting rate was 1/19. Smoothing methods that take into account the information of surrounding cells should eliminate these isolated values. It is likely that the species generally does not occur in this cell, except as a vagrant out of its range. The weighted average between the smoothed detection probabilities and the observed reporting rates tend to insert isolated values back into the distribution maps almost immediately (Fig. 11d). This is a disadvantage.

Another problem with this approach is that, for many of the grid cells with large numbers of checklists, these were contributed by a single observer. Not only is there then a lack of independence between checklists (so that the effective number of checklists is smaller than it appears to be), but also the observed reporting rate might be biased by an "observer skills effect" (see Harrison & Underhill 1997). Thus the process of using weighted averages between observed and smoothed reporting rates may not be a sensible approach to the problem of getting closer to the "true" detection probabilities in each grid cell. We therefore do not use this method in subsequent Chapters 4 and 5 of this thesis.

The Whiterumped Swift *Apus caffer*, discussed in Chapter 2, has a wide distribution that covers large areas of southern Africa. Its distribution thus includes areas of both high and low intensity of fieldwork. We focus on two specific areas of high and low coverage to illustrate how the use of the weighted average suggested in equation (1) depends on the number of checklists on which the reporting rates are based. The area of high coverage is the eastern highlands of Zimbabwe, south of 18°S along 32.5°E, whereas central Mozambique depicts an area where the coverage was patchy and low (see Fig. 1 of Chapter 2). Fig. 7 of Chapter 2 shows the distribution of observed reporting rates for the Whiterumped Swift. It clearly reflects the patchy coverage in central Mozambique, though in the grid cells that were surveyed, the observed reporting rates are relatively high. In the eastern highlands of Zimbabwe, the reporting rates are much lower, in fact below 19.6%, with the exception of two grid cells. We would ideally want a model that will share out the high observed rates in central Mozambique to neighbouring areas that were not covered but that will leave

the reporting rates in the intensely covered eastern highlands of Zimbabwe unchanged. The distribution based on the smoothed detection probabilities (Fig. 8 of Chapter 2) does expand the central Mozambique distribution of the Whiterumped Swift. However, it also increases the rates in the eastern highlands of Zimbabwe uniformly across the area. Figs 12a and b show the distribution maps based on a weighted average between observed reporting rates and smoothed detection probabilities when $\alpha = 0.5$ and $\alpha = 0.05$, respectively. When $\alpha = 0.5$, the distribution map (Fig. 12a) closely resembles the distribution of observed rates, though in central Mozambique the distribution has already widened. This shows that in the areas of low coverage, where n_i is small, the modelled detection probabilities carry more weight than the observed reporting rates even with a relatively high value for α. Fig. 12b shows that $\alpha = 0.05$ gives much more weight to the modelled detection probabilities and results in a smoother distribution overall. However, the rates in the eastern highlands of Zimbabwe are closer to the observed reporting rates (Fig. 7 of Chapter 2) than to the smoothed detection probabilities (Fig. 8 of Chapter 2). This illustrates how even a small α allows the strength of large n_i to give more weight to the observed rates in areas of high coverage while at the same time smoothing the distributions in the sparsely covered areas.

The use of a weighted average thus enables us to treat different areas differently when smoothing the species distributions. We can very easily apply different degrees of smoothing on different areas based on the degree of the intensity of the coverage in the areas. At the same time the specific choice of alpha will depend on the degree of fragmentation of the species distributions. Further research is required to look at incorporating the estimated value of the variogram directly in the estimation of α.

Zero Cut-off points

When developing the smoothing model, Erni (1998) chose to set all smoothed detection probabilities less than 0.5% to zero. This approach assumes a species to be absent if the smoothed detection probability is so small that the species is expected to be recorded only once in more than 200 checklists. It makes sense that grid cells with very small reporting rates should not be counted as part of the distribution range of a species. However, for very sparsely distributed species and possibly highly endangered species that occur in small numbers, truncation of small detection probabilities may lead to a disproportionately large reduction in the extent of the species range.

To illustrate, we focus attention on the distribution of the Thrush Nightingale *Luscinia luscinia*, a Palearctic migrant to the northern parts of the study area. It was recorded in 92 grid cells with observed non zero reporting rates ranging from 0.19% to 100%. The mean reporting rate was 13% and the median reporting rate was 6.7%. This species was recorded in only two grid cells in South Africa (Fig. 13). The standard smoothing algorithm that truncates all detection probabilities less than 0.5% to zero (Fig. 14), results in 179 non zero grid cells with none of these falling within the borders of South Africa. Fig. 15 illustrates the smoothed distribution when we have chosen to illustrate all non zero detection probabilities without any truncation of the very small values. The species distribution range now covers 463 grid cells though many of these are of very small magnitude. Choosing to truncate probabilities less than 0.1% to zero, results in 385 non-zero grid cells (Fig. 16). One of the South African grid cells has re-appeared, while in the map with no truncation (Fig. 15) the two observed non zero rates have been expanded to cover most or all of the surrounding blocks of nine grid cells.

The approach to mapping the smoothed distributions so far used, illustrates all detection probabilities less than 2% with an $\times$, rather than using a fully shaded grid cell. However, once again this approach may not be

appropriate for a sparsely distributed species like the Thrush Nightingale. In fact, the smoothing algorithm only generates 58 detection probabilities in excess of 2%. Thus 87.5% of the range of detection probabilities for the Thrush Nightingale will be illustrated with an × (Fig. 17). An alternative approach to using the absolute value of 2% would be to illustrate the lower 5% of the range of detection probabilities for a species with an ×. Fig. 18 shows the difference when this is done for the Thrush Nightingale. Cognizance should be taken of the very low predicted probabilities associated with the darker shades for these sparse distribution maps.

While we believe that the original approach of setting all detection probabilities less than 0.5% to zero and illustrating remaining detection probabilities less than 2% with an ×, the above two variants of this approach are available and easily implemented for species for which extremely small reporting rates may be of interest.

Conclusion

In this chapter, we have presented three variants to our approach for generating smoothed distribution maps of detection probabilities for species. The adjustments are aimed at coping with specific issues related to the mapping of species distributions. The absolute subdivision of the range of detection probabilities allows for the generation of maps that are easy to interpret and that can be used in field guides for the identification of species. Using a weighted average between observed reporting rates and smoothed detection probabilities allows us to apply different degrees of smoothing, depending on both the degree of fragmentation of the species distribution and the intensity of the field work in different areas. The different approaches for dealing with small detection probabilities may be of use when it is important to take into account even the occasional sightings of very scarce species.

All three adjustments to the smoothing approach are very easy to implement. While these adjustments may each be of value in the special circumstances described above, the general smoothing model as presented in Chapter 2 remains the preferred method for generating maps with accurate information for most species regarding their detection probability in given areas.

References

Cressie NAC 1993. Statistics for Spatial Data. Johyn Wiley & Sons. New York.

Erni B 1998. Analysis of Distribution Maps from the Bird Atlas Data. Unpublished M.Sc thesis, University of Cape Town.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997a. The Atlas of Southern African Birds. Vol 1: Non-passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997b. The Atlas of Southern African Birds. Vol 2: Passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Underhill LG 1997. Introduction and methods. In: Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds). The Atlas of Southern African Birds. Vol 1: Non-passerines, xliii-lxiv. BirdLife South Africa, Johannesburg.

Figure 1: Cape Weaver - Distribution of observed rates using a relative scale for colour shading

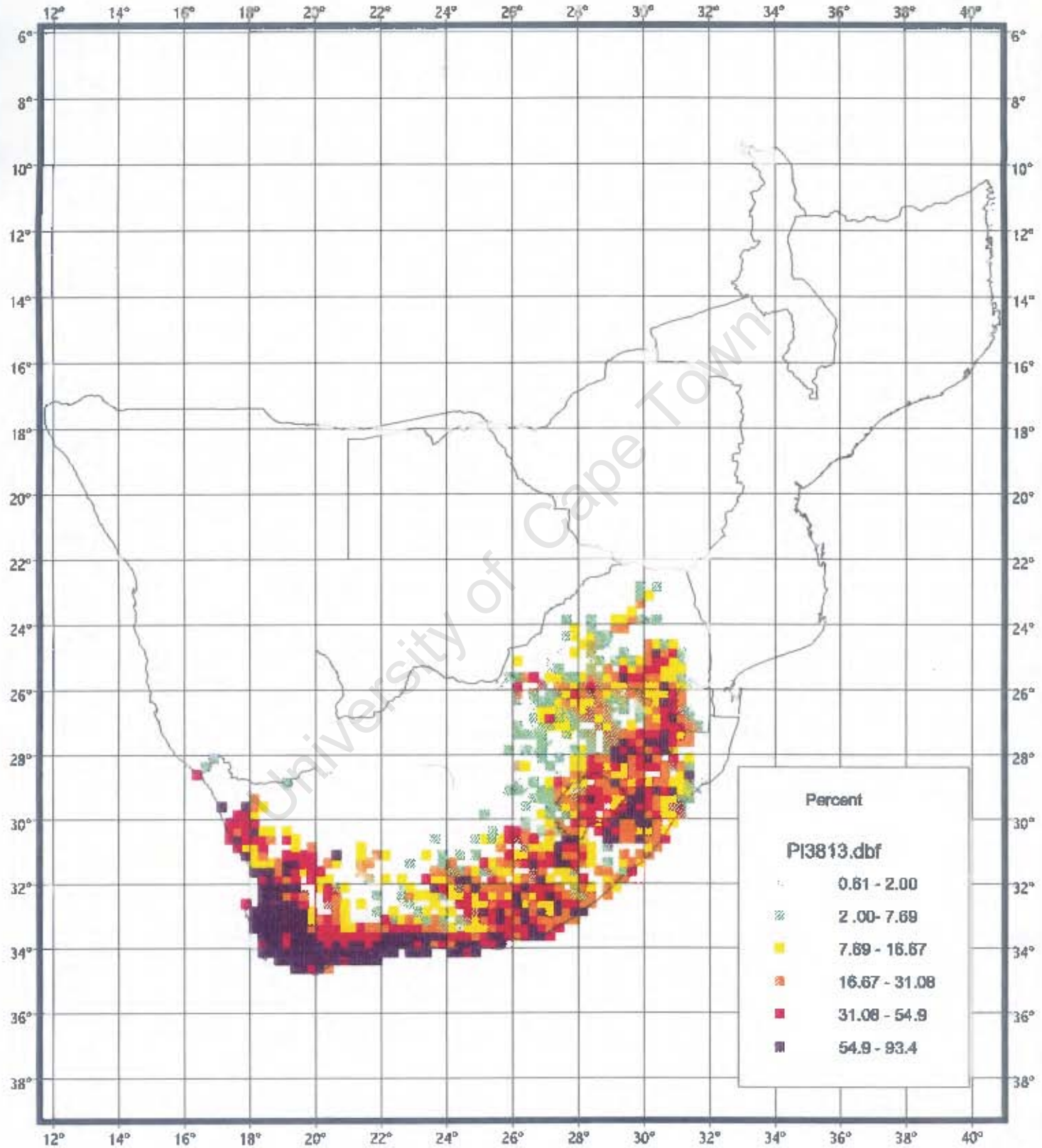


Figure 2: Cape Weaver - Smoothed distribution using a relative scale for colour shading

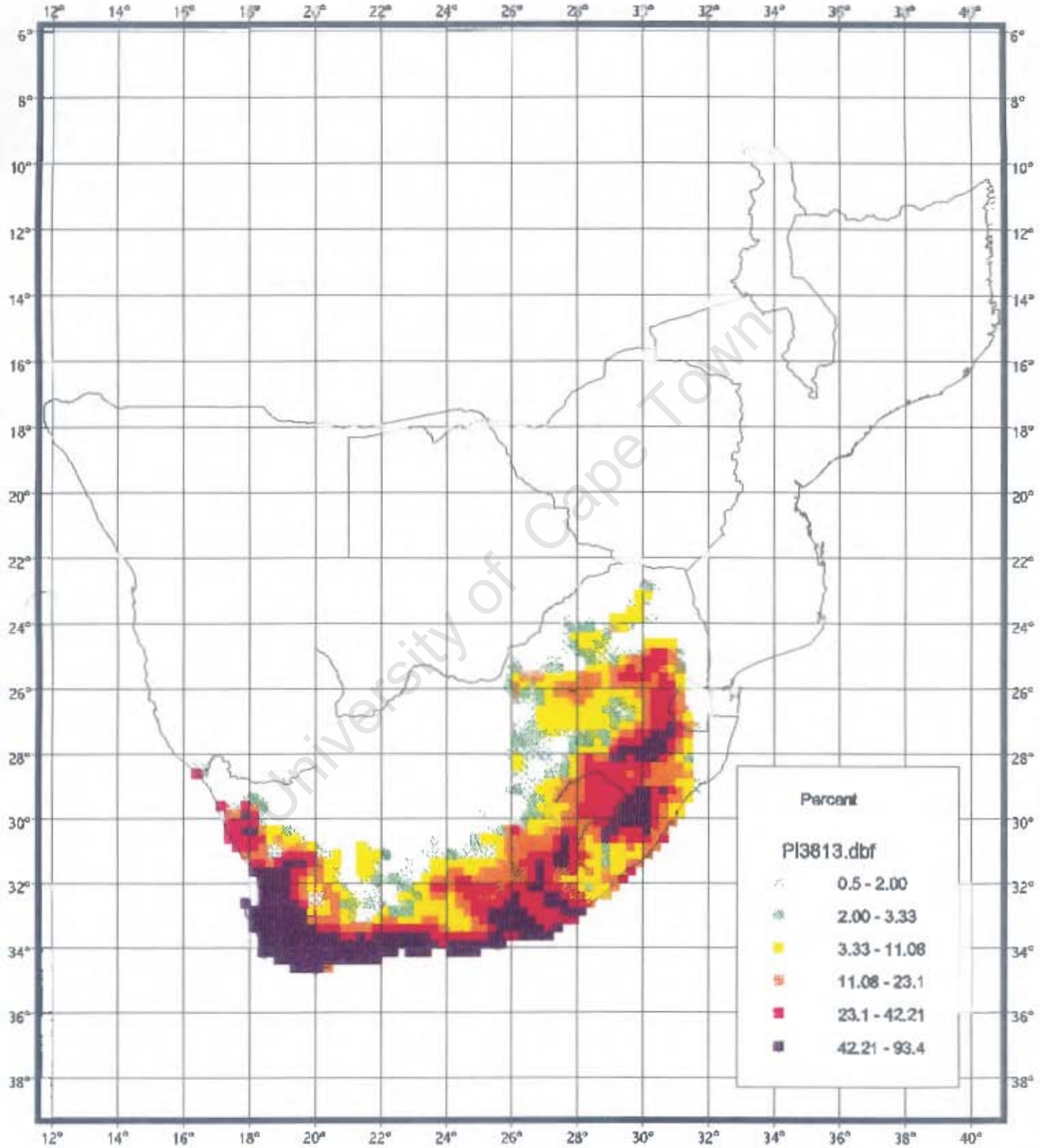


Figure 3: Cape Weaver - Smoothed distribution using an absolute scale for colour shading

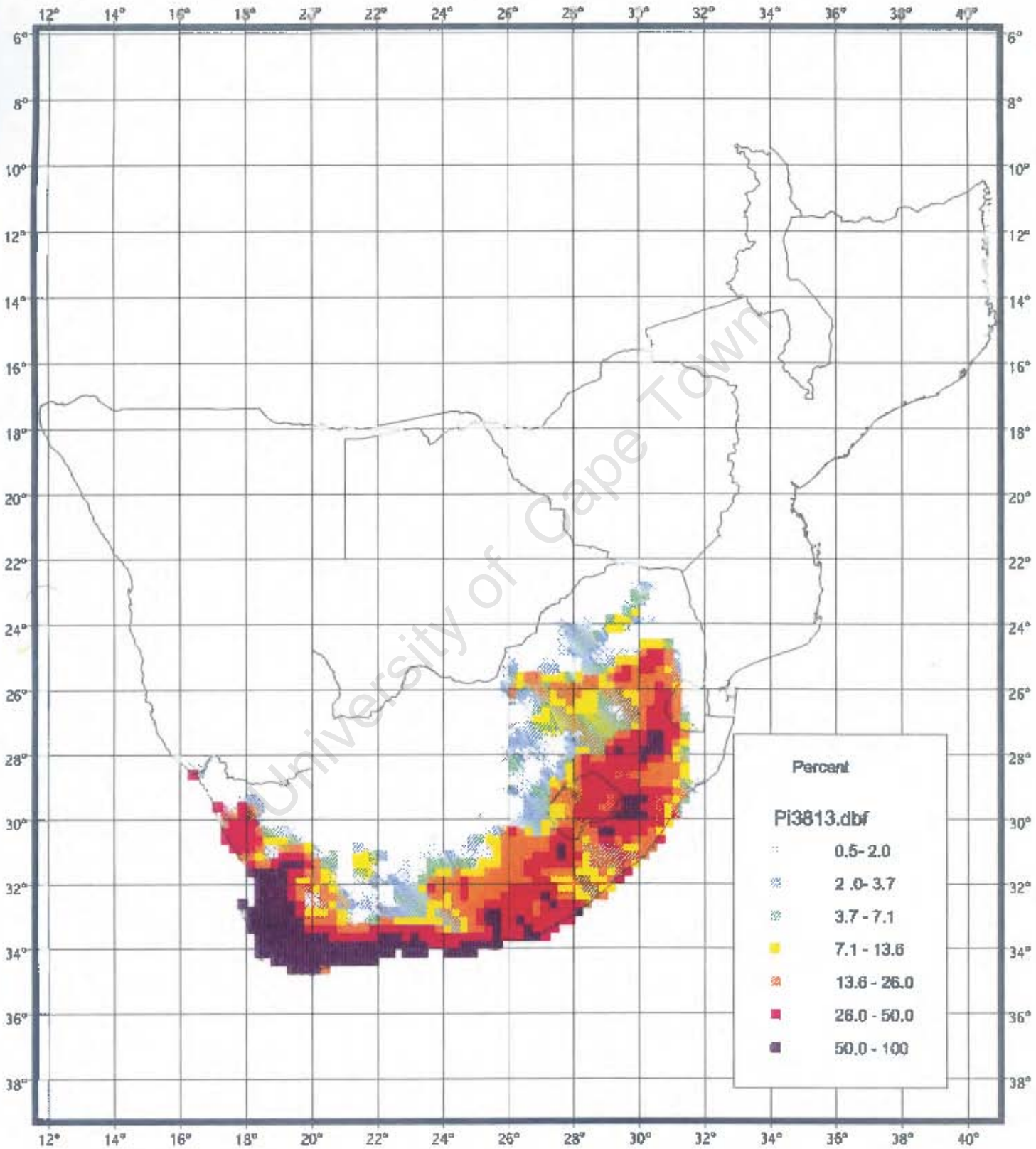


Figure 4: Blackheaded Canary - Distribution of observed reporting rates using a relative scale for colour shading

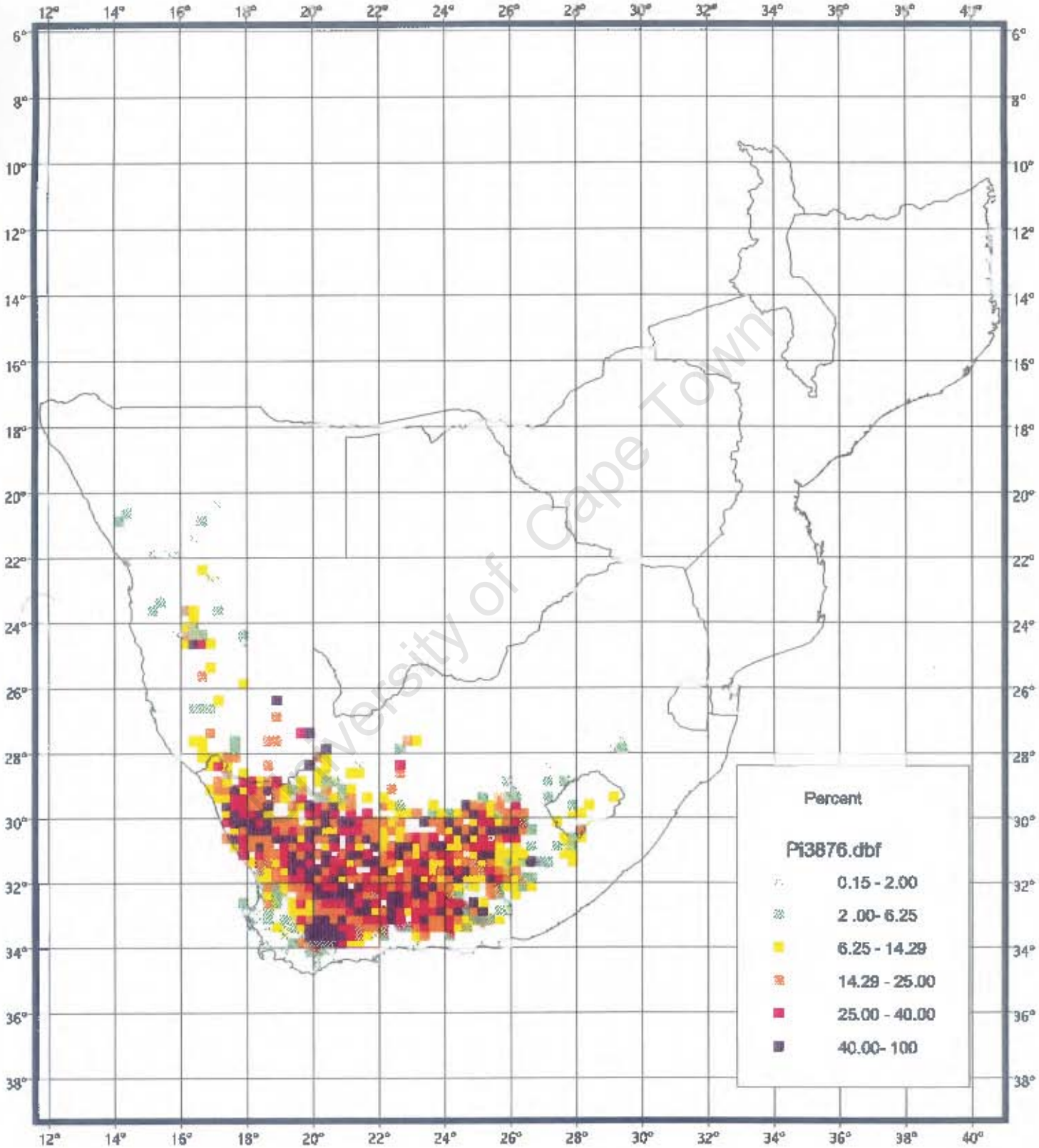


Figure 5: Blackheaded Canary - Smoothed distribution using a relative scale for colour shading

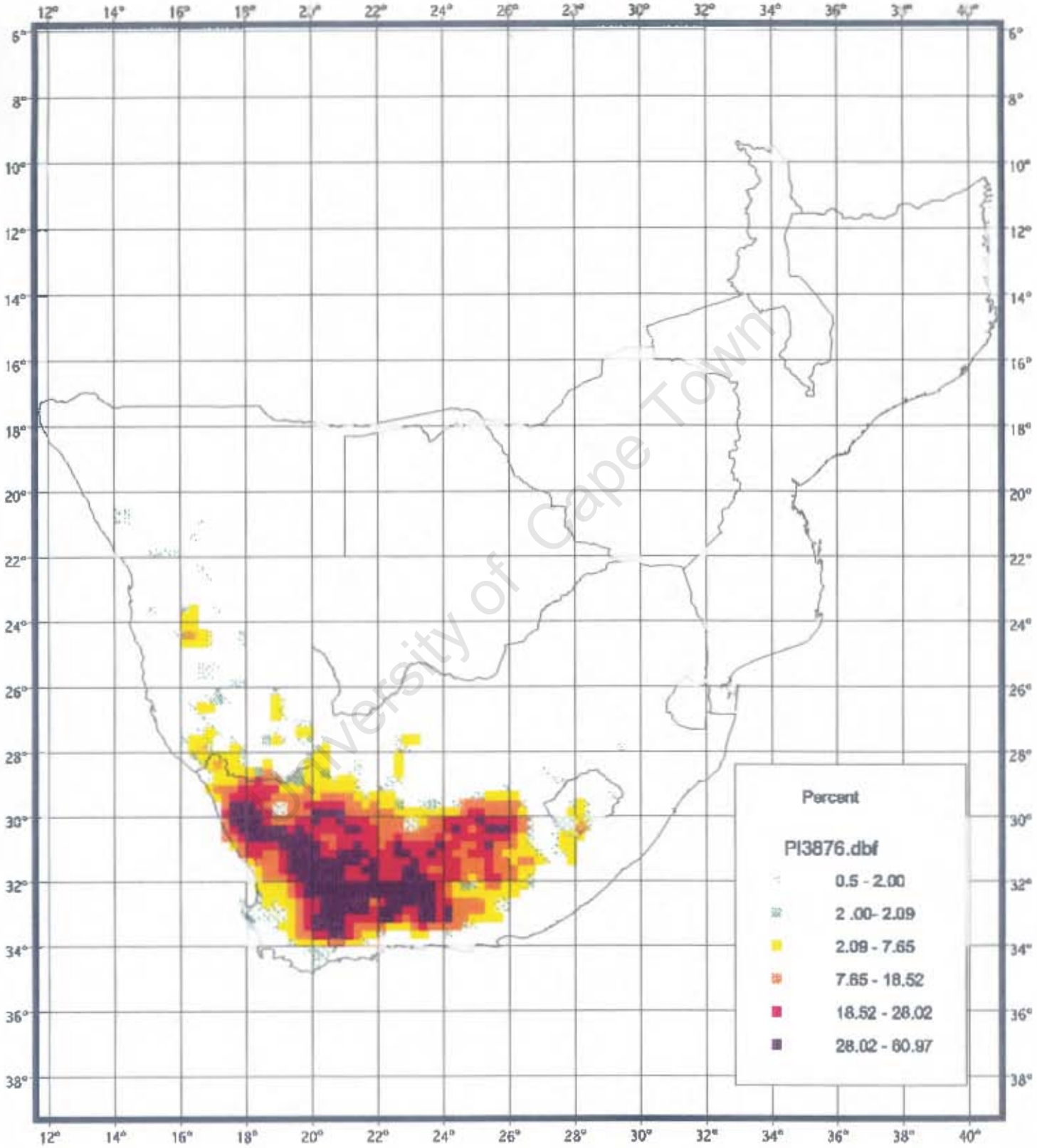


Figure 6: Blackheaded Canary - Smoothed distribution using an absolute scale for colour shading

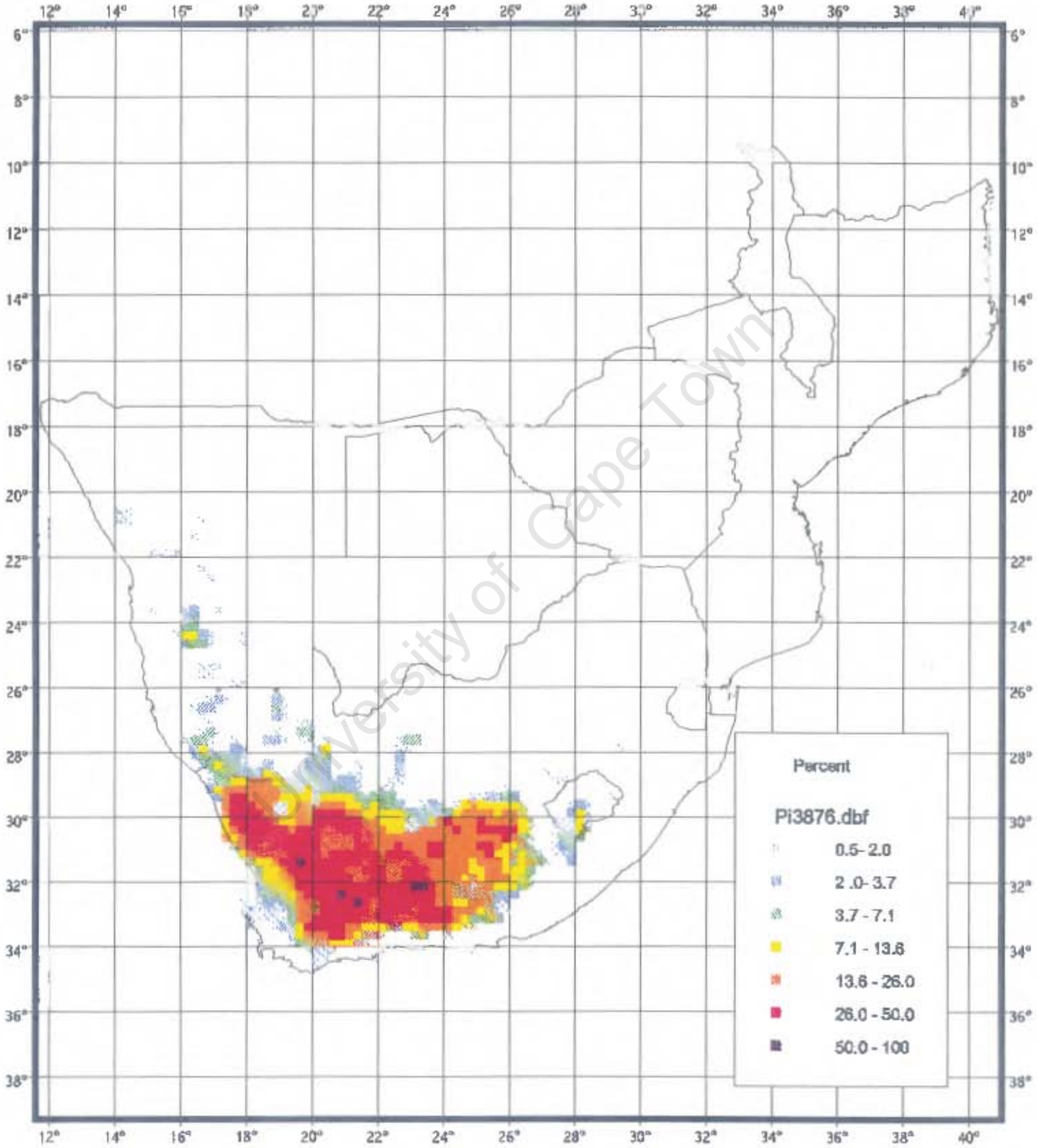


Figure 7: Cloud Cisticola - Smoothed distribution using a relative scale for colour shading

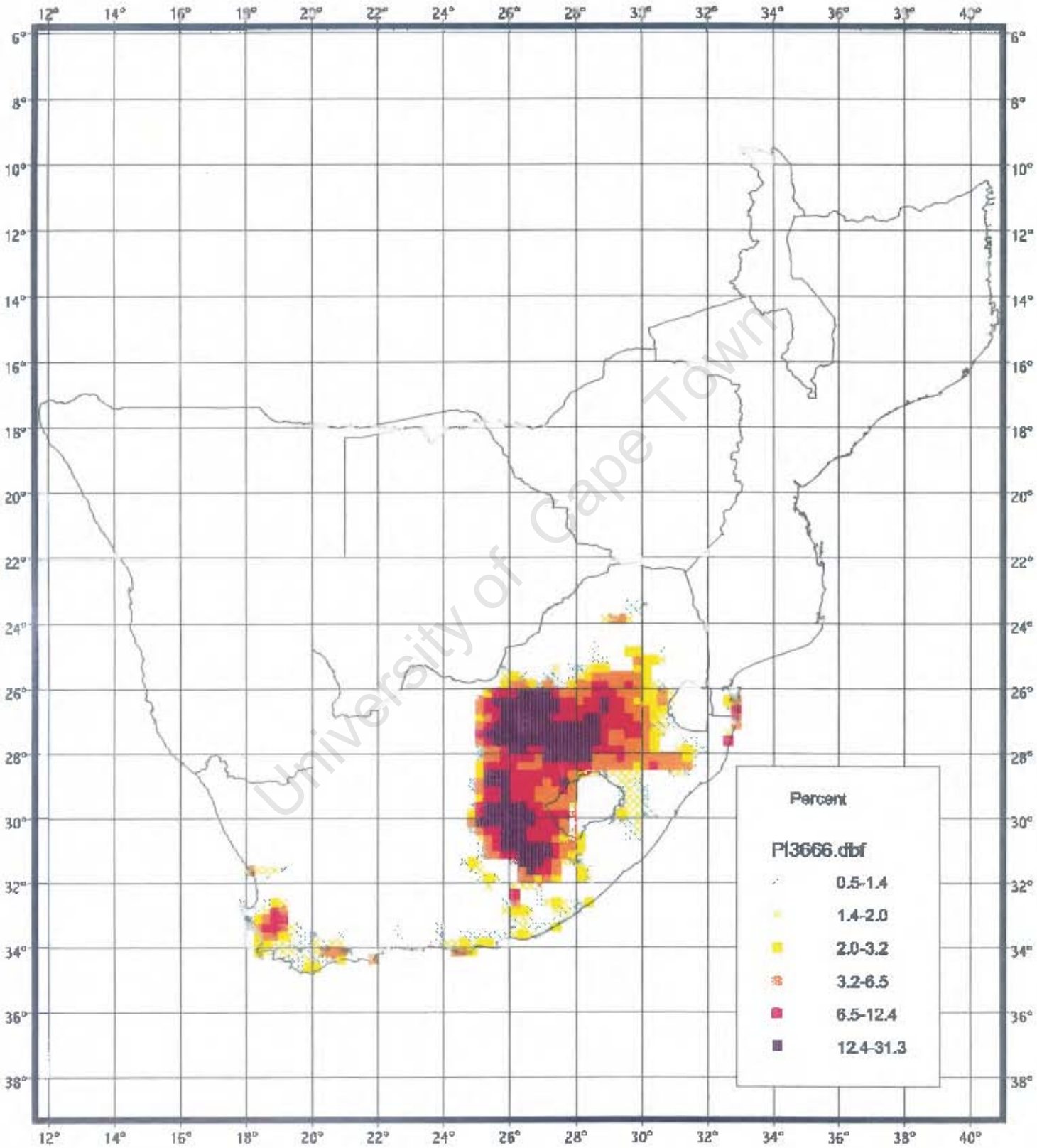


Figure 8: Cloud Cisticola - Smoothed distribution using an absolute scale for colour shading

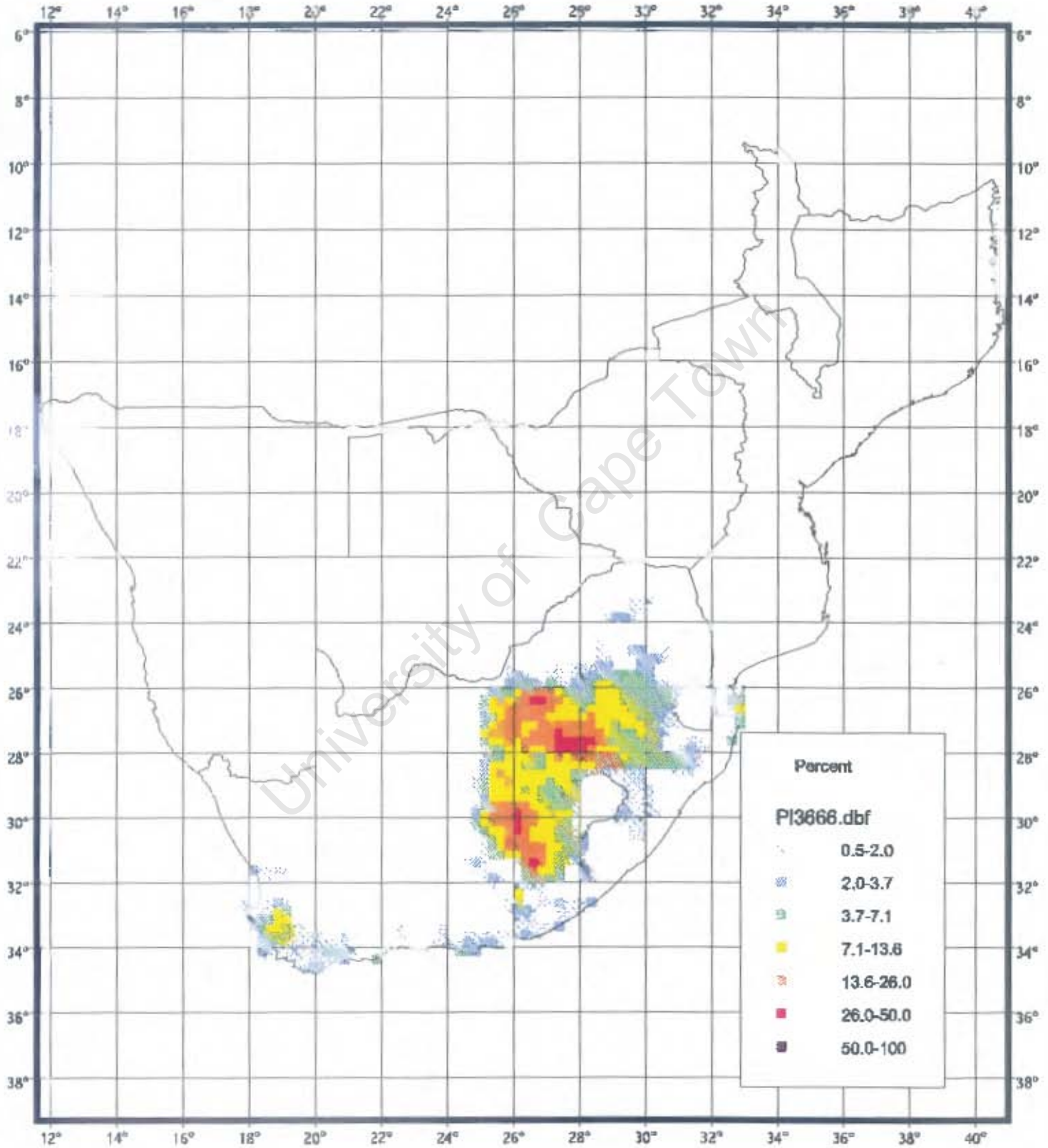


Figure 9: Blue Waxbill - Smoothed distribution using a relative scale for colour shading

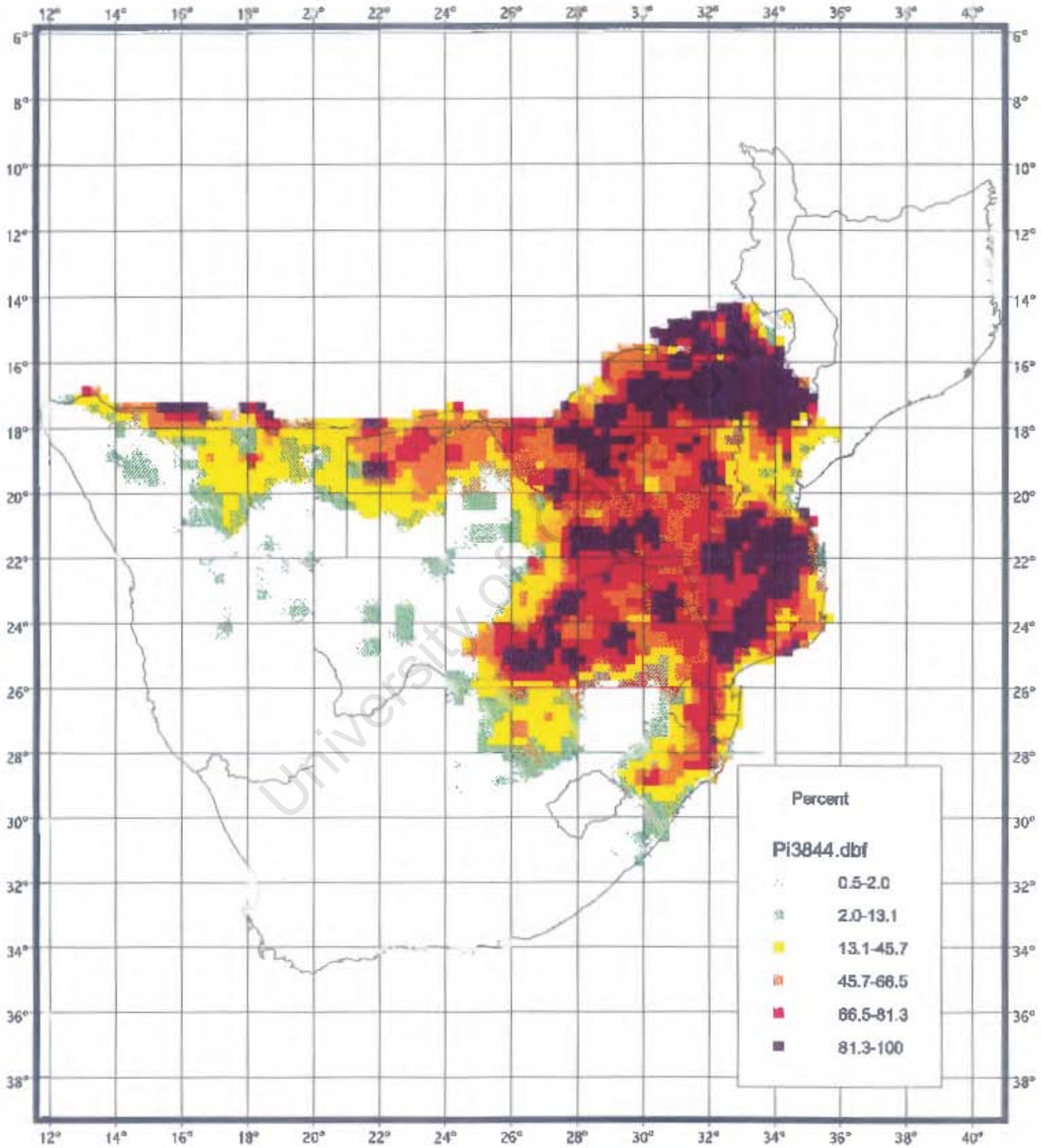


Figure 10: Blue Waxbill - Smoothed distribution using an absolute scale for colour shading

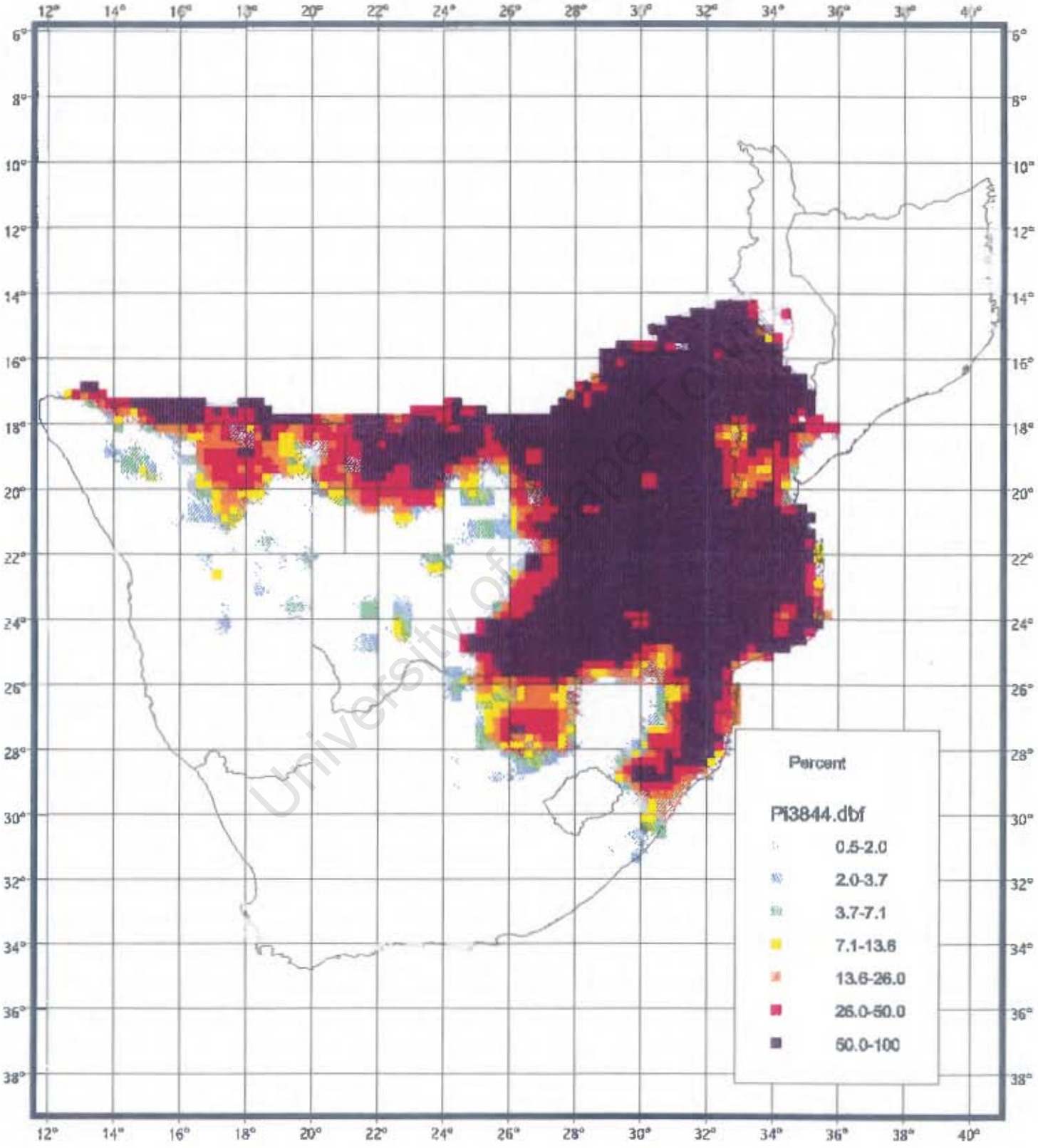


Figure 11a: Cape Weaver - Distribution based on weighted average between observed reporting rates and smoothed detection probabilities using alpha=0.05

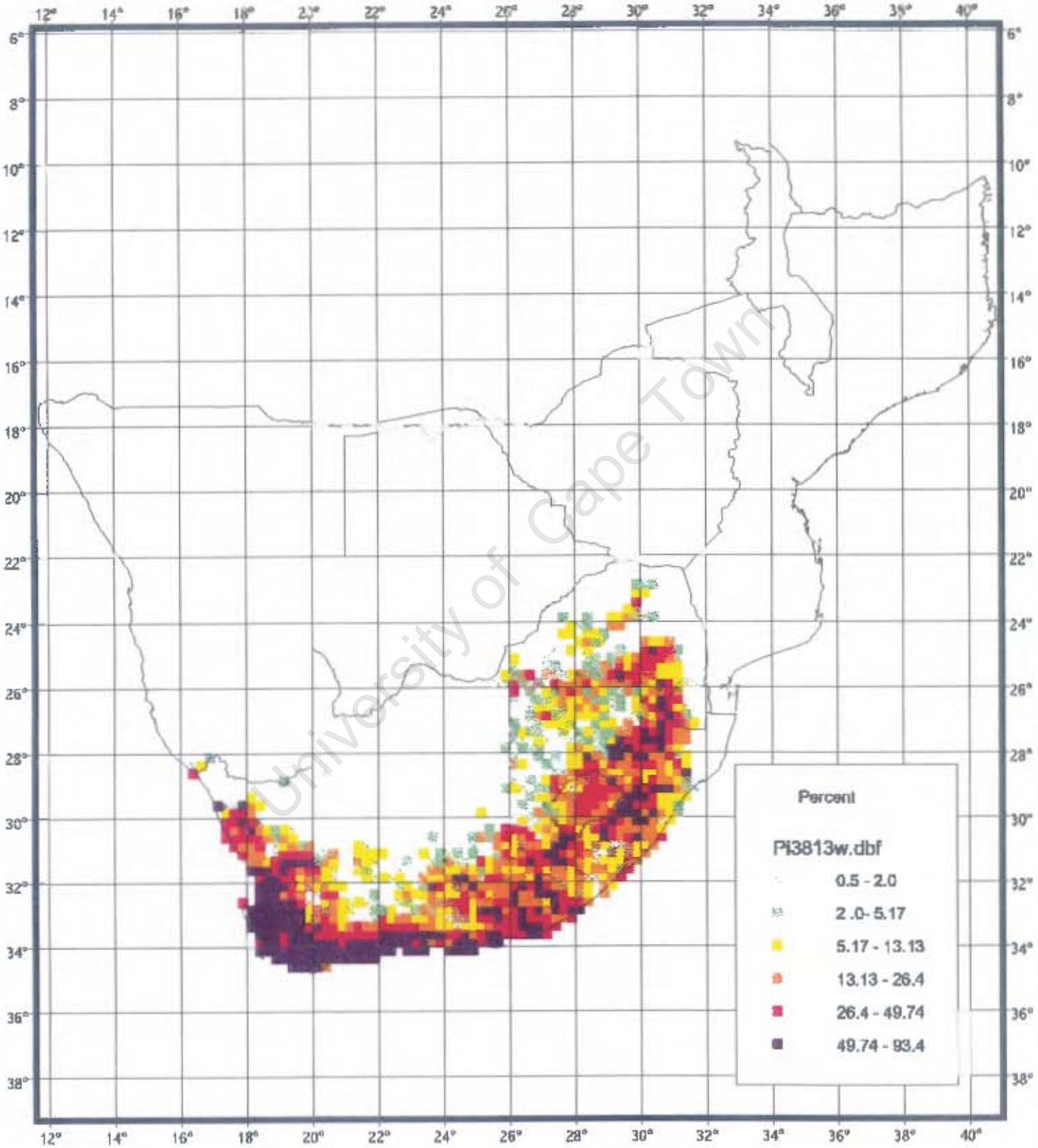


Figure 11b: Cape Weaver - Distribution based on weighted average between observed reporting rates and smoothed detection probabilities using alpha=0.10

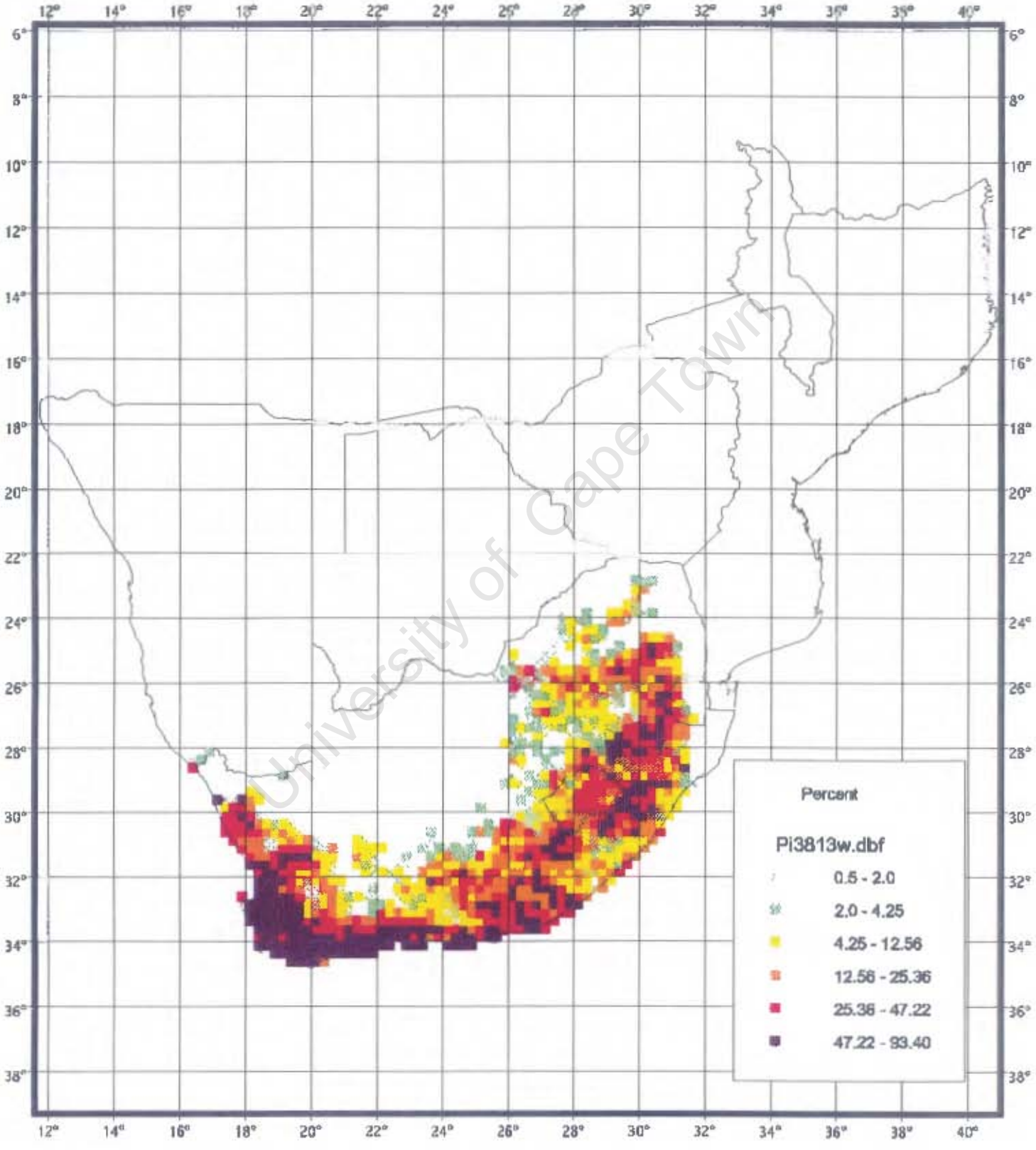


Figure 11c: Cape Weaver - Distribution based on weighted average between observed reporting rates and smoothed detection probabilities using alpha=0.25

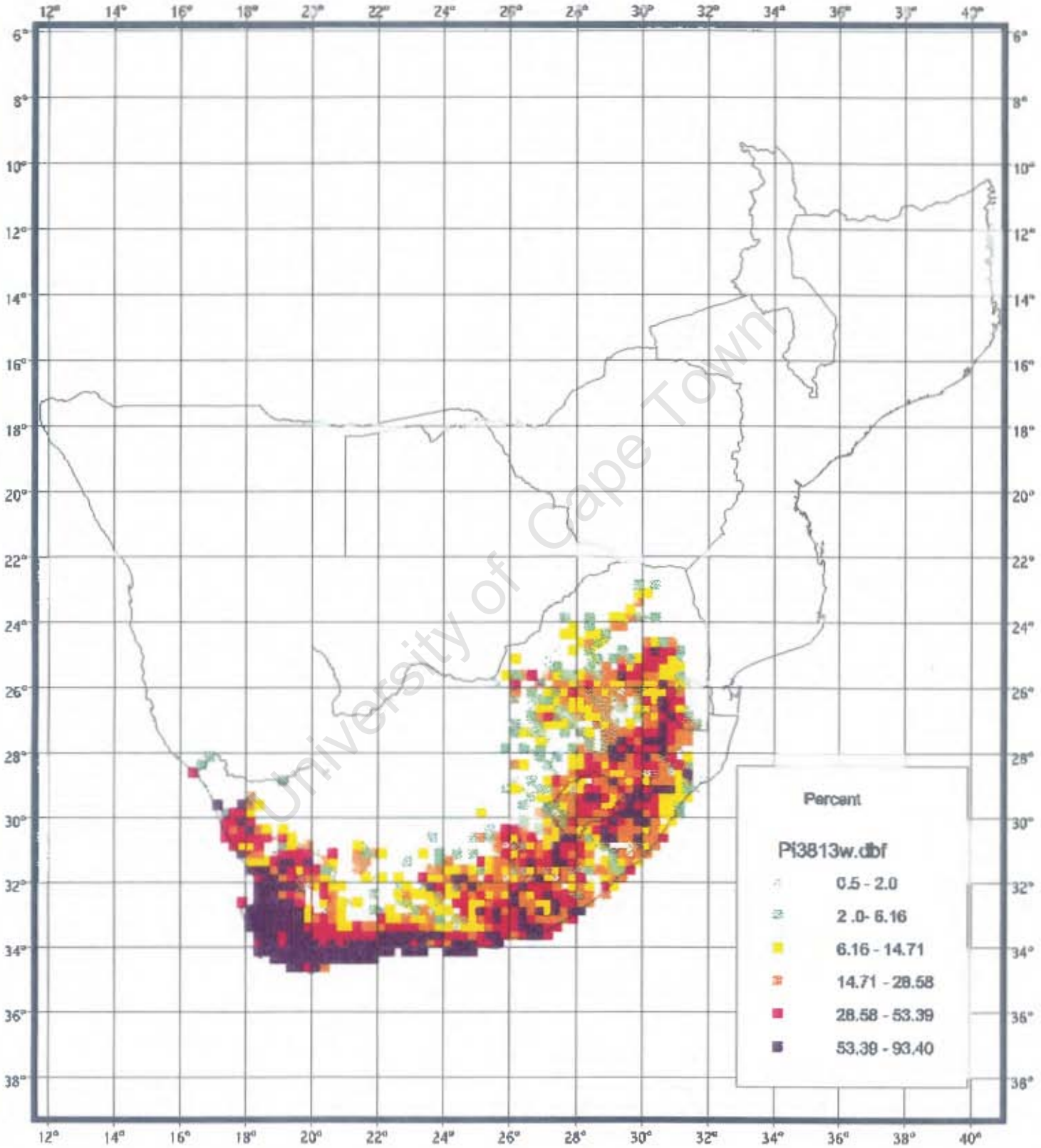


Figure 11d: Cape Weaver - Distribution based on weighted average between observed reporting rates and smoothed detection probabilities using alpha=0.50

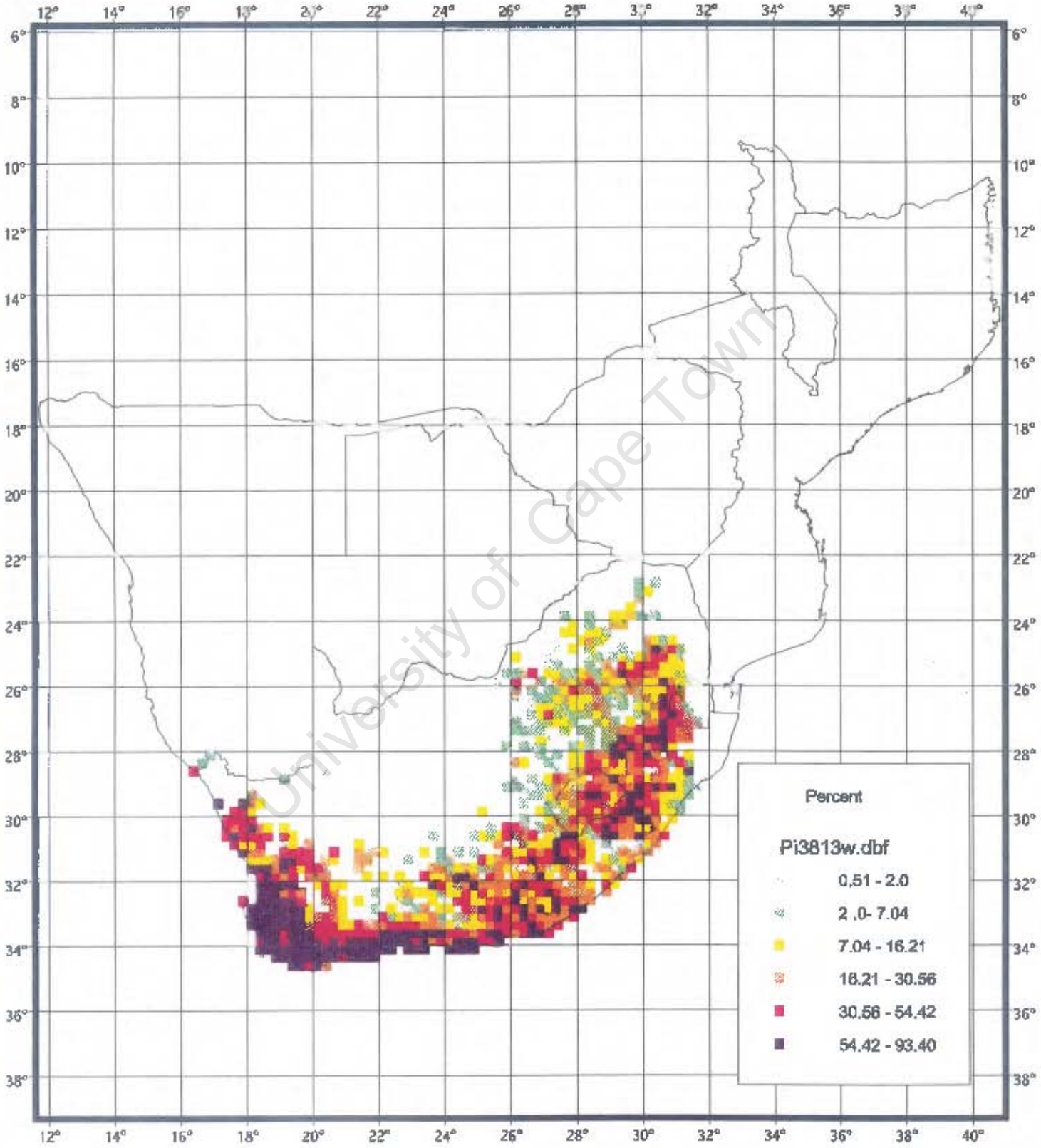


Figure 12a: Whiterumped Swift - Distribution based on weighted average between observed reporting rates and smoothed detection probabilities using alpha=0.50

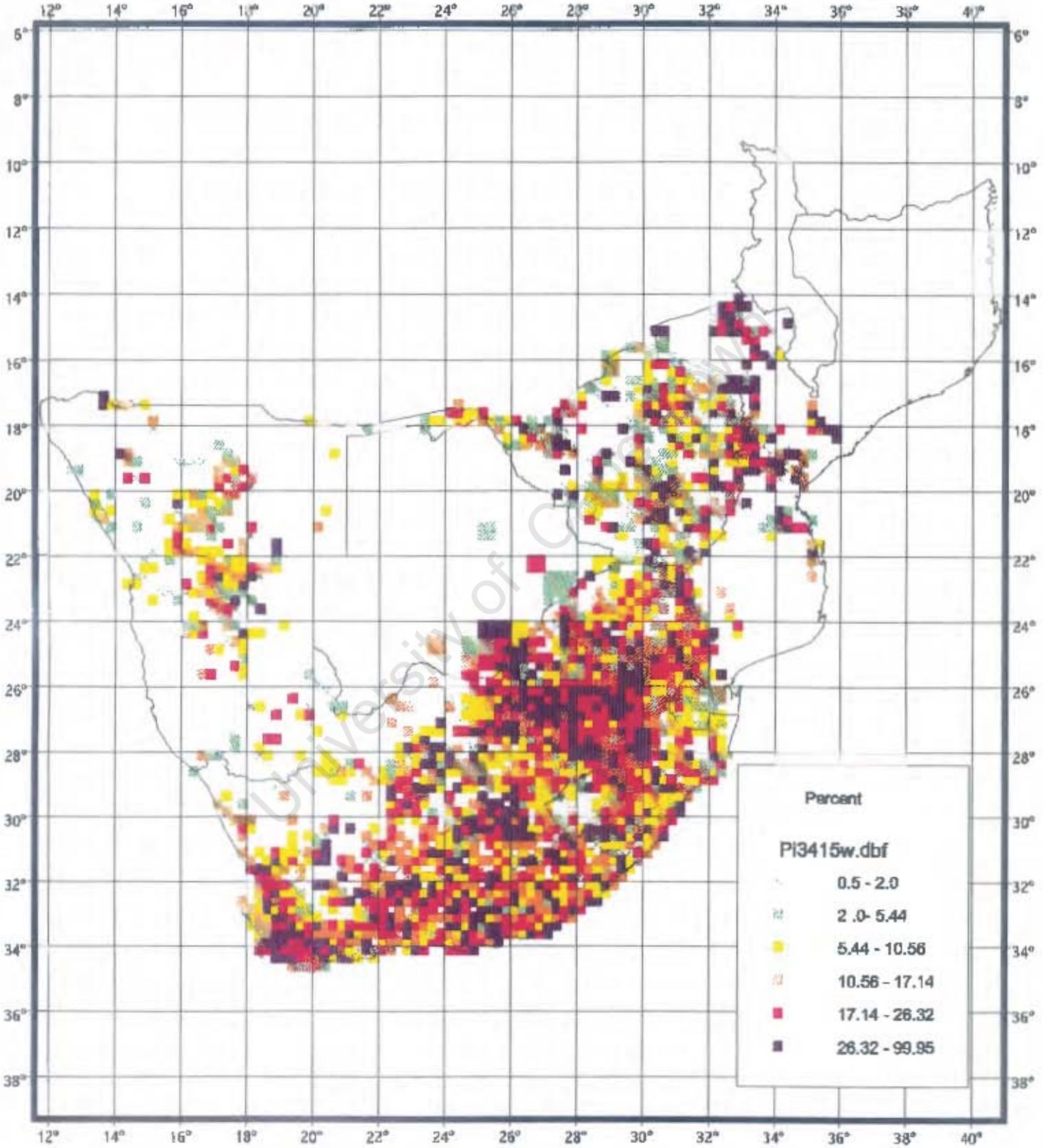


Figure 12b: Whiterumped Swift - Distribution based on weighted average between observed reporting rates and smoothed detection probabilities using alpha=0.05

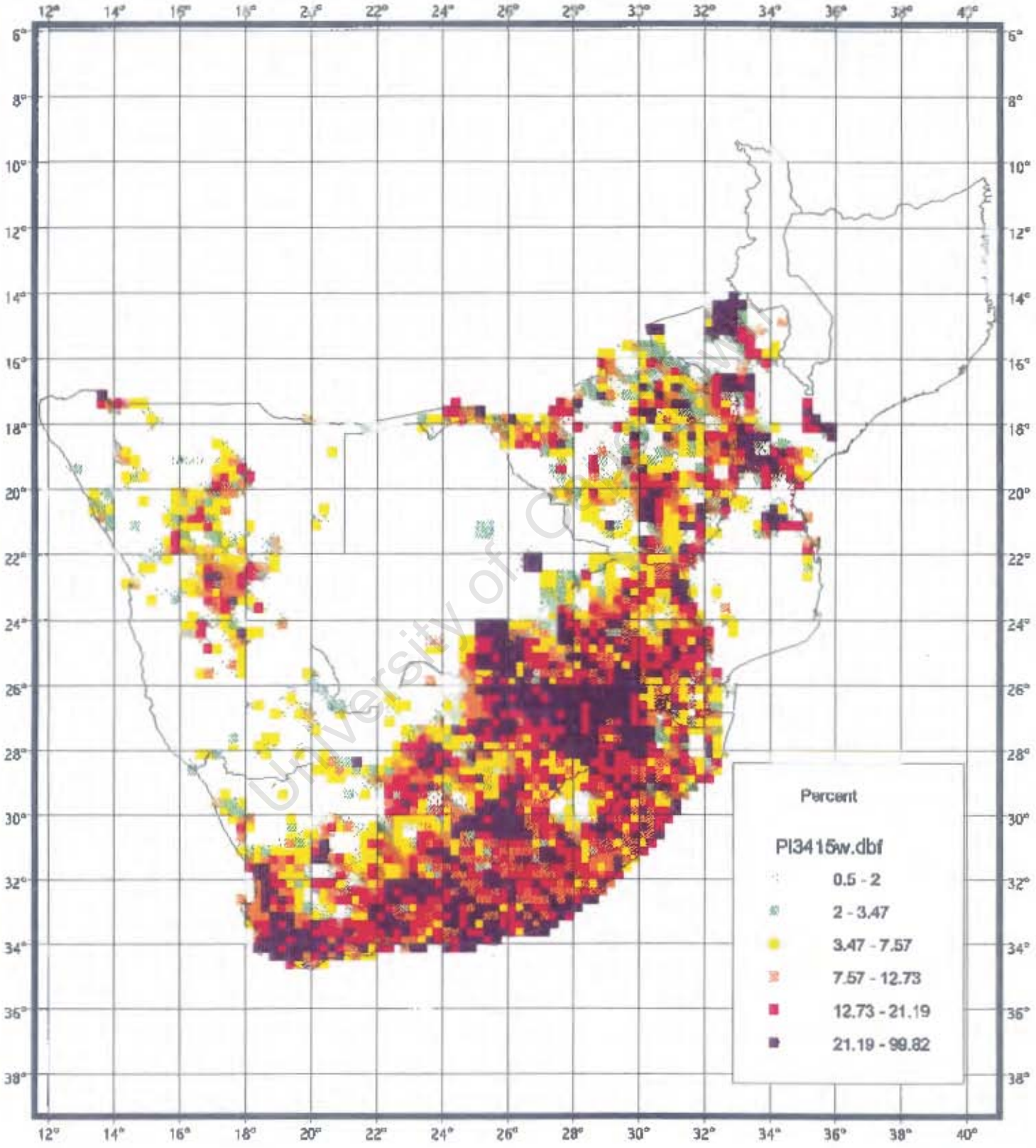


Figure 13: Thrush Nightingale - Distribution of observed reporting rates

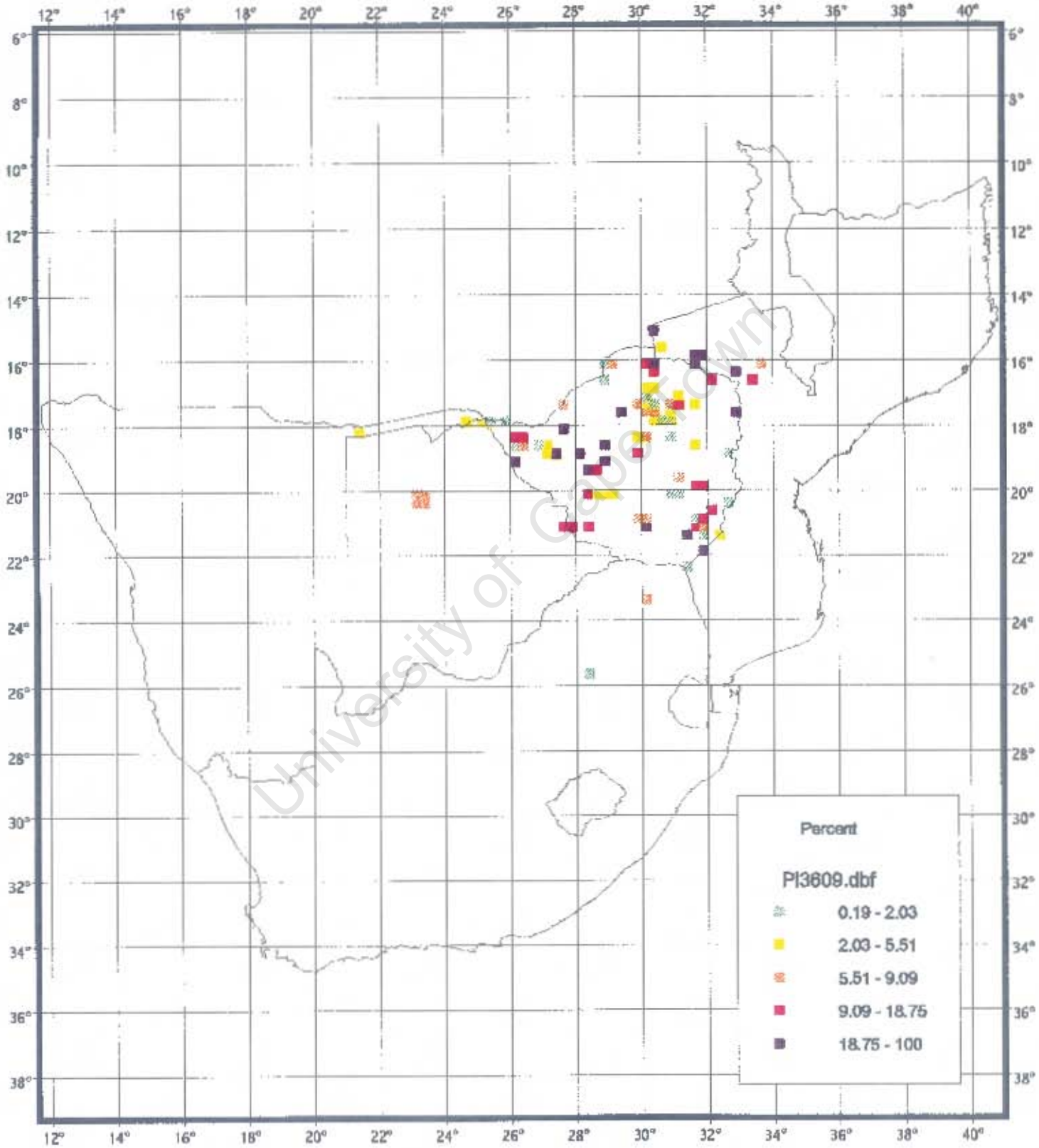


Figure 14: Thrush Nightingale - Smoothed distribution using 0.5% as zero cut-off

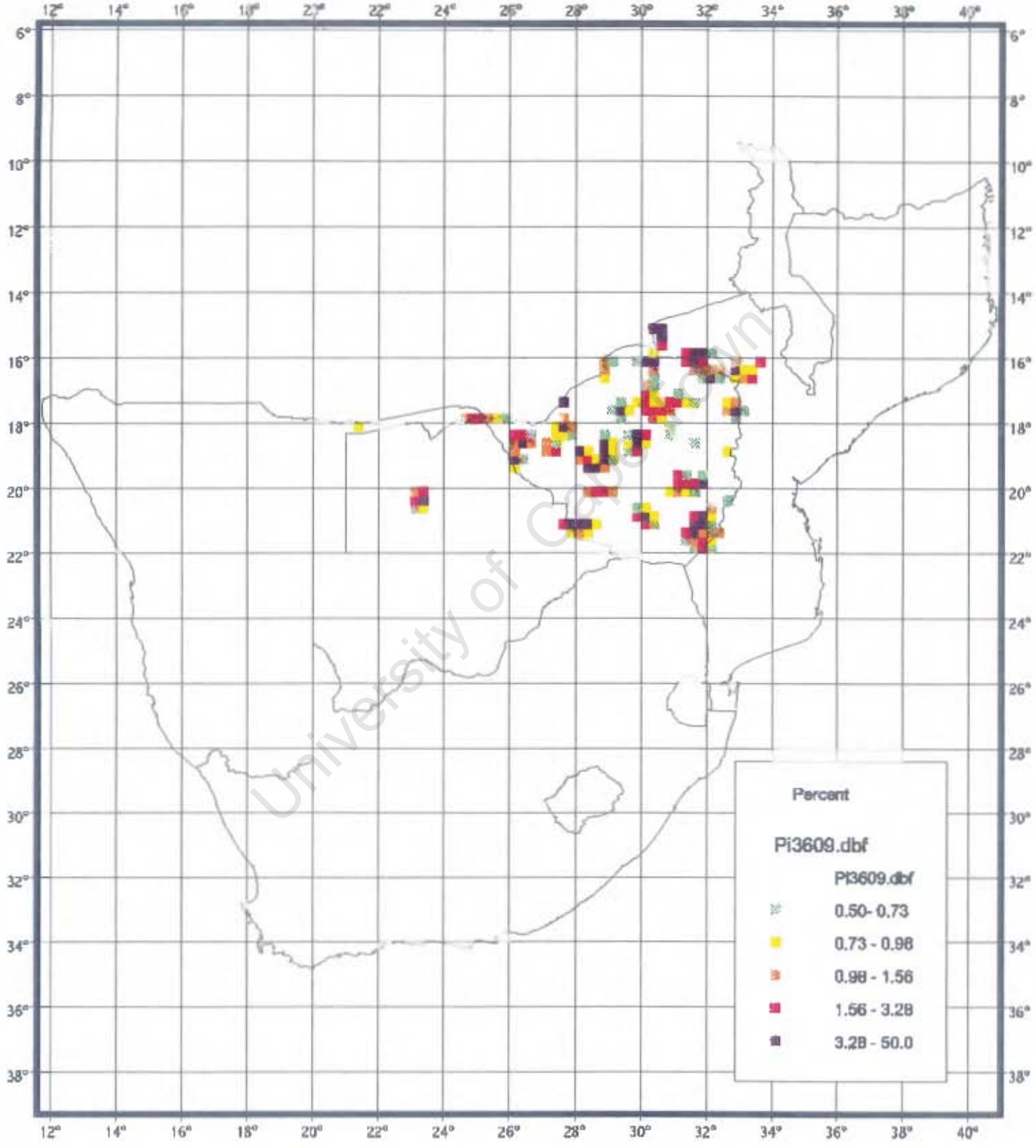


Figure 15: Thrush Nightingale - Smoothed distribution using 0% as zero cut-off

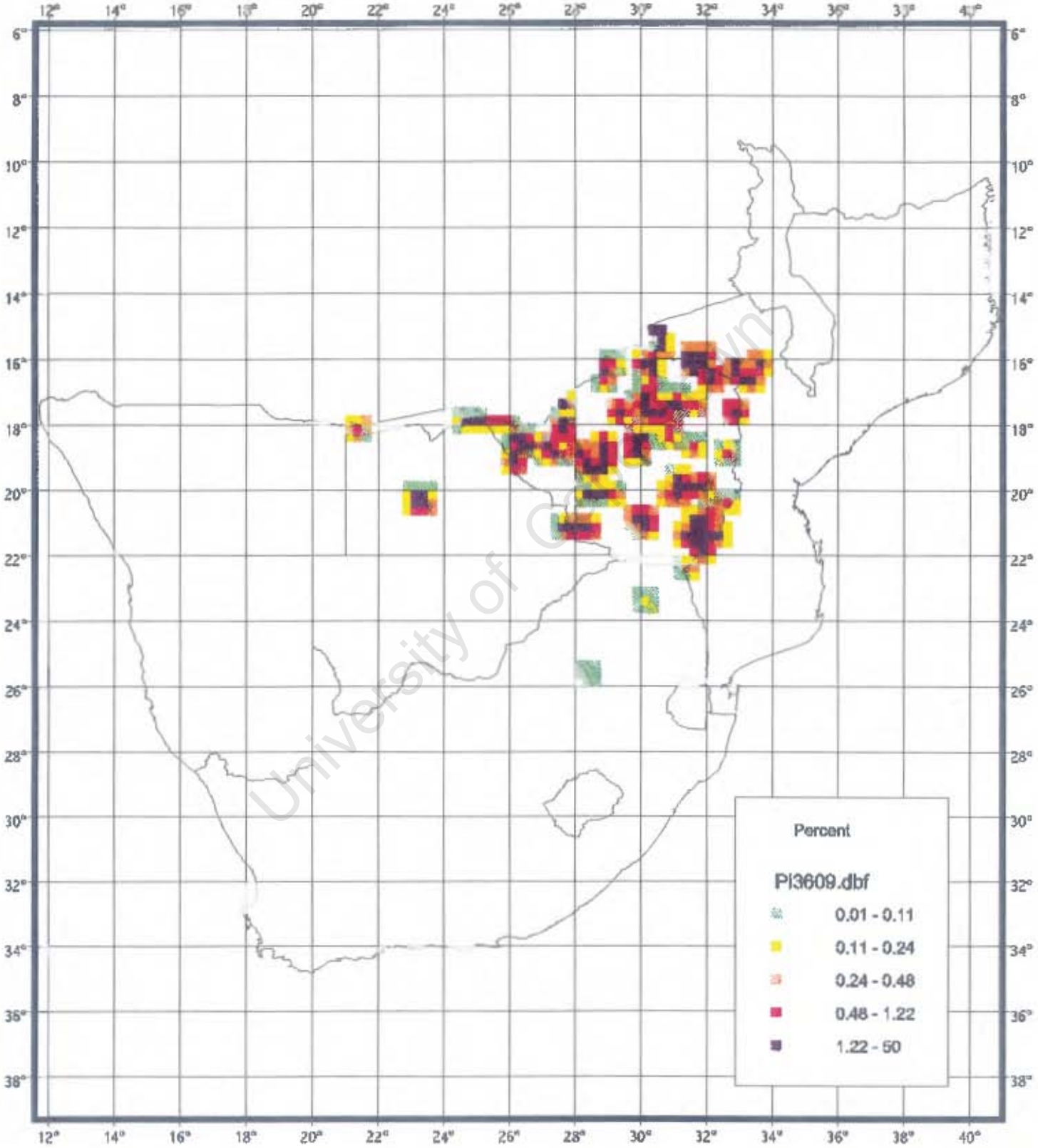


Figure 16: Thrush Nightingale - Smoothed distribution using 0.1% as zero cut-off

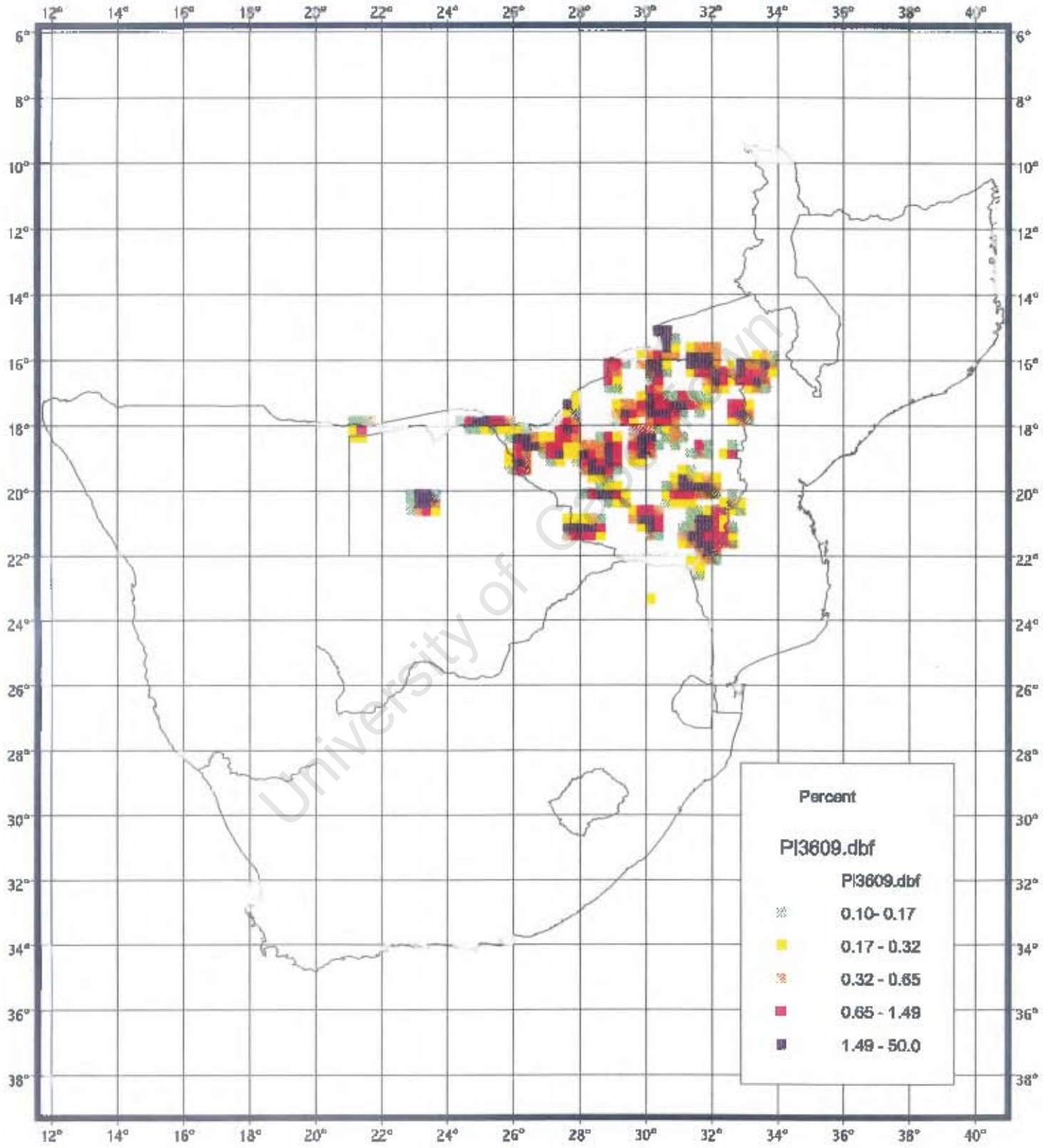


Figure 17: Thrush Nightingale - Smoothed distribution illustrating all probabilities < 2% with an x

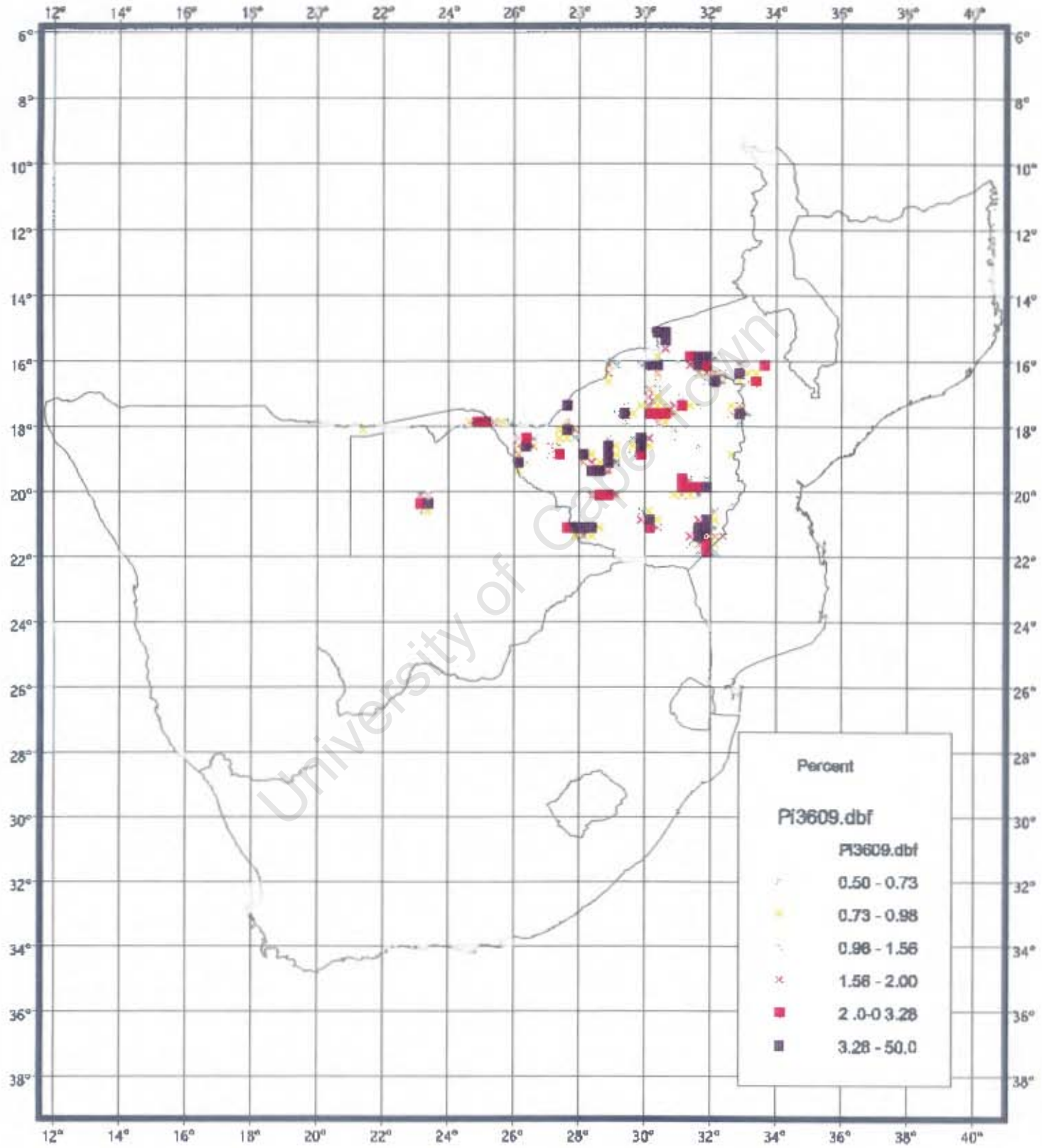
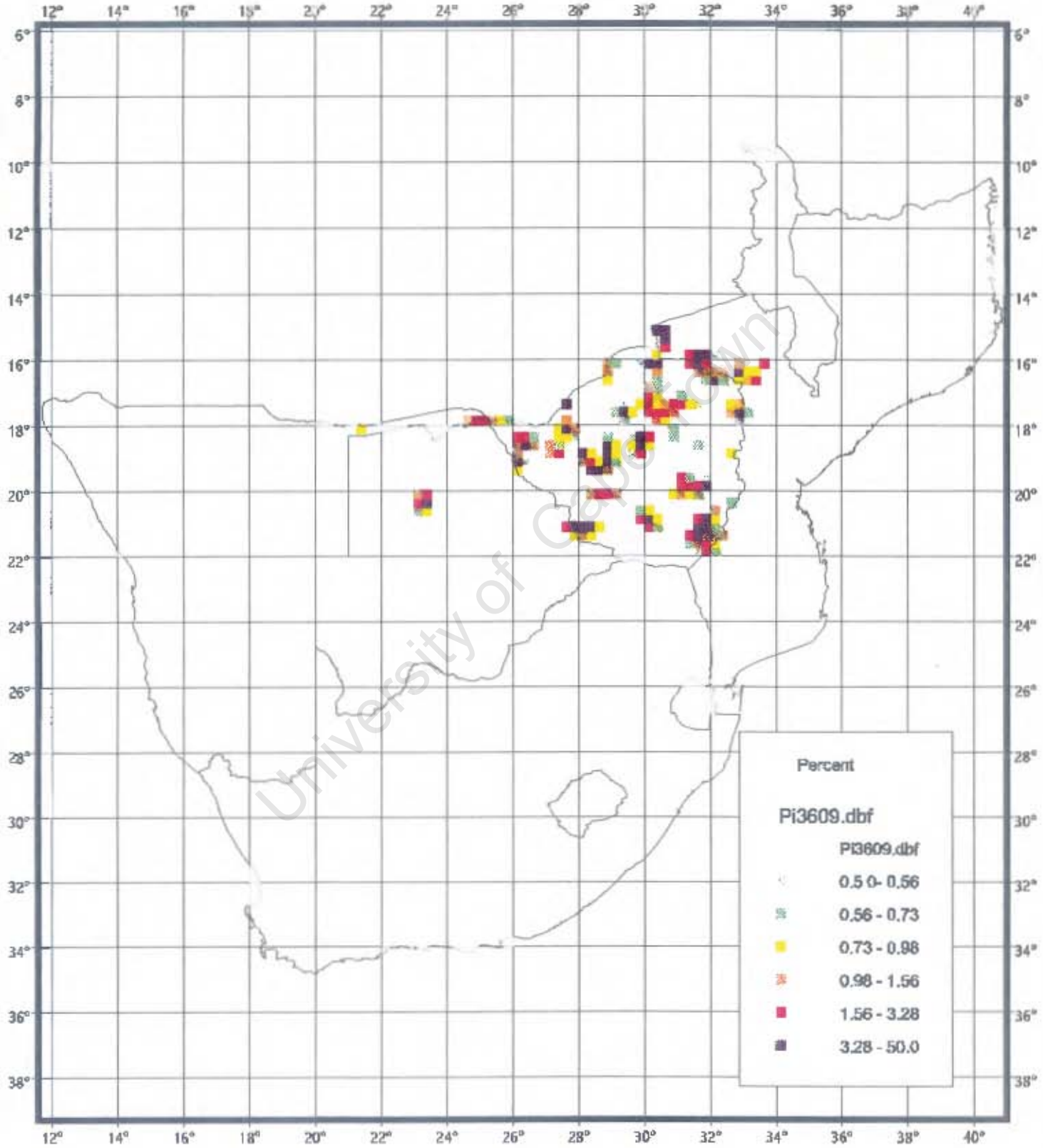


Figure 18: Thrush Nightingale - Smoothed distribution illustrating all probabilities in lower 5% with an x



Chapter 4

The Analysis of Species Richness and Species Endemism Incorporating Reporting Rates from Bird Atlas Data

F. LITTLE AND L.G. UNDERHILL

Avian Demography Unit, Department of Statistical Sciences, University of Cape Town, Rondebosch, 7701 South Africa

Introduction

The identification of patterns of distributions for species forms an important focus of biogeography (Linder 2001). More specifically, interest is often centered around the distribution of endemic species, i.e., species restricted to certain geographical areas. This requires defining the degree to which a species is confined to a certain area, as well as defining the boundaries of the area. To avoid these difficulties, one can look at species with narrow ranges, rather than species that are restricted to a given area (De Klerk 1998). It is often these species that are most at risk of extinction and hence when selecting areas for conservation one would ideally like to include many such range-restricted species, also referred to as "narrow-endemics". Areas that have many such species are known as "hotspots" for narrow endemism or "centres of endemism". Patterns of species richness, without regard to the extent of the species range, simply refer to areas with many species present and measure both diversity and carrying capacity (De Klerk 1998).

Hotspots or centres of endemism, as described above, differ from so-called "areas of endemism" which are defined as areas that have congruent distributions for at least two species of restricted range (Linder 2001). Identifying and selecting areas of endemism require many subjective decisions relating to the degree of congruency, the delimitation of study areas and the degree of restriction of species ranges to these areas.

Most earlier analyses of species distribution patterns have been based on presence-absence data. Lawes and Piper (1998) criticised this approach for several reasons. Among these are the observation that "the occurrence of a species in a given grid cell is not enough to ensure its continued survival" and the fact that presence-absence data do not take sampling effort into account. In their reply to this criticism, Van Jaarsveld *et al.* (1998) argued that before one can contemplate discontinuing the use of presence-absence data, one should provide the better data as well as the analytical methods for analysing these data.

The reporting rates from the bird atlas data replace binary presence-absence data with information regarding the observed reporting rates of a species in a given grid cell. These data are further improved through our smoothing model introduced in Chapter 2 which replaces the observed reporting rates with detection probabilities which take into account information from neighbouring grid cells and thereby reduces errors due to differential sampling effort or observer effects. Our objective here is to initiate an exploration of the benefits of using this range of detection probabilities in comparison with the use of presence-absence data to identify areas rich in species and rich in narrow-endemic species.

Material and Methods

In this chapter we shall use the following notation. The definitions and explanations are given in the following sections.

- r_{ij} = observed reporting rate for species i in grid cell j
- p_{ij} = probability of species i being present in grid cell j
- π_{ij} = smoothed detection probability for species i in grid cell j

- γ_{ij} = detection probability decile for species i in grid cell j
- s_j = species richness for grid cell j based on presence-absence data
- S_j = generalized species richness for grid cell j based on detection probability data
- d_i = range for species i based on presence-absence data as measured by the number of grid cells in which it occurs
- D_i = generalized range for species i based on detection probability data
- A_k = subset of species with $d_i < k$ grid cells
- $z(k)_j$ = species richness in grid cell j with ranges restricted to k grid cells based on presence-absence data
- $S(\alpha)_j$ = number of species in grid cell j in the α%-core of their range
- $Z(\alpha, k)_j$ = number of species in grid cell j with α%-core ranges restricted to k grid cells
- e_j = measure of narrow endemism for grid cell j based on presence-absence data
- E_j = generalized measure of narrow endemism for grid cell j based on detection probability data
- $v(k)_j = z(k)_j/s_j$ = ratio of species richness of 'restricted distribution species' to overall species richness based on presence-absence data
- $V(\alpha, k)_j = Z(\alpha, k)_j/S(\alpha)_j$ = ratio of species richness with restricted α%-core ranges to overall number of species in α%-core of their range
- $w_j = e_j/s_j$ = ratio of narrow endemism to overall species richness based on presence-absence data
- $W_j = E_j/S_j$ = ratio of generalized narrow endemism to generalized overall species richness based on detection probability data

Existing Measures of Species Richness and Endemism based on Presence-Absence Data

Previously, a measure for species richness has simply been a count of the number of species present in a grid cell, i.e., for grid cell j,

$$s_j = \sum_{i=1}^{n} p_{ij}, \tag{1}$$

where $p_{ij} = 1$ if species i is present in grid cell j and $p_{ij} = 0$ if species i is absent in grid cell j.

Linder (2001) summarised existing measures for quantifying richness of range-restricted species per area. He refers to three methods:

1. Consider a species to be "range-restricted" if it occurs in k or fewer grid cells. Count the number of grid cells d_i in which the species occurs:

$$d_i = \sum_{j=1}^{m} p_{ij}. \tag{2}$$

Now let A_k be the subset of species for which $d_i \leq k$. Then for grid cell j determine

$$z(k)_j = \sum_{i \in A_k} p_{ij}, \tag{3}$$

the number of species in a grid cell which are range-restricted in such a way that they occur in k or fewer grid cells.

2. Weight each species by the inverse of its distribution range and sum these inverses for each grid cell, i.e, determine

$$e_j = \sum_{i=1}^{n} \left(\frac{p_{ij}}{d_i} \right), \tag{4}$$

where d_i is defined as in equation (2) and p_{ij} as for equation (1).

3. Linder (2001) mentioned that each of the above two methods can be corrected by the total species richness, thus obtaining "the proportion of the biota that is endemic to the area". By this, we presume he meant that each of the above two measures can be expressed as a proportion of the total species richness in the grid cell giving

$$v(k)_j = \frac{z(k)_j}{s_j} \tag{5}$$

$$\text{and } w_j = \frac{e_j}{s_j}. \tag{6}$$

These measures of species richness and species endemism are derived from presence-absence data and give equal weight to all species in all grid cells for which they are present, whether they occur there in abundance or not and whether the grid cell lies within the centre of their distribution or along the edges. These counts thus include vagrant species far out of their usual ranges and species occurring occasionally near the edges of their distributions. The edges of species distributions are typically arbitrarily defined and fuzzy and it is thus questionable whether a vagrant species or a species at the edge of its range should be counted as a single unit towards species richness. This suggests the need for fractional counts, and the development of generalized measures of species richness which do not have the simple interpretation of being a count of the species which are present in an area. Linder's (2001) measures (3), (4), (5) and (6) are study-area dependent. If the study area is extended, the values of d_i will change for all species that occur in the extension to the study area, and the set of species A_k will change correspondingly.

Bird Atlas Data

The Atlas of Southern African Birds (Harrison *et al.* 1997a, b) contains distribution maps for bird species in southern Africa. These distribution maps are not based on presence-absence data, but on observed reporting rates calculated as the ratio of the number of sightings of a given species in a particular grid cell to the number of checklists for that grid cell. These reporting rates thus have a range between zero and one. Through a smoothing model which reduces sampling errors by making use of information in adjoining grid cells (see Chapter 2), we transform these observed proportions into improved estimates of detection probabilities. We are thus working with data that are better than presence-absence data in two ways: (1) we have available observed reporting rates, and (2) we have reduced sampling bias in these reporting rates through a judicious smoothing model.

We base our analysis of the patterns of distributions for bird species in southern Africa on these smoothed detection probabilities. Other approaches to smoothing may also be developed, and the method described here could also be used with the observed reporting rates. Our study area includes Botswana, Lesotho, southern and central Mozambique, Namibia, South Africa, Swaziland and Zimbabwe (see Chapter 2 for further detail). We will use our proposed measures of species richness and species endemism to generate distribution maps for all terrestrial species in southern Africa. This analysis is purely illustrative, because it does not acknowledge that the ranges of any species extend farther north in Africa. The inclusion of such areas, e.g. northern Mozambique, Zambia and Angola, will alter the results obtained over the entire region. We then identify a subset of species that are endemic and near-endemic to southern Africa and repeat the analysis for this subset. We compare the results from our analyses based on detection probabilities to analyses using binary presence-absence data.

Detection probabilities cannot be compared between species. Some naturally rare species have small detection probabilities even at the core of their range. For others, problems of conspicuousness may result in smaller detection probabilities than for the more conspicuous species (Harrison & Underhill 1997). It is thus

inappropriate to devise generalized measures of species richness using sums of detection probabilities. To take this into account, we replace the actual detection probabilities with their decile values. To achieve this, we divide the entire non-zero range of detection probabilities for a species into 10 intervals with equal numbers of probabilities in each interval. The detection probability for a species in a given grid cell will then fall in one of these 10 intervals. We replace the actual probability value by one of 10 possible numbers 1, 2, ..., 10, depending on whether it lies within the first, second, ...or tenth interval. Finally, we divide these integer values by 10 to convert them back to the probability range between zero and one. In addition we assign a value of zero to grid cells in areas from which the species is absent. The individual detection probabilities have thus been replaced by one of 11 values 0, 0.1, 0.2, ..., 1.0 depending on their relative magnitude within the species probability range. We refer to these as the detection probability deciles γ_{ij}. Clearly, the choice of 10 intervals for doing this transformation is arbitrary, and either more or fewer intervals could be used. Extensive trials have demonstrated that no benefits are obtained by using more than 10 intervals (LGU, unpubl.). The published distribution maps in Harrison *et al.* (1997a,b) employed this principle using three intervals. Equal numbers of grid cells were shaded as light, medium and dark to depict the smallest, middle and largest third of the reporting rates for each species, independent of the actual absolute values of the reporting rates.

There is a critical caveat to the use of this approach. This transformation can only be made if the study area is large enough to include both the core of the range of a species and an area from which it is absent, so as to include considerable variation in the abundance of the species. If the study area is so small that a species is abundant across the entire area, the suggested transformation will result in misleading downweighting of large detection probabilities. For example, if all detection probabilities for a common species in the study area exceed 0.9, say, then it does not make sense to rescale these over a range that covers the interval 0 to 1.

Extending the Measures for Species Richness by incorporating Reporting Rate Concepts

We can replace the probabilities of species presence in equation (1) with the detection probability deciles (γ_{ij}), so that for each grid cell j, we have the generalized species richness

$$S_j = \sum_{i=1}^{n} \gamma_{ij}. \tag{7}$$

This is not a simple count of the number of species present in a given grid cell (see equation 1), but a count where species presence has been weighted by the magnitude of the relative probability of species detection in that grid cell. Large values for this summation will reflect grid cells where a large number of species have detection probabilities that are in the higher decile intervals of their detection probability ranges, i.e, these species are in the core of their distribution ranges within the study area. As such equation (7) can be viewed as an indicator of core concentration and will highlight areas where many species have their cores in common. It is important to realize that S_j is dependent on the study area. Modifying the study area leads to changes in γ_{ij} in (7), and therefore to the generalized species richness.

We can easily convert this range of 11 possible decile values for each grid cell to a binary indicator of species presence or absence. By grouping all non zero values together we get back to the simple presence-absence data where a species is regarded as present in a grid cell as long as its detection probability is non zero, regardless of the actual magnitude of this probability. This is equivalent to reducing the number of intervals until there is only one interval. Mathematically, we can express this as $p(0)_{ij} = 1$ if $\pi_{ij} > 0$, where π_{ij} refers to the smoothed detection probability for species i in grid cell j. Summing these values for each grid cell over all species will result in a simple count of the number of species detected in each grid cell, similar to that referred to in equation (1), but based on smoothed detection probabilities rather than on observed presence-absence, viz.,

$$S(0)_j = \sum_{i=1}^{n} p(0)_{ij}. \tag{8}$$

We can also vary the cut-off value that determines species presence from "greater than zero" to any of the other decile probability values. For example, using $p(20)_{ij} = 1$ if $\gamma_{ij} \geq 0.2$, would put a slightly stricter requirement on identifying species presence, in that it would regard all detection probabilities in the lower 20% of the range as effectively zero. Increasing the cut-off value will result in ever stricter conditions being put on species presence. We could, for example, require that species i should be in the top 50% of its range to be regarded as present for grid cell j, in wich case we will use $p(50)_{ij} = 1$ if $\gamma_{ij} \geq 0.5$. We can then sum these binary values over all species for each grid cell, viz.,

$$S(\alpha)_j = \sum_{i=1}^{n} p(\alpha)_{ij}, \tag{9}$$

where

$$\begin{aligned} p(\alpha)_{ij} &= 1 \qquad \text{if } \gamma_{ij} \geq \frac{\alpha}{100} \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{10}$$

for $0 < \alpha \leq 100$. .

The resulting summations will reflect the number of species in that grid cell that have a detection probability of at least a certain relative magnitude for that grid cell, i.e., the number of species in the core of their distributions where these cores will be increasingly narrow. S_j is not a generalized species richness, but an "ordinary" species richness with less commonly recorded species omitted from the sum. Because it is based on the values of γ_{ij}, it is dependent on the study area, and will change if the study area is modified.

The core of the range of a species is loosely defined as the area where the reporting rates or detection probabilities are largest in relation to the range of reporting rates obtained for the study area. We tighten and quantify this definition by considering the "α%-core" of the range to consist of those grid cells where the detection probability is above the α% percentile of the reporting rates. Thus the "50%-core" is the set of grid cells for a species for which the detection probability is above the median, the 50% percentile. The 0%-core consists of all grid cells where the detection probability is positive, the 20%-core consists of grid cells where the detection probability is above the 20% percentile, etc. The absolute values of the actual detection probabilities defining these cores will differ from species to species. For very conspicuous and highly abundant species, they will typically be large, whereas for the naturally rare or inconspicuous species they will be smaller. Furthermore, the core ranges will change as the study area is augmented by adjacent areas in which the species occurs. Ideally the method should be applied when there are complete distributions available for all species.

Extended Measures of Species Endemism based on Reporting Rate Concepts

The four existing measures of species endemism given in equations (3), (4), (5) and (6) can be extended to incorporate the concept of detection probabilities, or more specifically, their corresponding decile values. The first method is based on determining species richness for each grid cell for a subset of range-restricted species, where the degree of range-restriction is determined by an arbitrary cut-off for the number of grid cells in which a species is present. We have a choice in the way in which we choose to limit the range-size of the species. We can limit the overall range-size, i.e., the number of grid cells for which the species has a non zero detection probability. However, a very small detection probability in a given grid cell implies limited chances of survival for the species in that grid cell. If we base our restricted ranges on this simple presence-absence approach, we will exclude species which have viable probabilities of detection for only a limited number of grid cells. To avoid this, we can specify a cut-off value to define a viable probability that may be required for continued survival and limit the range-size for which a species has this level of detection probability.

Using the decile detection probabilities allows us to put restrictions on both the relative probability of detection and the number of grid cells for which a species reached this minimum relative probability of detection. Equation (10) given above defines the α%-core of the species range. We determine the number of grid cells for which species i has a detection probability in the α%-core of its range as

$$S(\alpha)_i = \sum_{j=1}^{m} p(\alpha)_{ij}. \tag{11}$$

Then we determine for grid cell j, the number of species in the α%-core of their ranges which have restricted ranges such that there are k grid cells for which a species is in its α%-core. Mathematically, this is equivalent to

$$Z(\alpha, k)_j = \sum_{i \in A(\alpha)_k} p(\alpha)_{ij}, \tag{12}$$

where $A(\alpha)_k$ is the subset of species for which $S(\alpha)_i \leq k$, the cut-off limit. Putting α equal to zero equates to the presence-absence approach. Equation (12) thus generalizes equation (3). We shall refer to this measure as the "restricted core-range species count".

A generalized measure of narrow endemism needs to take into account the distribution range of a species and needs to give greater weight to species with narrow distributions. We extend the measure of narrow endemism based on presence-absence data given in equation (4). Firstly, we replace the zero-one values that indicate species absence or presence with one of the 11 detection probability deciles, γ_{ij}. Secondly, we weight this decile value for each species by the inverse of a measure of its distribution range. However, this measure is not simply a count of the number of grid cells in which a species is present. Instead we use a summation of the detection probability deciles over all grid cells for species i,

$$D_i = \sum_{j=1}^{m} \gamma_{ij}. \tag{13}$$

D_i will be large if species i occurs in a large number of grid cells. It will be small if the species occurs in only a few grid cells and has a narrow range. In general D_i will approximately be equal to half the total number of grid cells for which a species is present. This is because of the fact that the γ_{ij} replace the non zero probabilities of detection by one of 10 values between 0.1 and 1. On the average each non zero probability value is thus replaced by 0.5. Summing all the 0.5 values for all the non zero grid cells will thus equal the total number of grid cells where the species is present, divided by two. D_i will be larger than this if the species has detection probabilities in the upper half of its range for a majority of the non zero grid cells and vice versa. The ratio γ_{ij}/D_i will be small if species i has a wide range reflected in the large value for D_i. More importantly, this ratio will be large for species with narrow ranges and relatively large detection probabilities in a given grid cell. The ratio is thus a measure of range-restrictedness for each species that also takes into account the relative probability of species detection in a given grid cell. Large values of this ratio identify grid cells where a restricted range species occurs with a relatively large detection probability.

We combine these species-specific measures of the degree of endemism by summing these measures over all species for each grid cell,

$$E_j = \sum_{i=1}^{n} \left(\frac{\gamma_{ij}}{D_i} \right). \tag{14}$$

This extension of equation (4) gives a measure which takes relative probabilities of detection into account. E_j has large values if grid cell j is the common core of many restricted range species, and may be interpreted as "hotspots" or "centres of endemism".

Finally, in order to generalize equations (5) and (6), we express the measures in equations (12) and (14) as a proportion of the overall species richness. For the restricted core-range species counts, (equation 12), the appropriate measure of overall species richness will be $S(\alpha)_j$, defined by equation (9), so that the ratio which generalizes equation (5) becomes

$$V(\alpha, k)_j = \frac{Z(\alpha, k)_j}{S(\alpha)_j}. \tag{15}$$

For the measure of endemism based on detection probability deciles (equation 14), the appropriate measure to use as denominator will be S_j defined in equation (7), and hence the ratio which generalizes equation (6) is

$$W_j = \frac{E_j}{S_j}. \tag{16}$$

Illustrating the Measures of Species Richness and Narrow Endemism

We illustrate both the measures of species richness and of narrow endemism using colour-coded maps. For this purpose, we subdivide the ranges of these measures into sub-intervals so that we can use different colours for mapping. The sub-division is made at the quintile, so that subranges reflect 20% of the grid cells in the study area. Purple identifies the top 20% of the grid cells with respect to the measure being illustrated, followed by red, orange, yellow and green for the bottom 20% of the grid cells. For the maps showing the distribution of narrow endemism, we further subdivide the top quintile into four quartiles. In this way we can identify the top 5% of grid cells with respect to the measure of narrow endemism. This has traditionally been regarded in the literature as the criterion for defining "hotspots" for narrow endemism (De Klerk 1998).

To help us to identify the sometimes subtle differences between distribution maps based on different types of data and on different measures for species richness and species endemism, we generate maps to highlight these differences visually. To do this, we allocate a value to each grid cell based on its relative shading and thus on the range of values that it represents for the measure being mapped. Thus we let

$F_{lj} = 1$ if grid cell j of map l is shaded green for the lowest quintile of the range,

$F_{lj} = 2$ if grid cell j of map l is shaded yellow for the second lowest quintile of the range,

$F_{lj} = 3$ if grid cell j of map l is shaded orange for the middle quintile of the range,

$F_{lj} = 4$ if grid cell j of map l is shaded red for the second highest quintile of the range,

$F_{lj} = 5$ if grid cell j of map l is shaded purple for the top quintile of the range.

We then subtract these values, generated by the relative shading of the two maps which we wish to compare, from each other for each grid cell. This will generate a new value for each grid cell, $g_j = F_{lj} - F_{kj}$. g_j will equal zero if grid cell j has the same shade in maps l and k. g_j will be less than zero if grid cell j has a darker shade in map k than in map l and g_j will be larger than zero if the reverse holds. The maximum difference for grid cell j will be -4 or $+4$ depending on which map has the darker shade for this grid cell. We allocate colour shades to illustrate these differences. A value of zero that depicts no difference between the two maps is left white. The negative values that depict grid cells for which map k has a darker shade than map l are coloured in shades of blue that get darker as the difference gets larger. The positive values that depict grid cells for which map l has a darker shade than map k are coloured in shades of red that get darker as the difference gets larger. Overall the darker shades of red and blue will thus highlight grid cells for which the two distribution maps differ substantially.

Results

Species Richness for all Terrestrial Species in southern Africa

Overall Species Richness

Fig. 1 illustrates the distribution of generalized species richness for all terrestrial species based on the sum of the decile detection probabilities (equation 7) treating southern Africa as a discrete entity. The darker shades indicate areas where a large number of species have detection probabilities in the higher deciles of their probability ranges, i.e., where they are in the cores of their distributions. This covers a large part of Zimbabwe, extending south into northern parts of South Africa and in a band along the Escarpment to the east coast in KwaZulu-Natal. In addition, the area around the Okavango delta in northern Botswana is highlighted, as well as an area in Mozambique along the Zambezi valley. By contrast, the Northern Cape region and most of Namibia do not have many species with relatively high detection probabilities, in the cores of their ranges.

It is interesting to compare this map to a traditional species richness map (Fig. 2) based on presence-absence data. Here the dark shades identify areas where a large number of species were recorded, regardless of the magnitude of the detection probabilities. Species that occur at small detection probabilities at the peripheral edges of their distributions are thus also counted with the same weight as when they are in the core of their ranges. Differences between the distributions in Figs 1 and 2 are due to subtleties in the patterning of detection probabilities. These differences are highlighted in Fig. 3 which compares the two maps visually. We notice for example, the relatively darker shading of northern Namibia in Fig. 2 compared to Fig. 1. This implies that many of the species counted in this area as present in Fig. 2 occur at relatively low detection probabilities, hence the down-weighting apparent in Fig. 1. The reverse is true for a large part of southern Botswana, where the two types of data generated the most strikingly different results, as illustrated by the darker red shades in Fig. 3. A large area of Botswana has relatively small species richness (Fig. 2). However, in Fig. 1 many of these areas are depicted as relatively rich in species, or more accurately, as areas where many species have their common core concentration. The reason is that the species which do have non zero detection probabilities for these areas tend to have relatively large detection probabilities that indicate that in these areas the species are in the core of their ranges. The same is true for parts of Mozambique, especially in the north and along the Save River valley.

In South Africa, the differences between the two representations are subtle. Two areas where species richness, based on detection probability deciles, lies below that based on presence-absence stand out. One is the interior of the former Transkei in the Eastern Cape. The other is along the Escarpment in Mpumalanga. The Transkei interior is overgrazed with a dense human population. We refer to Underhill (1999), "The 'Transkei hole' is a distinct feature of the distributions of many species in the atlas, especially those that nest on the ground (Allan *et al.* 1997b). The Transkei was compelled to support unsustainably high human densities, and was inevitably ecologically degraded. Although political situations can undergo instant transformations, the environment operates to a different time-scale, especially when it is impacted as hard and as long as happened in the Transkei and the so-called homelands of South Africa, in which 67% of the nation's population lived on 14% of the land surface." The so-called "Transkei hole" did not generally give rise to a complete absence of the species but was apparent as reduced reporting rates. Thus it is encouraging that the generalized species richness map (Fig. 1) detects and displays this subtle result, whereas the species richness map (Fig. 2) does not. Large areas of the Mpumalanga Escarpment are afforested with monocultures of alien trees which has replaced the natural grassland vegetation (see Allan *et al.* 1997a). As in the Transkei, a large proportion of this has resulted in a reduction in the reporting rates for a suite of grassland species, rather than a complete absence. In both cases, the results presented here suggest that the impact of these land use patterns is to reduce detection probabilities for species rather than to render them extinct in these areas.

In many grid cells of the Western Cape, in contrast, species richness based on detection probability deciles

is greater than that based on presence-absence data (Fig. 3). This suggests that the relatively small number of species that do occur in this region, which includes the Fynbos Biome and much of the Karoo, are in the cores of their distributions here.

Species Richness for Core Distributions

Figs 4 to 8 show the species richness distributions for terrestrial species where the criterion for species presence requires an increasingly greater level of relative detection probability. Fig. 4 counts species in the top 80% of their detection probability ranges, or equivalently the 20%-core of their distributions. Fig. 5 counts all species in the 40%-core of their detection ranges. Fig. 6 requires species to fall in the top 50% of their detection ranges, Fig. 7 the top 40% and Fig. 8 the top 20%, or equivalently, the 80%-core of their distributions. Blank cells in these maps indicate that no species have large enough relative detection probabilities to be counted as present for these grid cells.

Comparing Fig. 2 to Figs 4 to 8 essentially compares counts of the number of species present in a given grid cell, regardless of the likelihood of that presence as measured by the detection probability (Fig. 2) with the number of species in each grid cell that fall within an increasingly narrow core of their detection probability ranges. As we increase the level of the relative detection probability that defines the core, areas with many species, shaded in dark purple in Fig. 2, become increasingly lighter in Figs 4 to 8. This indicates that many of the species with non zero detection probabilities for these grid cells, have relatively low detection probabilities. These areas include large parts of Zimbabwe, Mozambique and the east coast of KwaZulu-Natal. At the same time, some lightly shaded areas in Fig. 2 are more darkly shaded in Figs 4 to 8, indicating that though these areas do not have that many species, those which are there, are in the core of their ranges, i.e., they are where they want to be. These areas include grid cells in parts of Botswana, Mozambique and the Western Cape.

Figs 4 to 8 also show the darkly shaded areas getting more fragmented. More unshaded cells appear, reflecting areas where no species reached the respective large detection probabilities required to be counted as present on the different maps. Central and south-western Botswana emerges more and more strongly as an area where relatively few species occur but many of those which do occur there, do so with large detection probabilities. The same trend is observed for the Western Cape. This area corresponds to the Fynbos Biome. We also note the emergence of the grid cells in a curve along 32°S and 22°W, where relatively few species have non zero detection probabilities (Fig. 2) but many of those present in this area are in the core of their distributions (see, for example, Fig. 8). A further interesting trend is the increasingly lighter shading of the area north of 20° S and east of 26° E. This area has a large number of species but increasingly fewer species in the α%-core of their distributions as we increase the level of α.

Though the distribution maps illustrated in Figs 4 to 8 divide the species into presence-absence categories, they do so based on the range of detection probabilities generated by our smoothing model. Fig. 1 which bases species richness on the sum of the decile detection probabilities is a compromise between the species richness distribution based on presence-absence data that gives equal weight to all species regardless of the magnitude of their detection probabilities for a given grid cell, and the species richness distributions of Figs 4 to 8 that require species to be in the core of their distributions to be counted as present. Fig. 1 does not exclude species from the count due to small detection probabilities but down-weights the contribution of these species.

Species Endemism for all Terrestrial Species in southern Africa

In our analysis of species endemism we first concentrate on the measure as given in equation (14) in which species with narrower ranges are weighted larger than species with broad ranges. We determine the distribution of this measure for all terrestrial species treating the study area as an entity, ignoring the fact that

many species have distributions that continue farther north into Africa. Then we repeat it for a subset of terrestrial species known to be endemic or near-endemic to southern Africa. For this subset, we also derive the distributions of the other suggested measures of restricted range species given in equations (12), (15) and (16), and compare the results generated by these different measures.

Fig. 9 shows the distribution of terrestrial species in southern Africa with respect to the degree to which they are "narrow endemics", as defined by equation (14). The dark purple shade indicates the grid cells in the top 20% of the range for this measure. It is informative to compare this distribution with the generalized species richness distribution (Fig. 1). While, at a glance, these maps appear similar, there are important differences. These are highlighted by the comparison of the two distributions (Fig. 10); grid cells shaded blue indicate areas where the quintiles of generalized species richness exceeds the quintiles of apparent narrow endemism. Grid cells shaded red contrast the opposite comparison. Blank areas indicate grid cells in which both characteristics fell within the same quintile. The most striking differences between the distributions of Figs 1 and 9 are in north-western Namibia, an area with relatively few species but with a relatively large number of species with narrow distributions. This is also true for parts of Mozambique. Large parts of the area in the centre of southern Africa are shaded in light blue, indicating an area that scored marginally higher with respect to overall species richness compared to the richness of species with narrow distributions.

We analyse the top-scoring areas in Fig. 9 in more detail. We divide the top quintile into four further intervals, each of which represents 5% of the distribution range (Fig. 11). The grid cells shaded in dark purple thus represent the top 5% of the area with respect to the measure given in equation (14) that takes range-size into account. Table 1 analyses the five top scoring grid cells. We list these grid cells in order of descending magnitude for the endemism measure. For each grid cell, we list the 10 species that contribute most towards each cell's total. We tabulate the species-specific contribution towards the total for these cells, their percentage contribution, their range as measured by the sum of the detection probability deciles, (D_i, equation 13), as well as in terms of the number of grid cells for which they have non zero detection probabilities and we express the latter as a percentage of the total number of grid cells in the study area, again treating the study area as if it were an island. We notice that most of the species listed have very narrow distributions, often covering less than 1% of the total area. This shows the sensitivity of this measure to highlight species with narrow distributions. The majority of the top grid cells in Fig. 11 fall in Mozambique or in eastern Zimbabwe. The species that contribute the most to these cells all have very restricted distributions in southern Africa, but much wider distributions stretching farther north into Africa. Figs 9 and 11 demonstrate clearly how and where species with a distribution centered in more tropical areas to the north of the study area trickle into southern Africa (M. Herremans, pers. comm.).

These analyses have treated the study area as an island that assumes that complete distributions are available for all species. The aim of these analyses has been to define the method that will identify high concentrations of species with narrow distributions. However, to validate our measure of endemism bio-geographically, we repeat our analysis on a subset of species known to be endemic or near-endemic to southern Africa.

Distributions of Southern African Endemic Species

Clancey (1986) identified 152 species to be endemic or near-endemic to southern Africa (Appendix). For these species our study area cover their complete or near-complete distributions. We analyse this subset with respect to the measures given in equations (12) and (14). We also consider the overall species richness for this subset and finally we analyse the relationship between species richness and species endemism by calculating the ratios of these measures as given in equations (15) and (16).

Species Richness for Southern African Endemic Species

Fig. 12 illustrates the distribution of generalized species richness for the 152 southern African endemic species, based on the sum of the detection probability deciles as given in equation (7). Fig. 12 is an analogue of Fig. 1, restricted to southern African endemic species. It shows that a large number of endemic species occur in the Western Cape, extending northwards into the Northern Cape and the Northwest Province, and along the border of South Africa and Botswana, with some isolated areas in Botswana and Namibia. Fig. 13 shows the traditional species richness distribution based on presence-absence data, the analogue of Fig. 2. The grid cells thus reflect a count of the number of endemics with non zero detection probabilities. As for the overall terrestrial species analysis, the striking differences between the two distributions, based on probability deciles and based on presence-absence data respectively (Fig. 14), lie in Botswana. Fig. 13 shows that relatively few endemic species occur in Botswana. However, when we take the detection probability deciles into account, large areas of Botswana lie in the top quintile of the distribution of generalized species richness (Fig. 12). The relative importance of the Northern Cape also increases when using the data based on detection probabilities. In contrast, large parts of Namibia, the Northwest and Limpopo Provinces and KwaZulu-Natal decrease in relative importance when detection probabilities are taken into account, compared to when we simply count the number of species with non zero detection probabilities.

Centres of Endemism

Fig. 15 shows the distribution for the subset of southern African endemic species based on the measure given in equation (14) which weights the detection probability deciles for each species by the inverse of a generalized measure of range-size. The differences between Fig. 15 and Fig. 9, which treats the study area as an island, are enormous. The previously high-scoring areas in Zimbabwe, Botswana and Mozambique (Fig. 9) now fall within the lower ranges of the distribution of narrow endemic species, except for small isolated areas. As might be anticipated, the southern parts of the study area contain most of Clancey's endemic species.

It is also interesting to look at the changes in the distribution maps as we move from (1) a simple count of the number of endemic species with non zero detection probabilities (Fig. 13), to (2) weighting this count by the detection probability deciles (Fig. 12), to (3) weighting these detection probability deciles by the inverse of the range-size (Fig. 15). The main difference between Figs 13 and 12 (shown in Fig. 14) is the effect of down-weighting the peripheral edges of the species distributions and concentrating on the location of the cores of these distributions. The difference between Figs 12 and 15 (shown in Fig. 16) is the location of the core distributions of all endemic species to those with restricted ranges. We discuss these changes for different parts of the study area. Table 2 analyses the species composition of some of the high-scoring grid cells in Fig. 15. We selected the grid cells for this table based on their relative ranking in the distribution underlying the map, as well as to include all high-scoring geographical areas.

Namibia has relatively many endemic species in the central area north of 25°S but few species along the southwest coast (Fig. 13). Once we take into account the detection probability deciles (Fig. 12), we find that few of the species have the core of their ranges in this area. As we put more weight on the extent of the species range (Fig. 15), we note that a few narrow endemics have their cores along the southwest coast, north of 26°S. Grid cell 2615AA in Table 2 indicates that this area is dominated by the presence of the Dune Lark *Certhilauda erythrochlamys*.

The Northern Cape has a relatively large number of endemic species as can be seen by the red shading of this area in Fig. 13. A number of these species occur in the core of their distributions and many of these have narrow distributions. This is evident in the increasing relative importance of this area as we move from Fig. 13 to Fig. 12 to Fig. 15. The narrow endemic species with core distributions in this area include Sclater's Lark *Spizocorys sclateri* and the Red Lark *Certhilauda burra* (Table 2, grid cells 3020BD and 3020AD).

Similarly, the Western Cape has a relatively large number of endemic species, with many in the core of their ranges and with narrow distributions. Table 2 lists six species for the grid cells representative of this area (3418BD, 3418AB, 3322BD), all of which are characteristic of the Fynbos Biome.

The Karoo is an area with relatively many species, many of which are in the core of their distributions and many which have narrow distributions, though there is a marginal decline in the relative importance of this area as we move from a simple count, to core distributions, to core distributions of range-restricted species. The Eastern Province and KwaZulu-Natal do not have many endemic species, except in the Drakensberg area. The area's relative importance decreases further when we focus on core distributions, indicating that the species which do occur here, are not in the core of their distributions. However, when we take range-size into account, this entire area becomes very important and extends further north along the Escarpment. The same pattern is seen for the distribution that takes range size into account but is based on presence-absence data (equation 4) and as such does not concentrate on core distributions (Fig. 17). It follows that this is an area where many endemic species with narrow distributions occur and they can be, but are not necessarily, in the core of their distributions. These species include the Spotted Thrush *Zoothera guttata*, Brown Robin *Erythropygia signata*, Knysna Woodpecker *Campethera notata*, Knysna Warbler *Bradypterus sylvaticus* in the Eastern Cape (grid cells 3228DA, 3228BD in Table 2); the Shorttailed Pipit *Anthus brachyurus*, Gurney's Sugarbird *Promerops gurneyi*, Bush Blackcap *Lioptilus nigricapillus* and Bearded Vulture *Gypaetus barbatus* in the Drakensberg (grid cells 2829CC and 2629DD); Green Barbet *Cryptolybia woodwardi* in the KwaZulu-Natal grid cells (2831DC and 2831DD); and the Yellowbreasted Pipit *Hemimacronyx chloris*, Rudd's Lark *Heteromirafra ruddi* and Botha's Lark *Spizocorys fringillaris* in grid cells 3127DC, 3028BD, 2730AC, 2730AA, 2729BB, 2630CC, 2629DD.

The Northwest and Limpopo provinces have a large number of endemic species (Fig. 13) but fewer of these species are in the core of their distributions (Fig. 12) and even fewer have restricted ranges (Fig. 15). Botswana has relatively few endemics (Fig. 12) but many of the species that do occur here have relatively large detection probabilities, indicating that they are in the core of their distributions. This is especially true for the area around the Okavango delta, a diagonal strip from 21°S, 21°E to 24°S, 24°E, most of the area between 24°E, 26°E and 22°S, 26°S, as well as an area in the southwest along the border with Namibia. When we take range size into account, the relative importance of these areas decreases marginally. A closer inspection of the species composition for some of this area (specifically for grid cells 2021CB and 2021CD, not shown in Table 2) shows that the species with non zero detection probabilities in this area include Orange River Francolin *Francolinus levaillantoides*, Tinkling Cisticola *Cisticola rufilata*, Barred Warbler *Calamonastes fasciolatus*, Cape Penduline Tit *Anthoscopus minutus* and Clapper Lark *Mirafra apiata*. The ranges of these species are between 850 and 1900 grid cells, larger than the range sizes for the species listed in Table 2. However, all of these species have detection probabilities that fall into the top decile of their detection probability ranges in this area. Thus, though some of the species in these areas do have relatively narrow distributions, their large relative detection probabilities are the key feature of their distribution in Botswana.

Most of the species that occur in Zimbabwe are not southern African endemics. This can be seen by comparing the species richness for all terrestrial species (Fig. 1) with the species richness for the southern African endemic subset (Fig. 13). The endemic species that do occur in Zimbabwe, do so with relatively small probabilities and are not species with narrow distributions. Thus this is an area that reflects the edges of wider distributions for species with cores elsewhere in the study area. The exception is the Eastern Highlands which contains a significant number of narrow endemic species in the core of their distributions: Starred Robin *Pogonocichla stellata*, Olive Bush Shrike *Telophorus olivaceus*, Barratt's Warbler *Bradypterus barratti*, Striped Pipit *Anthus lineiventris*, Gurney's Sugarbird *Promerops gurneyi* and the Cape Batis *Batis capensis* (grid cells 1932BB and 1932BA in Table 2). The same comments hold for Mozambique where

only between 1 and 32 species had positive detection probabilities for the individual grid cells (Fig. 13). However, there are two isolated areas within Mozambique where range-restricted species have the core of their distributions (Fig. 15). The Shorttailed Pipit *Anthus brachyurus* has the core of its very restricted distribution in some of the grid cells along the Zambezi valley (grid cell 1835DA in Table 2).

A further comparison of interest is between the endemism distribution based on the range of detection probabilities (Fig. 15) and the endemism distribution based on presence-absence data (Fig. 17). Once again there are significant differences, highlighted in Fig. 18. The presence-absence data seem to generate a smoother map which is particularly noticeable in Botswana. Fig. 18 illustrates that in Botswana the measure that uses a range of detection probabilities based on the reporting rate data highlights regions that scored relatively low values when using the presence-absence data. By contrast, most of Namibia and large parts in the north-east of South Africa score higher when using the presence-absence data than when using the detection probability deciles. As briefly mentioned above, when discussing the distribution patterns in Botswana, both these maps highlight range-restricted species, but Fig. 15 in addition focuses on the core of the narrow distributions of these species.

An analysis of the species composition of some of the grid cells in areas which score relatively highly in these maps highlights the impact of using different measures and different types of data (detection probability deciles versus presence-absence). Figs 19 and 20 illustrate the top 20% of the endemism distribution based on a range of detection probabilities and on presence-absence data, respectively. Fig. 21 visually compares these two maps. Tables 2 and 3 analyse the species composition of some of these high-scoring grid cells. We selected the grid cells to include in these tables based on their relative ranking in the distributions underlying these maps, as well as to include all high-scoring geographical areas. Both tables include the same grid cells to facilitate comparison.

The differences between the measure of endemism based on detection probability deciles (equation 14) and the measure of endemism based on presence-absence data (equation 4) are

(1) for presence-absence data the numerator stays constant at 1 for those species that contribute towards the sum, while when using a range of detection probability deciles, the numerator varies between 0.1 and 1.0.

(2) using the sum of the detection probability deciles to measure the extent of the species range shrinks the denominator because it downweights the cells with relatively small detection probabilities. These two features combined have the effect of narrowing the species distributions and concentrating on the grid cells with high relative probabilities within the range for that species. This effect is clearly seen in the differences between the ranking and composition of grid cells in Tables 2 and 3.

The top-scoring grid cell for both Figs 19 and 20 is grid cell 2831DC. This cell is dominated by the Green Barbet that has a non zero detection probability for only two grid cells in the study area. The detection probabilities for this species in grid cell 2831DC is 0.0484 and in cell 2831DD is 0.0067. We notice how the sum of the detection probability deciles has narrowed the effective range of this species from two to 1.2. When using the range of detection probabilities (Fig. 19 and Table 2) only the cell with high probability is ranked among the top endemism cells, while the analysis based on presence-absence data (Fig. 21 and Table 3) treats the two cells similarly, regardless of the different probabilities of detecting the species in these cells.

Grid cell 3028BD is ranked second in importance in Table 2 (based on detection probability deciles) but only 11th in Table 3 (based on presence-absence data). The main narrow endemic species in this cell are Rudd's Lark and the Yellowbreasted Pipit. Rudd's and Botha's Larks account for the large endemic values of cells 2730AA, 2730AC and 2630CC. These species have different contributions to the endemism totals for

these cells for the analysis based on the range of probability values, but the same constant contribution when using presence-absence data. This lack of discrimination between cells where species are present regardless of relative probabilities of detection account for the smoother distribution of endemism when using presence-absence data. It also means that other grid cells in other areas with different narrow endemic species are pushed lower down the relative rank of importance, with the result that they may not fall within the top percentiles of the distribution. When using a percentile cut-off to identify so-called endemic "hotspots", these cells may then be omitted. This is clearly illustrated in Fig. 21, for the area in the Northern Cape around grid cells 3020AD and 3020BD, which form the centre of the narrow distributions for Sclater's Lark and the Red Lark. The analysis based on presence-absence data reflects the relatively narrow distribution of these species but these grid cells are only ranked in the third 5% sub-interval, as opposed to the top 5%. The same is true for high-scoring grid cells in the Eastern Highlands of Zimbabwe where the Starred Robin, Olive Bush Shrike and Barratt's Warbler have the cores of their relatively narrow distributions. Also in grid cell 1615AA along the west coast of Namibia, the core of the very narrow distribution of the Dune Lark, is not easily picked up by the analysis based on presence-absence data.

It is thus clear how incorporating a range of probabilities derived from a continuous smooth scale of detection probabilities increases the sensitivy of the measure of narrow endemism. This allows us to identify areas of narrow endemism based on the core ranges of species with narrow distributions.

Centres of Endemism using the Restricted Core-Range Species Counts

We return to the measures of endemism that restrict the distribution ranges of species explicitly, as given in equations (11) and (12). Recall that the data based on reporting rates enable us to also specify the level of the probability of detection and hence the so-called percentage core of the species distributions. We look at different combinations of cut-off values that define the core of the distribution and limit the extent of the range. We look at species in the 80%-core, 50%-core and 20%-core of their distributions, as well as the presence-absence approach where any species with a non zero detection probability is considered as present. Our study area covered 5057 grid cells. We restrict our attention to those species which have firstly a core range less than 500 grid cells (roughly 10% of the study area) and secondly a core range less than 250 grid cells (roughly 5% of the study area). As we increase the percentage specification to define the core and decrease the range cut-off limit, we focus more and more on species with very narrow distributions in the core of these distributions.

Fig. 22 illustrates the overall unrestricted distributions, as well as the restricted-range distributions. The trend as we move down the page concentrates more and more on the "core" and increasingly highlights grid cells where species can be detected with relatively high probability. The change as we increase the precentage-core specification is substantial. For the overall unrestricted distributions the original high numbers of species in most of South Africa and large parts of Namibia decrease as we increase the core specification. The eastern area fades away and Botswana, the Northern Cape and parts of the Western Cape are highlighted. These are areas where endemic species occur with high relative probabilities of detection. When we restrict the species range to fewer than 500 and fewer than 250 grid cells, the maps based on the presence-absence approach highlight the coastal regions along the west, south and east, continuing northwards along the Escarpment, as well as the Eastern Highlands of Zimbabwe. The maps based on the 20%-core have larger areas shaded and count more species, because some species which have wider distributions when counting all grid cells with non zero probability now have fewer than either 500 or 250 cells in which they are in their 20%-cores. Hence areas that are left blank in the presence-absence maps, for example Botswana, are shaded in the 20% core maps. Species which occur in these areas have wide overall distributions but much smaller areas where they occur with relatively large probabilities.

The choice of which combination of core-specification and range-restriction to use depends on the objective of the analysis. It is interesting to note that the maps based on the 20%-core distributions correspond best

to the distribution of endemism based on the continuous measure given in equation (14) and illustrated in Figs 15 and 19. This adds to the interpretation of that measure in that it seems to highlight areas where species occur with detection probabilities at least in the 20%-core of their ranges and which have these cores restricted to between 10% and 5% of the entire study area.

One could argue that a detection probability in the 20%-core range should be sufficient to ensure continued survival. Gaston & Rodrigues (in press) considered "peaks of abundance" to be grid cells "with reporting rates within 80% of their maximum value observed for a given species" and range-restricted species as "those present in less than 8.8% of the 1858 grid cells". We thus analyse the distribution of species for which the 20%-core was restricted to fewer than 500 grid cells, less than 10% of the study area. This distribution map is illustrated in Fig. 23 using a larger scale. The areas shaded in red and purple on this map have between 4 and 20 species in the core of their distributions with these cores restricted to fewer than 500 grid cells. These areas correspond reasonably well to the purple shaded areas in Fig. 15 which is based on the more continuous measure of endemism given in equation (14). Differences between these two distributions are due to the dependence of the shading in Fig. 23 on the actual count of species, whereas if a grid cell contains even only one species with a relatively high detection probability and a relatively narrow range, it can have a high score for the endemism measure mapped in Fig. 15. Table 2 shows for example, how the relatively high score in Fig. 15 for grid cell 2615AA along the west coast of Namibia is due to the contribution of the Dune Lark, which contributes 91% towards the total endemism sum for this grid cell. Similarly, the Shorttailed Pipit, accounts for 99.7% of the total endemism score in Fig. 15 for grid cell 1835DA in Mozambique.

Table 4 lists the species composition of some of the top scoring grid cells in Fig. 23. Most of these grid cells are in the Western Cape and include all the Fynbos species. We list two of these grid cells in Table 4. These are followed by two grid cells in the Drakensberg area. The species listed for these grid cells correspond to the species that contributed significantly to the continous measure of endemism (equation 14) for grid cells in this area (for example, 2829CC), mapped in Fig. 15 and analysed in Table 2. Rudd's Lark, Botha's Lark and the Yellowbreasted Pipit are again the restricted range species for grid cells 2730AC and 3127DC. In the grid cell towards the northern end of the Escarpment (2430DD) we find the Starred Robin, Gurney's Sugarbird and Barratt's Warbler among the restricted range species. The Northern Cape grid cell (3020CB) again highlights the presence of the Red Lark and Sclater's Lark with their narrow distributions. Table 4 lists the same species for the grid cell in the Eastern Highlands of Zimbabwe (1932BB) as were listed in Table 2. The grid cell in northern Botswana (1922AA) has Dickinson's Kestrel *Falco dickinsoni* and Bennett's Woodpecker *Campethera bennetti* with relatively restricted distributions. In northern Zimbabwe (grid cell 1728BB) we find the narrow 20% cores of the Blackeared Canary *Serinus mennelli* and the Rackkettailed Roller *Coracias spatulata*, among others. Finally Table 4 confirms the presence of only two restricted-range species in grid cell 1835DA, the Shorttailed Pipit with a 20% core of only 4 grid cells and Stierling's Barred Warbler *Calamonastes stierlingi* with a much wider 20% core of 401 grid cells. The only restricted range species in western Namibia (grid cell 2615AA) is the Dune Lark.

Ratios of Richness of Restricted-Range Species to Overall Species Richness

Finally we look at the ratios of species endemism to species richness, as given in equations 15 and 16. Fig. 24 maps the distribution of the ratios for species that have restricted distributions for their 20%-cores. The grid cells shaded in purple will indicate areas where a large proportion of the species in the 20%-core of their distributions have these cores restricted to fewer than 500 grid cells. The west coast of Namibia has the darkest shading. This is an area where there were not that many species in the 20%-core of their ranges. Fig. 23 shows that these areas also have relatively few species with their 20%-core restricted to fewer than 500 grid cells. However, Fig. 24 shows that a large proportion of the few species with 20%-core ranges here, have restricted cores. Because this ratio may sometimes reflect one out of one or one out of two species, this map can be confusing. Other areas that differ in relative shading from those in Fig. 23 include the north-east

coast of KwaZulu-Natal, large parts of southern and central Mozambique and north-eastern Zimbabwe. All of these areas have relatively few endemic species but a relatively large proportion of those which do occur there have restricted distributions of fewer than 500 grid cells.

Fig. 25 expresses the two measures of endemism and species richness based on the detection probability deciles as a ratio, converted to a percentage. Effectively, we are comparing Fig. 15 to Fig. 12. The measures underlying the distributions in these figures are not simple counts and hence the percentage is not a direct percentage of the number of narrow endemics out of the total number of species. It rather expresses the degree of narrow endemism to the degree of species richness as a percentage. Some of the previously purple shaded areas for both the endemism (Fig. 15) and the species richness (Fig. 12) distributions are now shaded in red, indicating that the relatively large number of narrow endemics in these areas are mainly due to the large number of different species and that many of the species in these areas have wider distributions. Central and southern Botswana have a large number of species in the core of their distributions as indicated in Fig. 12. Relatively fewer of these have restricted distributions as indicated by the lighter shading in Fig. 15. This is confirmed in Fig. 25 where most of the grid cells in Botswana fall in the bottom quantile of the percentage distribution, shaded in green. We once again observe that the west coast of Namibia and most of Mozambique do not have many species in the cores of their ranges (Fig. 12), but that a significant percentage of those there have relatively narrow distributions.

Discussion

Our objective has been to investigate the benefits of using measures derived from reporting rate data, in comparison to the use of presence-absence data, to identify areas rich in species and rich in narrow-endemic species. We generalized existing measures of species richness and species endemism in two ways: by replacing presence-absence data with detection probability deciles that reflect the relative likelihood of detecting a species in a given grid cell and by limiting the area of presence to the core of the range, where the core was also defined in terms of detection probability deciles. In this way our measures give more weight to the areas where species have large relative detection probabilities, i.e., the areas where the species have the core of their distributions. At the same time the peripheral edges of the species distributions, where detection probabilities may be too small to guarantee continued survival, are down-weighted.

Interpretation of Detection Probability Deciles

The need for the transformation of the smoothed detection probabilities to their decile values arose because detection probabilities between species cannot be compared. This is because they are the smoothed version of the observed reporting rates that are subject to species characteristics such as differential conspicuousness and abundance (Harrison & Underhill 1997). For this reason we transform the detection probabilities so that the largest detection probabilities for all species have a common value. Ideally, the study area should be large enough to include complete distributions for the chosen species. It is essential that it includes considerable variation among the detection probabilities for a single species.

If the study area includes the complete distribution for a species, the addition of neighbouring areas will not affect the range of detection probabilities and hence the detection probability deciles will remain unchanged. If, on the other hand, the study area only covers part of the distribution range for a species, the detection probability deciles could change substantially as a result of extending the study area. This condition is necessary to ensure the independence of the results from the demarcation of the study area. This is clearly illustrated by a comparison of the distribution maps for all terrestrial species and for southern African endemic species (Figs 9 and 15). Fig. 9 shows the distribution of apparent narrow endemism based on data for all terrestrial species. It includes the southern limits of the distributions of species that have their cores farther north in the tropics. Should the study area be extended to include more of Africa, the range

of detection probabilities will change and the detection probabilities falling within the current study area may fall in the lower deciles of the entire detection probability range. The analysis of Fig. 9 weights these tropical species as if they had narrow ranges in the north of southern Africa. They illustrate effectively how and where the distributions of these tropical species extend into southern Africa. However, to make results independent of the chosen study area so that our measures would focus on the real cores of the species distributions, we chose a subset of endemic or near-endemic species. For this subset, the extension of the study area will not change the range of detection probabilities and hence the detection probability deciles will remain constant.

The second condition, that refers to the variation in detection probabilities for a species, is necessary to avoid the incorrect differential weighting of essentially uniform detection probabilities. In particular, should a species be uniformly abundant and uniformly conspicuous over the entire study area, the division of a short range of large detection probabilities will result in some large detection probabilities being replaced by small decile values, thus giving the wrong impression of low abundance. For example, in southern Mozambique the Blackeyed Bulbul *Pycnonotus barbatus* has a uniform distribution of relatively large detection probabilities ranging between 0.7589 and 0.9989 (Parker 1999). If our study area was thus restricted to southern Mozambique (only 253 grid cells), the transformation to detection probability deciles would divide this short range of relatively large detection probabilities into 10 subranges, thereby replacing the large probability value of 0.7589 with a decile value of 0.1. Though this does indicate the relative importance of these grid cells with respect to the range of detection probabilities in the area, it may give the erroneous impression of small detection probabilities. For the entire study area of southern Africa, as described in Chapter 2, the detection probabilities for the Blackeyed Bulbul ranged between 0.0061 and 0.9987 and the species was absent from the arid west (Harrison *et al.* 1997b). The most widespread species in southern Africa is the Cape Turtle Dove *Streptopelia capicola*, occurring in 91% of grid cells; but even this species is completely absent from the Namib Desert and scarce in the coastal forests of KwaZulu-Natal (Harrison *et al.* 1997a) and Mozambique (Parker 1999). A further potential problem may arise when a species is uniformly detectable in a substantial sub-region of the study area. This may result in a number of the probability deciles representing effectively equal detection probabilities. In such cases we may be differentiating too much between effectively equal probabilities and as a result we are potentially narrowing the core-range. However, for all the species considered, the southern African study area contained sufficiently many habitats to ensure a large amount of variation in reporting rates and hence detection probabilities. To quote Alan *et al.* (1997b), "the southern African subcontinent has a diversity of habitats probably unsurpassed by any other region of comparable size. Seven major terrestrial biomes are represented: forest, woodland (or savanna), grassland, fynbos, Nama Karoo, succulent Karoo and desert (Rutherford & Westfall 1986)." We are thus justified in transforming to deciles of detection probabilities.

Interpretability of Detection Probability Deciles over Time

The above discussions centered around the issue that reporting rate deciles generate results that are in relation to the chosen study area. The time axis provides another reference frame. Should the Bird Atlas Project be repeated , it is conceivable that the core ranges, as defined by us, may change even if the complete range for a species stays constant. This is because the core range is defined relative to the complete range and changes anywhere within this overall range may change the relative relationships within the range of detection probabilities. Information regarding such shifts in the concentration of a species is very valuable to bio-geographers and conservationists and the fact that the relative core ranges are sensitive to such shifts is an added bonus. An example of a species that has changed the location of its core range is the Blue Crane *Anthropoides paradiseus*. Traditionally its core range was in the grassland biome. However, due to various factors, including the loss of habitat because of afforestation, urbanisation and crop farming, it has declined in numbers in the grassland biome and expanded its range to the agricultural regions of the fynbos biome where it now has its core. While this may not be the preferred biome for the Blue Crane, the fact that farmlands provided a satisfactory alternative habitat for this species has been an important factor in its conservation (Shaw 2003).

Evaluation of New Measures of Species Concentration and Species Endemism

Our suggested new measures based on detection probability deciles fall into two classes. The measures of narrow endemism given in equations (7) and (14) respectively, has the advantage of not excluding any species or any grid cells which do not meet some (possibly subjective) criteria. Instead they take into account the relative importance of species and grid cells by essentially weighting species presence by the magnitude of the relative probability of species detection in a given grid cell. They do not represent simple counts and as such are strictly not measures that indicate overall species "richness" or "richness" of species with narrow distributions. They measure "relative" species richness and are thus indicators of areas where many species, or many species with narrow distributions, have their common core concentrations.

The second approach, as given in equation 12, defines the core ranges for species as areas where the detection probabilities are above a specified percentile, and then counts the number of species in these cores of their ranges for a given grid cell. These measures do present simple counts of species. They also have the advantages of an explicit definition of the relative level of detection probability required to define the core and an explicit cut-off for range-restriction. However, their disadvantage is the exclusion of both species and grid cells that do not satisfy these definitions. Though this second class of measures divides species into two categories that define presence or absence, they do so based on the range of detection probabilities generated by our smoothing model for the study area. They do, however, clearly illustrate that the use of a range of detection probabilities is an extension of the traditional presence-absence approach. When we reduce the number of deciles to one by setting all decile values greater than zero to one, we get back to presence-absence data.

The main difference between the measures of species endemism given in equations (7) and (14) and the measures based on core ranges given in equation (12), relates to the inclusion or exclusion of species and grid cells as already discussed. Secondly, the measures based on explicitly defined core-ranges require a relatively large number of species in their narrow cores in a grid cell for that grid cell to be ranked towards the top end of the range, whereas the more continuous measures will have a large value for a grid cell if only one species with a narrow range has a large detection probability for that grid cell.

Measures based on a range of detection probability deciles are more sensitive to species distribution patterns than those using presence-absence data, in that they treat grid cells for which a species has different relative detection probabilities differently. As a result they generate a more discriminating ranking of grid cells with respect to the measures of species richness and endemism. They also avoid relying too much on the peripheral edges of species distributions and thus avoid guiding conservation efforts to areas where the level of the species population may not be enough for sustainable conservation. The ability to define, determine and map the relative core-ranges for species provides a large degree of additional insight into species distribution patterns. Presence-absence data previously told us where a species could be found. Core-ranges show us where a species appears with a large relative probability of detection and where it finds its current more favourable habitat. This may not in fact be the true optimal habitat for a species but one where the species have moved to as a result of heavily transformed landscapes. The movement of the Blue Crane from the grassland biome to the farmlands of the fynbos biome is one such example. For this species these farmlands provide a viable alternative habitat and conservation efforts directed at this new habitat will be worthwhile. However, this is not always the case. For example, for the Nene goose of Hawaii, conservation efforts directed at the highlands where the only remnant population was found, proved less successful because the optimal habitat for this species was at lower altitudes (Black 1995 & BirdLife International 2000).

The ratios of species endemism to species richness add a further dimension to the understanding of distribution patterns in that they identify areas that have a proportionately large number of narrow endemic species

regardless of the underlying overall species richness. The use of these ratios when the overall number of species can vary between 1 and 80 for a given grid cell is questionable. However, we do not wish to choose one measure above another but rather wish to present ecologists and biogeographers with a choice of measures that will help them to understand the distribution patterns of species and to use this knowledge in their research and conservation efforts. The measure of endemism given in equation (14) that relies on the sum of a range of detection probability deciles weighted by the inverse of their range-sizes does seem to work well in highlighting areas where species with narrow distributions occur in the core of their ranges. It decreases the range size by down-weighting grid cells with small relative probabilities of detection, it identifies grid cells where a species has a large relative detection probability, and it does not exclude grid cells containing only one or two species with extremely narrow ranges.

Previous Research

Two previous studies into the patterns of distributions of endemic Afrotropical species have looked at measuring both species richness and narrow endemism for study areas that include the southern African region. Crowe & Crowe (1982) analysed "patterns of distribution, diversity and endemism" for Afrotropical terrestrial species. De Klerk (1998) looked at biogeographical patterns of distribution of terrestrial Afrotropical species. Both studies were based on presence-absence data that were extracted from existing range maps for the species. Both studies weighted the presence of each species by the inverse of the number of localities at which it was recorded. Crowe & Crowe (1982) used four-degree quadrats. De Klerk (1998) digitised individual species distributions into a one-degree grid cell system. Both studies performed a form of cluster analysis on the localities in an attempt to group these into areas based on their similarity with respect to species composition. De Klerk (1998) also mapped the distributions of species richness and richness of range-restricted species. The essential difference between these two studies and our own research is that while they adopted an inter-species approach, our analysis is based on looking at patterns of distribution within individual species ranges.

Biogeographical Patterns

While our focus has been on the derivation of methods to measure species richness and narrow endemism and not really on the biogeographical results, we do just summarise the results from a biogeographical perspective. Firstly, Figs 15 and 19 identified the so-called "hotspots " for endemism in southern Africa and Table 2 analysed the species composition of these areas. They are the Fynbos Biome in the Western Cape, and the various areas where the larks have the centres of their narrow distributions like Rudd's Lark and Botha's Lark in the eastern parts of South Africa, Sclater's Lark and the Red Lark in the Northern Cape and Dune Lark in western Namibia. Further hotspots include the Drakensberg region where we find the Drakensberg Siskin, Orange Breasted Rockjumper, Gurney's Sugarbird, Bush Blackcap, Bearded Vulture and Chorister Robin. Along the east coast of South Africa, we find the core distributions for the Spotted Thrush, Knysna Warbler and the Knysna Woodpecker. The Green Barbet has an extremely narrow range in KwaZulu-Natal, the Shorttailed Pipit has a similar narrow distribution in Mozambique and the Dune Lark has its narrow distribution in western Namibia. The Eastern Highlands of Zimbabwe is an important centre for species like the Starred Robin, Olive Bush Shrike, Barratt's Warbler and Striped Pipit. Secondly, by looking at the richness of species in increasingly narrow cores of their distributions (Fig. 22), Botswana has emerged as an area where a large number of species have very large relative detection probabilities in the top percentile of their ranges. This is most probably as a result of the fragmented landscape that consists of both good quality open Kalahari grassland habitat and dense tall thornbush as a result of bush encroachment in overgrazed areas. Each of these is the optimal habitat for a group of regional endemics (M. Herremans, pers. comm.). Some of these species include the Orange River Francolin, Tinkling Cisticola, Barred Warbler, Penduline Tit and Clapper Lark.

Hotspots for endemism indicate areas where many range-restricted species occur with relatively large detection probabilities, i.e. where they have their current common cores. We have seen how core ranges may

indicate true natural habitats for some species, for example for the fynbos biome in the Western Cape. For others species like the Blue Crane, the core ranges may point towards current, but not traditionally, favourable habitats. On the other hand, there may be some endemic species in southern Africa for whom, as for the Nene Goose of Hawaii, the current cores reflect their movement to marginal habitats not ideally suited for survival. Sometimes the distribution of endemism hotspots may point towards areas where land-use policies have created successful habitats for different groups of endemic species, as in western Botswana, that cannot be improved by further conservation attempts. In fact, areas with low detection probabilities for several endemic species, for example the Kalahari biome in Namibia and South Africa, may be better targets where successful conservation efforts may result in increased detection probabilities (M. Herremans, pers. comm.). The distribution maps of relative species richness and endemism provides exciting additional insight into patterns of species distribution in the region - insight that was not previously available from distributions based on presence-absence data. These new maps should be interpreted with care and they should only be used to guide conservation efforts in conjunction with other biological and geographical information. However, to conclude, we believe that we have met the challenge set by Van Jaarsveldt *et al.* (1998) to provide the analytical methods to analyse data based on reporting rates and that we have shown the benefits of using such data over simple presence-absence data.

References

Allan DG, Harrison JA, Navarro RA, Van Wilgen BW, Thompson MW 1997a. The impact of commercial afforestation on bird populations in Mpumalanga Province, South Africa - insights from bird-atlas data. Biological Conservation 79: 173-185.

Allan DG, Harrison JA, Herremans M, Navarro RA, Underhill LG 1997b. Southern African geography: its relevance to birds. In: Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds). The Atlas of Southern African Birds. Vol 1: Non-passerines,pp. xliii-lxiv. BirdLife South Africa, Johannesburg.

BirdLife International 2000. Threatened birds of the world. Barcelona and Cambridge, UK: Lynx Edicions and BirdLife International.

Black JM 1995. The Nene Branta sandvicensis Recovery Initiative: research against extinction. Ibis 137: S153-S160.

Clancey PA 1986. Endemicity in the southern African avifauna. Durban Museum Novitates 13: 245-284.

Crowe TM, Crowe A 1982. Patterns of distribution, diversity and endemism in Afrotropical birds. Journal of the Zoological Society of London 198: 417-442.

De Klerk HM 1998. Biogeography and Conservation of Terrestrial Afrotropical Birds. Unpublished Ph.D. thesis, University of Cape Town.

Gaston KJ, Rodrigues ASL (in press). Reserve selection in regions with poor biological data. Conservation Biology.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997a. The Atlas of Southern African Birds. Vol 1: Non-passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997b. The Atlas of Southern African Birds. Vol 2: Passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Underhill LG 1997. Introduction and methods. In: Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) The Atlas of Southern African Birds. Vol 1: Non-passerines, pp. xliii-lxiv. BirdLife South Africa, Johannesburg.

Lawes MJ, Piper SE 1998. There is less to binary maps than meets the eye: the use of species distribution data in the southern African sub-region. South African Journal of Science 94: 207-210.

Linder HP 2001. On areas of endemism, with an example from the African Restionaceae. Systematic Biology 50: 892-912.

Parker V 1999. The Atlas of the Birds of Sul do Save, Southern Mozambique. Avian Demography Unit and Endangered Wildlife Trust, Cape Town and Johannesburg.

Rutherford MC, Westfall RH 1986. Biomes of southern Africa - an objective categorization. Memoirs of the Botanical Survey of South Africa 54: 1-98.

Shaw K 2003. Blue Crane *Anthropoides paradiseus*. In: Big Birds on Farms: Mazda CAR Report 1993-2001. Young DJ, Harrison JA, Anderson MD, Colahan BD (eds), 32-37. Avian Demography Unit, Cape Town. Underhill LG 1999. Avian demography: statistics and ornithology. In: Adams NJ, Slotow RH (eds). Proceedings of the 22nd International Ornithological Congress, Durban. Ostrich 70: 61-70.

Van Jaarsveld AS, Gaston KJ, Chown SL, Freitag S 1998. Throwing biodiversity out with the binary data? South African Journal of Science 94: 210-214

Appendix

LIST OF 152 SPECIES ENDEMIC OR NEAR ENDEMIC TO SOUTHERN AFRICA (CLANCEY 1986)

Bearded Vulture	*Gypaetus barbatus*	Cape Vulture	*Gyps coprotheres*
Jackal Buzzard	*Buteo rufofuscus*	Pale Chanting Goshawk	*Melierax canorus*
Black Harrier	*Circus maurus*	Dickinson's Kestrel	*Falco dickinsoni*
Pygmy Falcon	*Polihierax semitorquatus*	Orange River Francolin	*Francolinus levaillantoides*
Redbilled Francolin	*Francolinus adspersus*	Cape Francolin	*Francolinus capensis*
Natal Francolin	*Francolinus natalensis*	Swainson Francolin	*Francolinus swainsonii*
Blue Crane	*Anthropoides paradiseus*	Ludwig's Bustard	*Neotis ludwigii*
Blue Korhaan	*Eupodotis caerulescens*	Karoo Korhaan	*Eupodotis vigorsii*
Redcrested Korhaan	*Eupodotis ruficrista*	Black Korhaan	*Eupodotis afra*
Namaqua Sandgrouse	*Pterocles namaqua*	Burchell Sandgrouse	*Pterocles burchelli*
Doublebanded Sandgrouse	*Pterocles bicinctus*	Grey Lourie	*Corythaixoides concolor*
Bradfield's Swift	*Apus bradfieldi*	Redfaced Mousebird	*Urocolius indicus*
Rackettailed Roller	*Coracias spatulata*	Pied Barbet	*Tricholaema leucomelas*
Green Barbet	*Cryptolybia woodwardi*	Redfronted Tinker Barbet	*Pogoniulus pusillus*
Crested Barbet	*Trachyphonus vaillantii*	Ground Woodpecker	*Geocolaptes olivaceus*
Bennett's Woodpecker	*Campethera bennettii*	Knysna Woodpecker	*Campethera notata*
Melodious Lark	*Mirafra cheniana*	Monotonous Lark	*Mirafra passerina*
Clapper Lark	*Mirafra apiata*	Sabota Lark	*Mirafra sabota*
Rudd's Lark	*Heteromirafra ruddi*	Longbilled Lark	*Certhilauda curvirostris*
Shortclawed Lark	*Certhilauda chuana*	Karoo Lark	*Certhilauda albescens*
Dune Lark	*Certhilauda erythrochlamys*	Red Lark	*Certhilauda burra*
Spikeheeled Lark	*Chersomanes albofasciata*	Pinkbilled Lark	*Spizocorys conirostris*
Botha's Lark	*Spizocorys fringillaris*	Sclater's Lark	*Spizocorys sclateri*
Stark's Lark	*Eremalauda starki*	Thickbilled Lark	*Galerida magnirostris*
Greybacked Finchlark	*Eremopterix verticalis*	Blackeared Finchlark	*Eremopterix australis*
Pearlbreasted Swallow	*Hirundo dimidiata*	Southern Grey Tit	*Parus afer*
Ashy Tit	*Parus cinerascens*	Southern Black tit	*Parus niger*
Cape Penduline Tit	*Anthoscopus minutus*	Pied Babbler	*Turdoides bicolor*
Bush Blackcap	*Lioptilus nigricapillus*	Cape Bulbul	*Pycnonotus capensis*
Redeyed Bulbul	*Pycnonotus nigricans*	Kurrichane Thrush	*Turdus libonyana*
Spotted Thrush	*Zoothera guttata*	Cape Rock Thrush	*Monticola rupestris*
Sentinal Rock Thrush	*Monticola explorator*	Shorttoed Rock Thrush	*Monticola brevipes*
Mountain Chat	*Oenanthe monticola*	Buffstreaked Chat	*Oenanthe bifasciata*
Tractrac Chat	*Cermomela tractrac*	Sicklewinged Chat	*Cercomela sinuata*
Karoo Chat	*Cercomela schlegelii*	Anteating Chat	*Myrmecocichla formicivora*
Chorister Robin	*Cossypha dichroa*	Whitethroated Robin	*Cossypha humeralis*
Starred Robin	*Pogonocichla stellata*	Cape Rockjumper	*Chaetops frenatus*
Orangebreasted Rockjumper	*Chaetops aurantius*	Karoo Robin	*Erythropygia coryphaeus*
Kalahari Robin	*Erythropygia paena*	Brown Robin	*Erythropygia signata*
Titbabbler	*Parisoma subcaeruleum*	Layard Titbabbler	*Parisoma layardi*
Barratt's Warbler	*Bradypterus barratti*	Knysna Warbler	*Bradypterus sylvaticus*
Victorin's Warbler	*Bradypterus victorini*	Rudd's Apalis	*Apalis ruddi*
Longbilled Crombec	*Sylvietta rufescens*	Karoo Eremomela	*Eremomela gregalis*
Burntnecked Eremomela	*Eremomela usticollis*	Barred Warbler	*Calamonastes fasciolatus*
Stierling's Barred Warbler	*Calamonastes stierlingi*	Cinnamonbreasted Warbler	*Euryptila subcinnamomea*
Grassbird	*Sphenoeacus afer*	Cloud Cisticola	*Cisticola textrix*

Greybacked Cisticola	*Cisticola subruficapilla*	Tinkling Cisticola	*Cisticola rufilata*
Blackchested Prinia	*Prinia flavicans*	Spotted Prinia	*Prinia hypoxantha*
Namaqua Prinia	*Phragmacia substriata*	Rufouseared Warbler	*Malcorus pectoralis*
Marico Flycatcher	*Melaenornis mariquensis*	Chat Flycatcher	*Melaenornis infuscatus*
Fiscal Flycatcher	*Sigelus silens*	Cape Batis	*Batis capensis*
Pririt Batis	*Batis pririt*	Woodward's Batis	*Batis fratrum*
Fairy Flycatcher	*Stenostira scita*	Striped Pipit	*Anthus lineiventris*
Rock Pipit	*Anthus crenatus*	Shorttailed Pipit	*Anthus brachyurus*
Yellowbreasted Pipit	*Hemimacronyx chloris*	Orangethroated Longclaw	*Macronyx capensis*
Southern Boubou	*Laniarius ferrugineus*	Crimsonbreasted Shrike	*Laniarius atrococcineus*
Southern Tchagra	*Tchagra tchagra*	Bokmakierie	*Telophorus zeylonus*
Olive Bush Shrike	*Telophorus olivaceus*	Whitecrowned Shrike	*Eurocephalus anguitimens*
Pied Starling	*Spreo bicolor*	Burchell's Starling	*Lamprotornis australis*
Longtailed Starling	*Lamprotornis mevesii*	Redwinged Starling	*Onychognathus morio*
Palewinged Starling	*Onychognathus nabouroup*	Cape Sugarbird	*Promerops cafer*
Gurney's Sugarbird	*Promerops gurneyi*	Orangebreasted Sunbird	*Nectarinia violacea*
Neergaard Sunbird	*Nectarinia neergaardi*	Lesser Double-collared Sunbird	*Nectarinia chalybea*
Greater Double-collared Sunbird	*Nectarinia afra*	Whitebellied Sunbird	*Nectarinia talatala*
Dusky Sunbird	*Nectarinia fusca*	Cape White-eye	*Zosterops pallidus*
Sociable Weaver	*Philetairus socius*	Cape Sparrow	*Passer melanunis*
Yellowthroated Sparrow	*Petronia superciliaris*	Scalyfeathered Finch	*Sporopipes squamifrons*
Cape Weaver	*Ploceus capensis*	Pinkthroated Twinspot	*Hypargos margaritatus*
Blue Waxbill	*Uraeginthus angolensis*	Violeteared Waxbill	*Uraeginthus granatinus*
Swee Waxbill	*Estrilda melanotis*	Redheaded Finch	*Amadina erythrocephala*
Shafttailed Whydah	*Vidua regia*	Lemonbreasted Canary	*Serinus citrinipectus*
Forest Canary	*Serinus scotops*	Cape Siskin	*Pseudochloroptila totta*
Drakensberg Siskin	*Pseudochloroptila symonsi*	Blackheaded Canary	*Serinus alario*
Yellow Canary	*Serinus flaviventris*	Whitethroated Canary	*Serinus albogularis*
Protea Canary	*Serinus leucopterus*	Blackeared Canary	*Serinus mennelli*
Cape Bunting	*Emberiza capensis*	Larklike Bunting	*Emberiza impetuani*

Table 1 : Analysis of Species Composition for Top Scoring Grid Cells in Figure 9

Grid Cell	Grid Total	Rank deciles	Species Details	Species Contribution Absolute	Species Contribution Percent	Range Di	No. of Grid Cells	Percent of Study Area
1935AB	3.973	1	Whitebreasted Alethe	0.3226	8.12	3.1	13	0.26
			Barred Cuckoo	0.2174	5.47	4.6	24	0.48
			Blackheaded Apalis	0.2174	5.47	4.6	23	0.46
			Eastern Honeyguide	0.2128	5.36	4.7	12	0.24
			Gunning's Robin	0.1961	4.94	5.1	16	0.32
			Yellowbrstd Hyliota	0.1961	4.94	5.1	14	0.28
			Delegorgue's Pigeon	0.1887	4.75	5.3	19	0.38
			Angola Pitta	0.1667	4.20	6.0	19	0.38
			Nyasa Seedcracker	0.1639	4.13	6.1	23	0.46
			Lesser Cuckoo	0.1538	3.87	6.5	13	0.26
1835CC	3.7868	2	Whitebreasted Alethe	0.3226	8.52	3.1	13	0.26
			Blackheaded Apalis	0.2174	5.74	4.6	23	0.46
			Barred Cuckoo	0.2174	5.74	4.6	24	0.48
			Gunning's Robin	0.1961	5.18	5.1	16	0.32
			Yellowbrstd Hyliota	0.1961	5.18	5.1	14	0.28
			Eastern Honeyguide	0.1915	5.06	4.7	12	0.24
			Delegorgue's Pigeon	0.1887	4.98	5.3	19	0.38
			Nyasa Seedcracker	0.1639	4.33	6.1	23	0.46
			Lesser Cuckoo	0.1538	4.06	6.5	13	0.26
			Angola Pitta	0.1500	3.96	6.0	19	0.38
1834AC	1.9522	3	Greenheaded Oriole	0.5556	28.46	1.8	4	0.08
			Swynnerton's Robin	0.1786	9.15	2.8	6	0.12
			Eastern Honeyguide	0.1277	6.54	4.7	12	0.24
			Nyasa Seedcracker	0.1148	5.88	6.1	23	0.46
			Moustached Warbler	0.0882	4.52	10.2	31	0.61
			Marsh Tchagra	0.0857	4.39	3.5	12	0.24
			Blckfrnt Bush Shrike	0.0750	3.84	8.0	30	0.59
			Vanga Flycatcher	0.0720	3.69	12.5	46	0.91
			Chirinda Apalis	0.0548	2.81	7.3	27	0.54
			Delegorgue's Pigeon	0.0377	1.93	5.3	19	0.38
1932BB	1.9304	4	Swynnerton's Robin	0.3571	18.50	2.8	6	0.12
			Redfaced Crimsonwing	0.1538	7.97	6.5	14	0.28
			Blckfrnt Bush Shrike	0.1250	6.48	8.0	30	0.59
			Chirinda Apalis	0.1233	6.39	7.3	27	0.54
			Orange Thrush	0.1149	5.95	8.7	48	0.95
			Whitetailed Flycatcher	0.0943	4.89	10.6	35	0.69
			Stripecheecked Bulbul	0.0935	4.84	10.7	34	0.67
			Robert's Prinia	0.0860	4.56	9.3	25	0.50
			Yellowstreaked Bulbul	0.0645	3.34	15.5	75	1.49
			East African Swee	0.0641	3.32	15.6	44	0.87
2033AA	1.8893	5	Eastern Honeyguide	0.1915	10.14	4.7	12	0.24
			Delegorgue's Pigeon	0.1132	5.99	5.3	19	0.38
			Lesser Cuckoo	0.1077	5.70	6.5	13	0.46
			Blckfrnt Bush Shrike	0.1000	5.29	8.0	30	0.59
			Vanga Flycatcher	0.0800	4.23	12.5	46	0.91
			Slender Bulbul	0.0725	3.84	6.9	36	0.71
			Barred Cuckoo	0.0652	3.45	4.6	24	0.48
			Yellowstreaked Bulbul	0.0645	3.41	15.5	75	1.49
			Whitetailed Flycatcher	0.0566	2.99	10.6	35	0.69
			Silverycheeked Hornbill	0.0496	2.63	14.1	55	1.09

Table 2 : Analysis of Species Composition for Selected Top Scoring Endemism Grid Cells based on deciles as illustrated in Figures 15 and 19

Grid Cell	Grid Total	Rank deciles	Species Details	Species Contribution		Range DI	No. of Grid Cells	Percent of Study Area
				Absolute	Percent			
3418BD	0.2961	7	Cape Rockjumper	0.0667	22.52	15.0	55	1.09
			Victorin's Warbler	0.0625	21.11	16.0	79	1.56
			Orangebrstd Sunbird	0.0298	10.05	33.6	149	2.95
			Protea Canary	0.0275	9.28	18.2	89	1.76
			Cape Sugarbird	0.0236	7.98	42.3	170	3.36
			Cape Siskin	0.0223	7.54	44.8	155	3.07
3418AB	0.0906	186	Orangebrstd Sunbird	0.0179	19.71	33.6	149	2.95
			Cape Sugarbird	0.0118	13.05	42.3	170	3.36
			Cape Siskin	0.0112	12.32	44.8	155	3.07
			Cape Francolin	0.0086	9.46	105.0	273	5.40
			Cape Bulbul	0.0069	7.67	129.5	289	5.71
			Grassbird	0.0048	5.31	124.8	647	12.79
3322BD	0.2178	10	Cape Rockjumper	0.0667	30.6093	15	55	1.09
			Protea Canary	0.0275	12.6139	18.2	89	1.76
			Victorin's Warbler	0.0188	8.6088	16	79	1.56
			Cape Siskin	0.0179	8.1988	44.8	155	3.07
			Orangebrstd Sunbird	0.0089	4.0996	33.6	149	2.95
3228DA	0.2633	8	Spotted Thrush	0.1250	47.47	8.0	33	0.65
			Brown Robin	0.0379	14.39	26.4	121	2.39
			Knysna Woodpecker	0.0373	14.18	24.1	123	2.43
			Chorister Robin	0.0158	5.99	50.7	256	5.06
			Rdfrtd Tinker Barbet	0.0090	3.42	111.1	285	5.64
3228BD	0.3113	6	Spotted Thrush	0.0750	24.09	8.0	33	0.65
			Knysna Warbler	0.0596	19.15	15.1	30	0.59
			Knysna Woodpecker	0.0415	13.33	24.1	123	2.43
			Brown Robin	0.0341	10.95	26.4	121	2.39
			Barrat's Warbler	0.0216	6.95	37.0	170	3.36
3127DC	0.144	53	Rudd's Lark	0.0667	46.2965	3	7	0.14
			Yellowbreasted Pipit	0.0253	17.5806	7.9	43	0.85
			Buch Blackcap	0.0074	5.1632	26.9	84	1.66
			Starred Robin	0.007	4.8563	28.6	186	3.68
			Rock Pipit	0.0039	2.6917	77.3999	386	7.63
3030BC	0.1549	43	Shorttailed Pipit	0.05	32.2789	2	7	0.14
			Brown Robin	0.0227	14.672	26.4	121	2.39
			Spotted Thrush	0.0125	8.0697	8	33	0.65
			Barratt's Warbler	0.0108	6.9793	37	170	3.36
			Rdfrtd Tinker Barbet	0.0081	5.2298	111.1	285	5.64
3028BD	0.5793	2	Rudd's Lark	0.3333	57.54	3.0	7	0.14
			Yellowbreasted Pipit	0.1266	21.85	7.9	43	0.85
			Orngbrstd Rockjumper	0.0226	3.90	31.0	92	1.82
			Buffstreaked Chat	0.0171	2.95	58.5	272	5.38
			Gurney's Sugarbird	0.0147	2.54	27.2	123	2.43
3020BD	0.0869	205	Sclater's Lark	0.0212	24.35	37.8	134	2.65
			Red Lark	0.0165	19.02	24.2	80	1.58
			Blackeared Finchlark	0.0073	8.41	123.2	463	9.16
			Tractrac Chat	0.0033	3.84	209.5	773	15.29
			Karoo Korhaan	0.0030	3.44	267.3	840	16.61
3020AD	0.0898	191	Sclater's Lark	0.0212	23.57	37.8	134	2.65
			Red Lark	0.0207	23.01	24.2	80	1.58
			Blackeared Finchlark	0.0049	5.42	123.2	463	9.16
			Tractrac Chat	0.0038	4.25	209.5	773	15.29
			Karoo Korhaan	0.0026	2.92	267.3	840	16.61

Grid Cell	Grid Total	Rank deciles	Species Details	Species Contribution		Range DI	No. of Grid Cells	Percent of Study Area
				Absolute	Percent			
2831DC	0.9157	1	Woodward's Barbet	0.8333	91.01	1.2	2	0.04
			Spotted Thrush	0.0500	5.46	8.0	33	0.65
			Chorister Robin	0.0059	0.65	50.7	256	5.06
			Rdfrtd Tinker Barbet	0.0054	0.59	111.1	285	5.64
			Brown Robin	0.0038	0.41	26.4	121	2.39
2831DD	0.2087	11	Woodward's Barbet	0.1667	79.8596	1.2	2	0.04
			Spotted Thrush	0.025	11.9789	8	33	0.65
			Starred Robin	0.0035	1.6756	28.6	186	3.68
			Rdfrtd Tinker Barbet	0.0027	1.2937	111.1	2	0.04
			Rudd's Apalis	0.0024	1.138	42.1	173	3.42
2829CC	0.1432	56	Gurney's Sugarbird	0.0184	12.8366	27.2	123	2.43
			Bush Blackcap	0.0149	10.3841	26.9	233	4.61
			Bearded Vulture	0.0116	8.0964	34.5	103	2.04
			Chorister Robin	0.0099	6.8869	50.7	256	5.06
			Orngbrstd Rockjumper	0.0097	6.7577	31	92	1.82
2730AC	0.2573	9	Rudd's Lark	0.1000	38.87	3.0	7	0.14
			Yellowbreasted Pipit	0.0506	19.68	7.9	43	0.85
			Botha's Lark	0.0317	12.34	6.3	17	0.34
			Bush Blackcap	0.0186	7.22	26.9	94	1.86
			Buffstreaked Chat	0.0103	3.99	58.5	272	5.38
2730AA	0.4291	4	Rudd's Lark	0.2000	46.61	3.0	7	0.14
			Botha's Lark	0.1270	29.59	6.3	17	0.34
			Yellowbreasted Pipit	0.0380	8.85	7.9	43	0.85
			Bush Blackcap	0.0149	3.47	26.9	94	1.86
			Buffstreaked Chat	0.0085	1.99	58.5	272	5.38
2729BB	0.1914	16	Botha's Lark	0.0952	49.7586	6.3	17	0.34
			Rudd's Lark	0.0667	34.8312	3	7	0.14
			Blue Korhaan	0.005	2.6123	140.0001	412	8.15
			Buffstreaked Chat	0.0034	1.7863	58.5	272	5.38
			Orangethrtd Longclaw	0.0022	1.1374	413.502	1145	22.64
2630CC	0.3861	5	Rudd's Lark	0.1667	43.17	3.0	7	0.14
			Botha's Lark	0.1587	41.11	6.3	17	0.34
			Yellowbreasted Pipit	0.0253	6.56	7.9	43	0.85
			Bush Blackcap	0.0074	1.93	26.9	94	1.86
			Blue Korhaan	0.0036	0.92	140.0	412	8.15
2629DD	0.1996	12	Botha's Lark	0.1111	55.6668	6.3	17	0.34
			Rudd's Lark	0.0667	33.4003	3	7	0.14
			Blue Korhaan	0.0036	1.7891	140.0001	412	8.15
			Orangethrtd Longclaw	0.0022	1.0907	413.502	1145	22.64
			Pinkbilled Lark	0.0021	1.0451	95.8998	52	1.03
2615AA	0.0659	392	Dune Lark	0.0602	91.41	16.6	64	1.27
			Greybacked Finchlark	0.0007	1.09	558.0	2505	49.54
			Southern GreyTit	0.0006	0.95	159.2	626	12.38
			Shrttoed Rock Thrush	0.0006	0.85	177.8	777	15.36
			Tractrac Chat	0.0005	0.72	209.5	773	15.29
1932BB	0.0825	234	Starred Robin	0.0350	42.38	28.6	186	3.68
			Olive Bush Shrike	0.0129	15.58	77.8	419	8.29
			Barratt's Warbler	0.0108	13.10	37.0	170	3.36
			Gurney's Sugarbird	0.0074	8.91	27.2	123	2.43
			Cape Batis	0.0054	6.54	148.3	610	12.06
1932BA	0.0827	231	Starred Robin	0.0350	42.28	28.6	186	3.68
			Olive Bush Shrike	0.0129	15.54	77.8	419	8.29
			Striped Pipit	0.0087	10.55	68.8	463	9.16
			Barratt's Warbler	0.0081	9.80	37.0	170	3.36
			Cape Batis	0.0054	6.52	148.3	610	12.06
1835DA	0.5014	3	Shorttailed Pipit	0.5000	99.72	2.0	7	0.14
			String beard Warbler	0.0011	0.23	177.1	928	18.35
			Kurrichane Thrush	0.0003	0.06	663.6	1877	37.12

Table 3 : Analysis of Species Composition for Selected Top Scoring Endemism Gird Cells based on presence-absence data as illustrated in Figures 16 and 20

Grid Cell	Grid Total	Rank	Species Details	Species Contribution		No. of Grid Cells	Percent of Study Area
				Absolute	Percent		
3418BD	0.0935	147	Cape Rockjumper	0.0182	19.45	55	1.09
			Victorin's Warbler	0.0127	13.54	79	1.56
			Protea Canary	0.0112	12.02	89	1.76
			Orangebrstd Sunbird	0.0067	7.18	149	2.95
			Cape Siskin	0.0065	6.90	155	3.07
			Cape Sugarbird	0.0059	6.29	170	3.36
3418AB	0.0539	555	Orangebrstd Sunbird	0.0067	12.45	149	2.95
			Cape Siskin	0.0065	11.97	155	3.07
			Cape Sugarbird	0.0059	10.91	170	3.36
			Cape Francolin	0.0037	6.80	273	5.40
			Cape Bulbul	0.0035	6.42	289	5.71
			Sentinal Rock Thrush	0.0034	6.40	290	5.73
3322BD	0.1283	39	Cape Rockjumper	0.0182	14.1715	55	1.09
			Victorin's Warbler	0.0127	9.8659	79	1.56
			Protea Canary	0.0112	8.7576	89	1.76
			Orangbrstd Sunbird	0.0067	5.2307	149	2.95
			Cape Siskin	0.0065	5.0288	155	3.07
3228DA	0.0719	275	Spotted Thrush	0.0303	42.146	33	0.65
			Brown Robin	0.0083	11.4937	121	2.39
			Knysna Woodpecker	0.0081	11.3074	123	2.43
			Chorister Robin	0.0039	5.4325	256	5.06
			Rdfrtd Tinker Barbet	0.0035	4.8804	285	5.64
3228BD	0.1220	56	Knysna Warbler	0.0333	27.32	30	0.59
			Spotted Thrush	0.0303	24.84	33	0.65
			Brown Robin	0.0083	6.77	121	2.39
			Knysna Woodpecker	0.0081	6.66	123	2.43
			Barratt's Warbler	0.0059	4.82	170	3.36
3127DC	0.2351	6	Rudd's Lark	0.1429	60.76	7	0.14
			Yellowbreasted Pipit	0.0233	9.89	43	0.85
			Bush Blackcap	0.0106	4.52	94	1.86
			Starred Robin	0.0054	2.29	186	3.68
			Chorister Robin	0.0039	1.66	256	5.06
3030BC	0.2333	8	Shorttailed Pipit	0.1429	61.23	7	0.14
			Spotted Thrush	0.0303	12.99	33	0.65
			Brown Robin	0.0083	3.54	121	2.39
			Barratt's Warbler	0.0059	2.52	170	3.36
			Starred Robin	0.0054	2.30	186	3.68
3028BD	0.2282	11	Rudd's Lark	0.1429	62.60	7	0.14
			Yellowbreasted Pipit	0.0233	10.19	43	0.85
			Orangbrstd Rockjumper	0.0109	4.76	92	1.82
			Bearded Vulture	0.0097	4.25	103	2.04
			Gurney's Sugarbird	0.0081	3.56	123	2.43
3020BD	0.0511	636	Red Lark	0.0125	24.46	80	1.58
			Sclater's Lark	0.0075	14.60	134	2.65
			Blackeared Finchlark	0.0022	4.23	463	9.16
			Stark's Lark	0.002	3.98	492	9.73
			Southern Grey Tit	0.0016	3.13	626	12.38
3020AD	0.0613	402	Red Lark	0.0125	20.39	80	1.58
			Sclater's Lark	0.0075	12.17	134	2.65
			Cape Francolin	0.0037	5.98	273	5.40
			Karoo Eremomela	0.0028	4.62	353	6.98
			Blackeared Finchlark	0.0022	3.52	463	9.16

Grid Cell	Grid Total	Rank	Species Details	Species Contribution		No. of	Percent of
				Absolute	Percent	Grid Cells	Study Area
2831DD	0.5611	2	Woodward's Barbet	0.5	89.11	2	0.04
			Spotted Thrush	0.0303	5.40	33	0.65
			Rudd's Apalis	0.0058	1.03	173	3.42
			Starred Robin	0.0054	0.96	186	3.68
			Chorister Robin	0.0039	0.70	256	5.06
2829CC	0.1164	69	Drakensberg Siskin	0.0179	15.3411	56	1.11
			Orangebwtd Rockjumper	0.0109	9.3385	92	1.82
			Bush Blackcap	0.0106	9.1392	94	1.86
			Bearded Vulture	0.0097	8.3411	103	2.04
			Gurney's Sugarbird	0.0081	6.9845	123	2.43
2831DC	0.5750	1	Woodward's Barbet	0.5	86.96	2	0.04
			Spotted Thrush	0.0303	5.27	33	0.65
			Brown Robin	0.0083	1.44	121	2.39
			Rudd's Apalis	0.0058	1.01	173	3.42
			Starred Robin	0.0054	0.94	186	3.68
2730AC	0.2831	4	Rudd's Lark	0.1429	50.46	7	0.14
			Botha's Lark	0.0588	20.78	17	0.34
			Yellowbreasted Pipit	0.0233	8.21	43	0.85
			Bush Blackcap	0.0106	3.76	94	1.86
			Chorister Robin	0.0039	1.38	256	5.06
2730AA	0.2871	3	Rudd's Lark	0.1429	49.76	7	0.14
			Botha's Lark	0.0588	20.49	17	0.34
			Yellowbreasted Pipit	0.0233	8.10	43	0.85
			Bush Blackcap	0.0106	3.71	94	1.86
			Barratt's Warbler	0.0059	2.05	170	3.36
2729BB	0.2341	7	Rudd's Lark	0.1429	61.02	7	0.14
			Botha's Lark	0.0588	25.13	17	0.34
			Buffstreaked Chat	0.0037	1.57	272	5.38
			Sentinal Rock Thrush	0.0034	1.47	290	5.73
			Blue Korhaan	0.0024	1.04	412	8.15
2630CC	0.2735	5	Rudd's Lark	0.1429	52.23	7	0.14
			Botha's Lark	0.0588	21.51	17	0.34
			Yellowbreasted Pipit	0.0233	8.50	43	0.85
			Bush Blackcap	0.0106	3.89	94	1.86
			Buffstreaked Chat	0.0037	1.34	272	5.38
2629DD	0.2302	9	Rudd's Lark	0.1429	62.06	7	0.14
			Botha's Lark	0.0588	25.55	17	0.34
			Buffstreaked Chat	0.0037	1.60	272	5.38
			Blue Korhaan	0.0024	1.05	412	8.15
			Pinkbilled Lark	0.0019	0.83	522	10.32
2615AA	0.0305	1708	Dune Lark	0.0156	51.23	64	1.27
			Southern Grey Tit	0.0016	5.24	626	12.38
			Tractrac Chat	0.0013	4.24	773	15.29
			Shorttoed Rock Thrush	0.0013	4.22	777	15.36
			Karoo Korhaan	0.0012	3.90	840	16.61
1932BB	0.0346	1408	Gurney's Sugarbird	0.0081	23.50	123	2.43
			Barratt's Warbler	0.0059	17.00	170	3.36
			Starred Robin	0.0054	15.54	186	3.68
			Blackeared Canary	0.0024	6.91	418	8.27
			Olive Bush Shrike	0.0024	6.90	419	8.29
1932BA	0.0368	288	Gurney's Sugarbird	0.0081	22.09	123	2.43
			Barratt's Warbler	0.0059	15.98	170	3.36
			Starred Robin	0.0054	14.61	186	3.68
			Blackeared Canary	0.0024	6.50	418	8.27
			Olive Bush Shrike	0.0024	6.49	419	8.29
1835DA	0.1445	22	Shorttailed Pipit	0.1429	98.86	7	0.14
			String Bared Warbler	0.0011	0.75	928	18.35
			Kurrichane Thrush	0.0005	0.37	1877	37.12

Table 4: Analysis of Species Composition for Selected Top Scoring Grid Cells in Figure 23

Grid Cell	No.of Species	Species Details	Range Size
3319DD	20	Victorin's Warbler	39
		Protea Canary	54
		Orangebrstd Sunbird	92
		Cape Siskin	105
		Cape Sugarbird	108
		Swee Waxbill	182
		Cape Francolin	190
		Southern Tchagra	212
		Black Harrier	223
		Cape Bulbul	235
		Grassbird	299
		Namaqua Prinia	321
		Pearlbreasted Swallow	342
		Grtr Dblclrd Sunbird	350
		Cape Rock Thrush	356
		Cape Batis	363
		Ground Woodpecker	365
		Southern Grey Tit	380
		Layard's Titbabbler	401
		Blue Crane	472
3322BD	19	Cape Rockjumper	37
		Victorin's Warbler	39
		Protea Canary	54
		Orangebrstd Sunbird	92
		CapeSiskin	105
		Cape Sugarbird	108
		Forest Canary	112
		Southern Tchagra	212
		Black Harrier	223
		Karoo Lark	232
		Cape Bulbul	235
		Grassbird	299
		Namaqua Prinia	321
		Pearlbreasted Swallow	342
		Cape Rock Thrush	356
		Cape Batis	363
		Ground Woodpecker	365
		Layard's Titbabbler	401
		Sicklewing Chat	453
2929AD	18	Drakensberg Siskin	41
		Orangebrstd Rockjumper	57
		Bush Blackcap	66
		Gurney's Sugarbird	69
		Bearded Vulture	70
		Sentinal Rock Thrush	72
		Barratt's Warbler	87
		Chorister Robin	116
		Buffstreaked Chat	144
		Swee Waxbill	182
		Black Harrier	223
		Grassbird	299
		Grtr Dblcrd Sunbird	350
		Cape Rock Thrush	356
		Cape Batis	363
		Groudn Woodpecker	365
		Cape Vulture	447
		Sicklewing Chat	453
2829CC	18	Drakensberg Siskin	41
		Orangebrstd Rockjumper	57
		Bush Blackcap	66
		Gurney's Sugarbird	69
		Bearded Vulture	70
		Sentinal Rock Thrush	72
		Barratt's Warbler	87
		Forest Canary	112
		Chorister Robin	116
		BuffstreakedChat	144
		Swee Waxbill	182
		Black Harrier	223
		Grassbird	299
		Grtr Dblcrd Sunbird	350
		Cape Rock Thrush	356
		Cape Batis	363
		Ground Woodpecker	365
		Cape Vulture	447
2730AC	16	Rudd's Lark	7
		Botha's Lark	17
		Yellowbrstd Pipit	18
		Bush Blackcap	66
		Sentinal Rock Thrush	72
		Buffstreaked Chat	144
		Rock Pipit	189
		Black Harrier	223
		Blue Korhaan	295
		Grassbird	299
		Grtr Dblclrd Sunbird	350
		Cape Rock Thrush	356
		Cape Batis	363
		Ground Woodpecker	365
		Cloud Cisticola	396
		Blue Crane	472
3127DC	12	Rudd's Lark	7
		Yellowbreasted Pipit	18
		Starred Robin	60
		Bush Blackcap	66
		Rock Pipit	189
		Rdfrtd Tinker Barbet	220
		Blue Korhaan	295
		Grtr Dblclrd Sunbird	350
		Cape Rock Thrush	356
		Cape Batis	363
		Ground Woodpecker	365
		Cape Vulture	447
2430DD	12	Starred Robin	60
		Gurney's Sugarbird	69
		Barratt's Warbler	87
		Forest Canary	112
		Chorister Robin	116
		Striped Pipit	137
		Swee Waxbill	182
		Olive Bush Shrike	186
		Grassbird	299
		Grtr Dblcrd Sunbird	350
		Cape Rock Thrush	356
		Cape Batis	363

3020CB	8	Red Lark	55
		Sclater's Lark	70
		Stark's Lark	191
		Karoo Eremomela	222
		Blackeared Finchlark	246
		Southrn Grey Tit	380
		Sicklewing Chat	453
		Tractrac Chat	492
1932BB	8	Starred Robin	60
		Gurney's Sugarbird	69
		Barratt's Warbler	87
		Striped Pipit	137
		Olive Bush Shrike	186
		Grassbird	299
		Cape Batis	363
		String Barrd Warbler	401
1922AA	8	Dickinson's Kestrel	217
		Bennett's Woodpecker	231
		Cape Penduline Tit	319
		Pearlbreasted Swallow	342
		Burntnckd Eremomela	365
		Barred Warbler	455
		Doublbnd Sandgrouse	466
		Longtailed Starling	499
1728BB	8	Blackeared Canary	86
		Racketttailed Roller	153
		Dickinson's Kestrel	217
		Bennett's Woodpecker	231
		Arnot's Chat	231
		String Barrd Warbler	401
		Doublbnd Sandgrouse	466
		Longtailed Starling	499
1835DA	2	Shorttaild Pipit	4
		String Barrd Warbler	401
2615AA	1	Dune Lark	37

Figure 1: Species richness for all terrestrial species using detection probability deciles (equation 7)

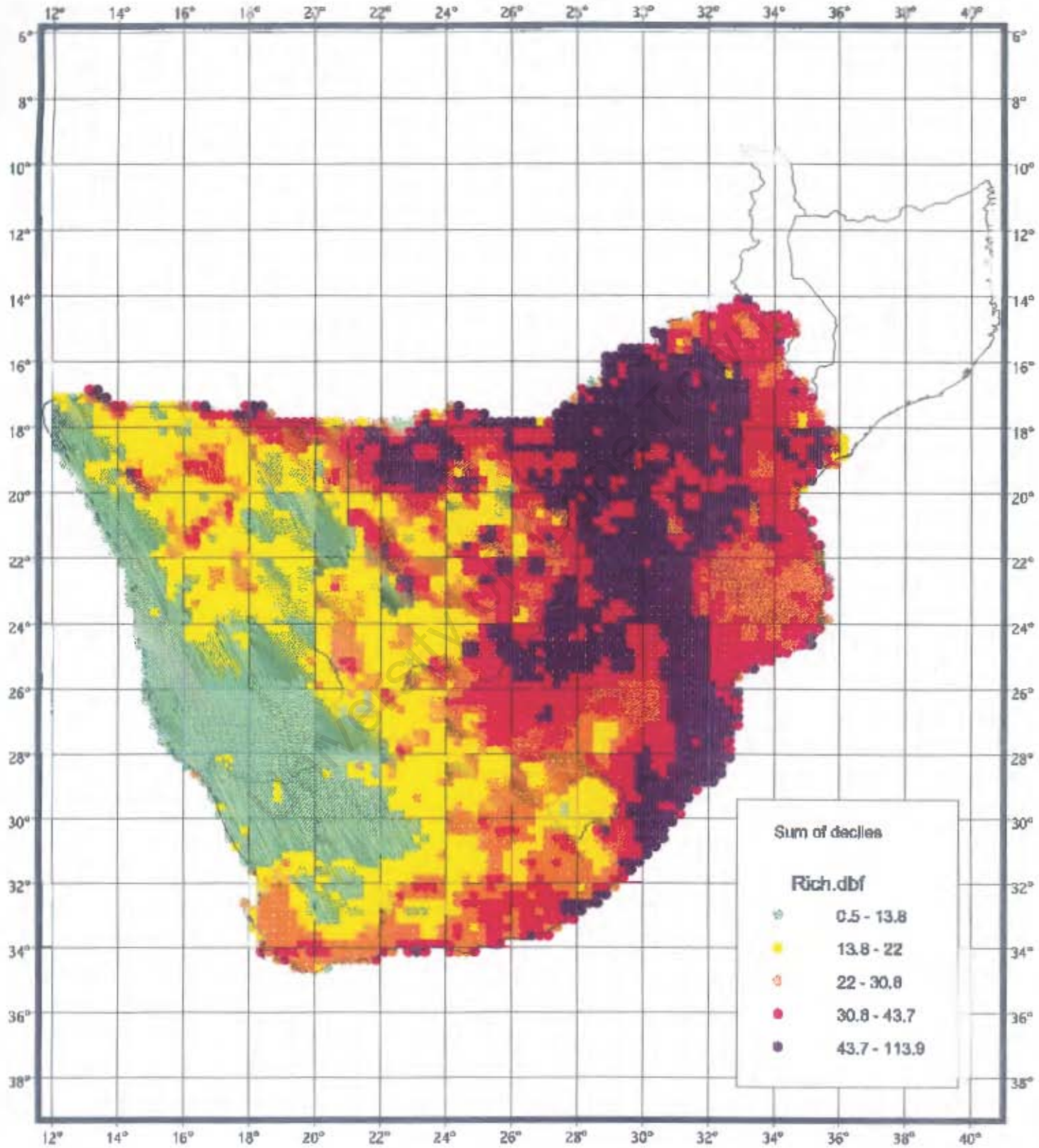


Figure 2: Species richness for all terrestrial species using presence-absence data (equation 1)

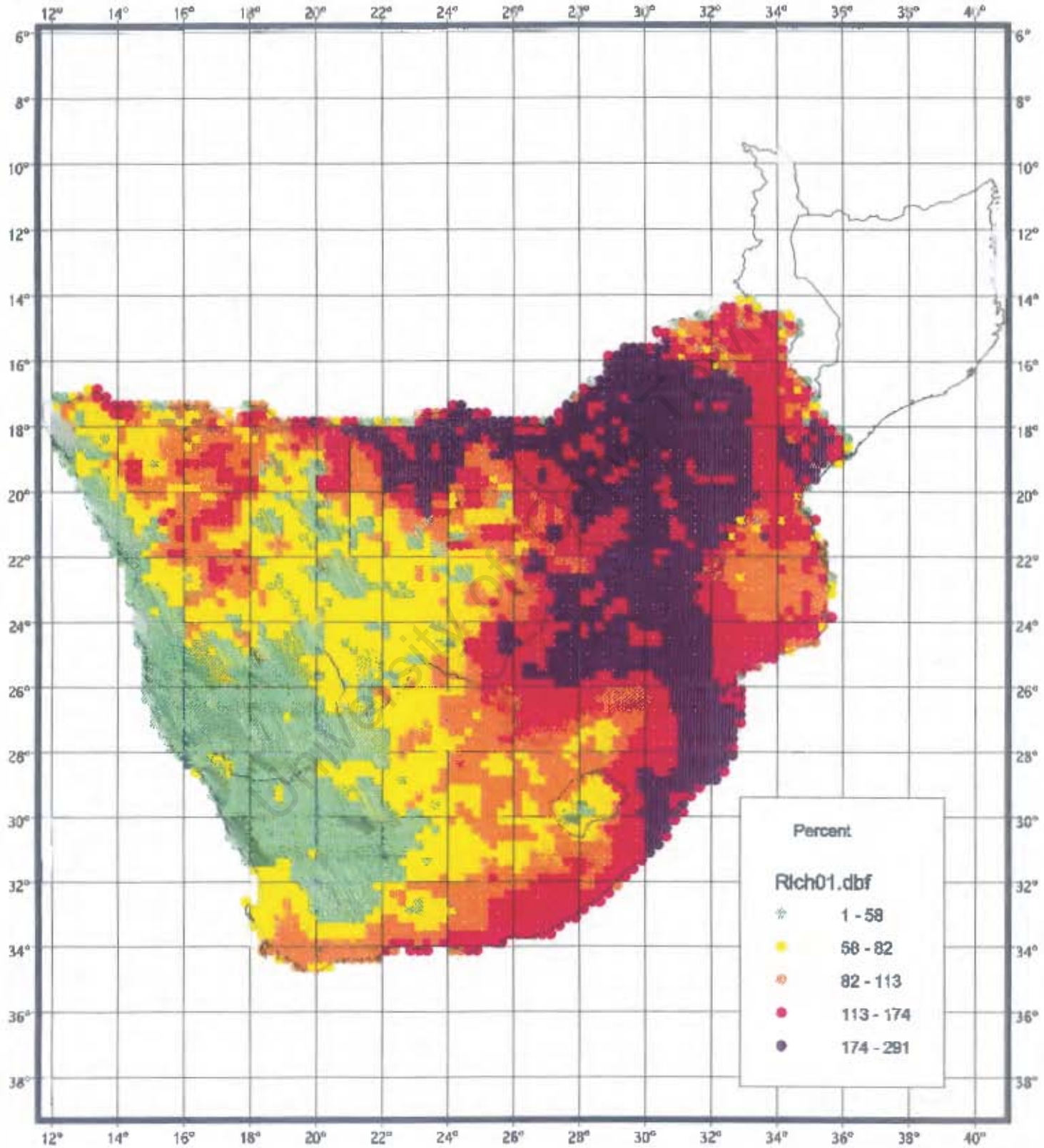


Figure 3: Comparing species richness using detection probability deciles and using presence-absence data (equation 7 versus equation 1)

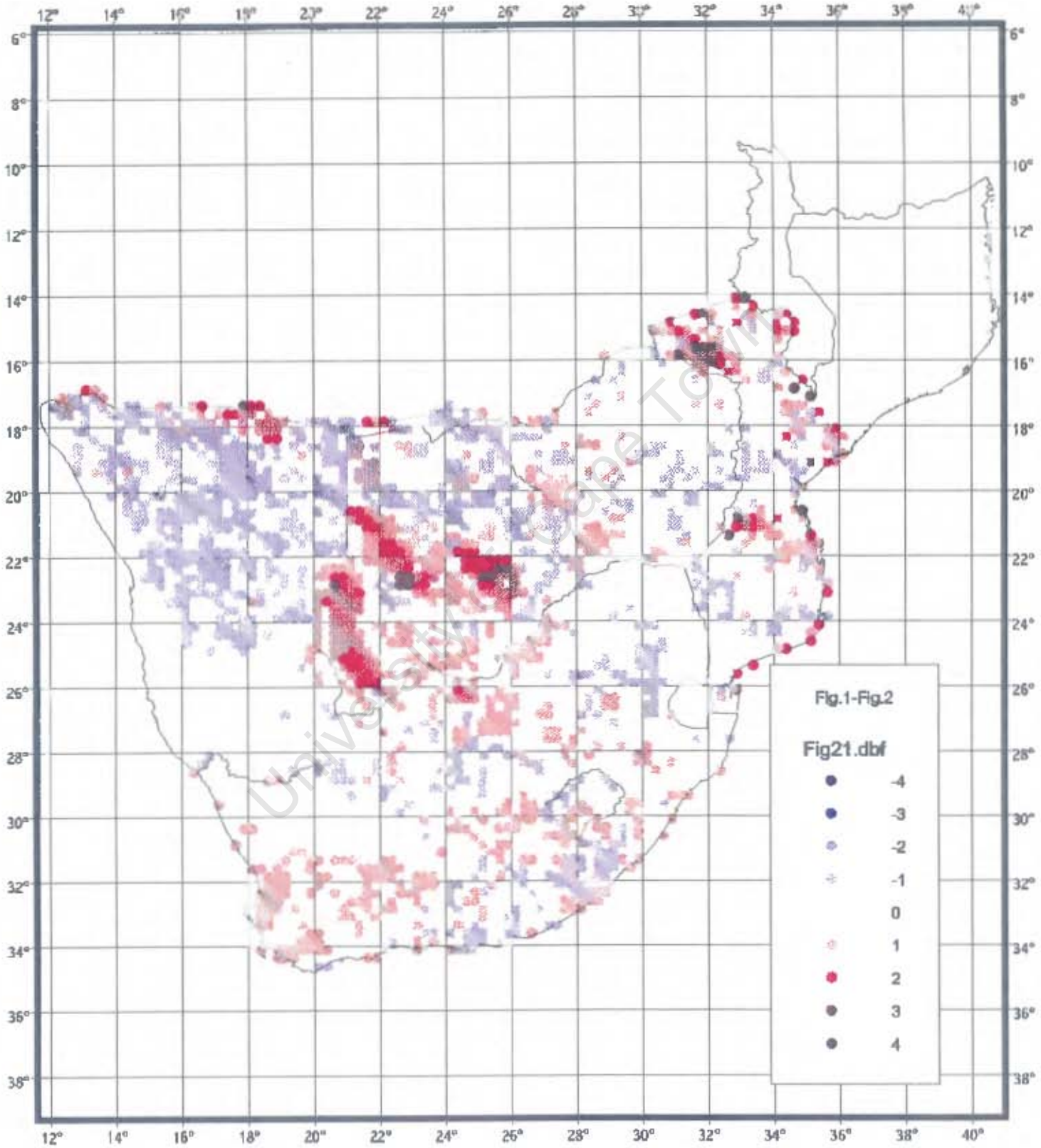


Figure 4: Species richness for all terrestrial species in the 20% core of their distributions derived from detection probability deciles (equation 9)

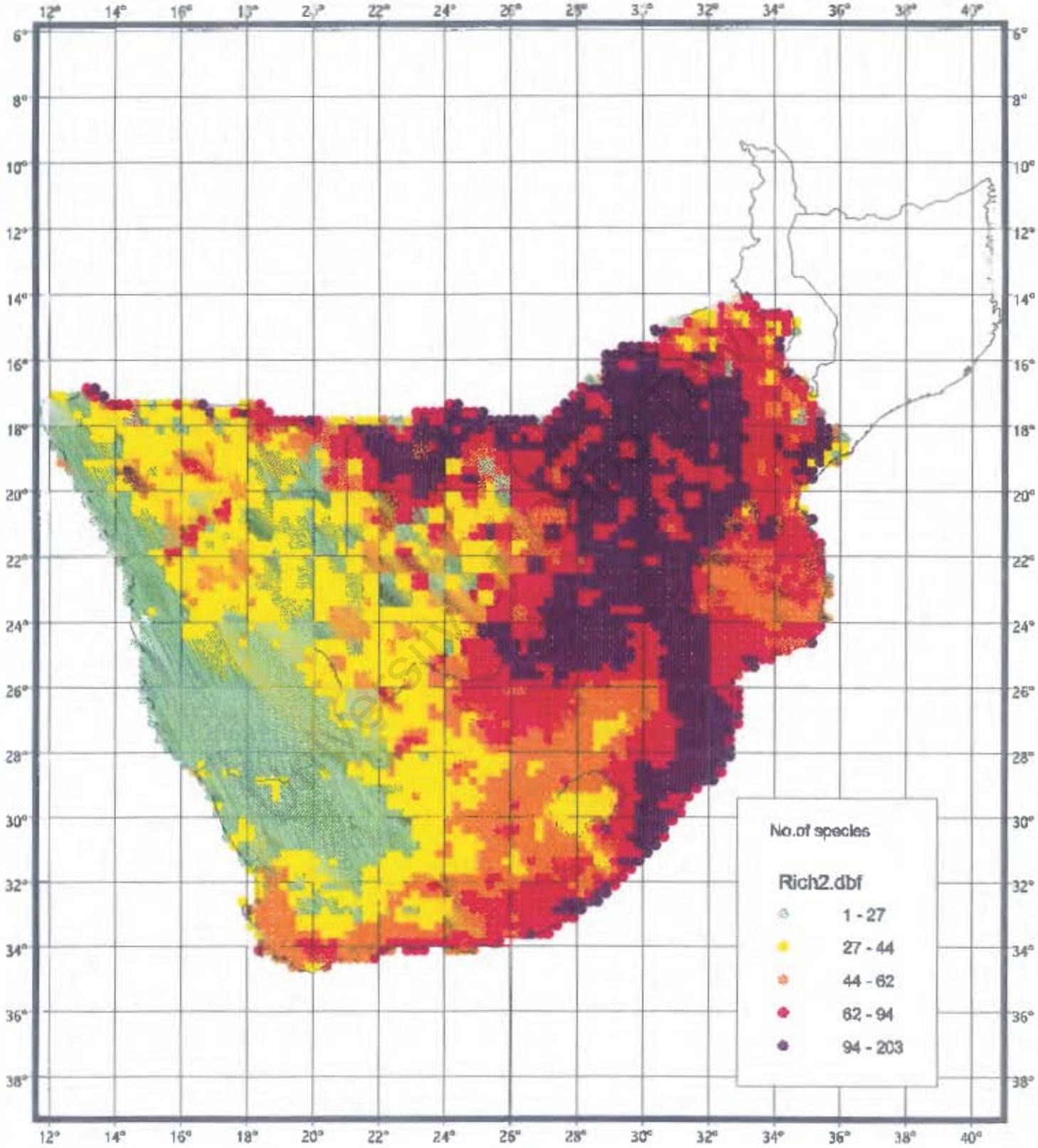


Figure 5: Species richness for all terrestrial species in the 40% core of their distributions derived from detection probability deciles (equation 9)

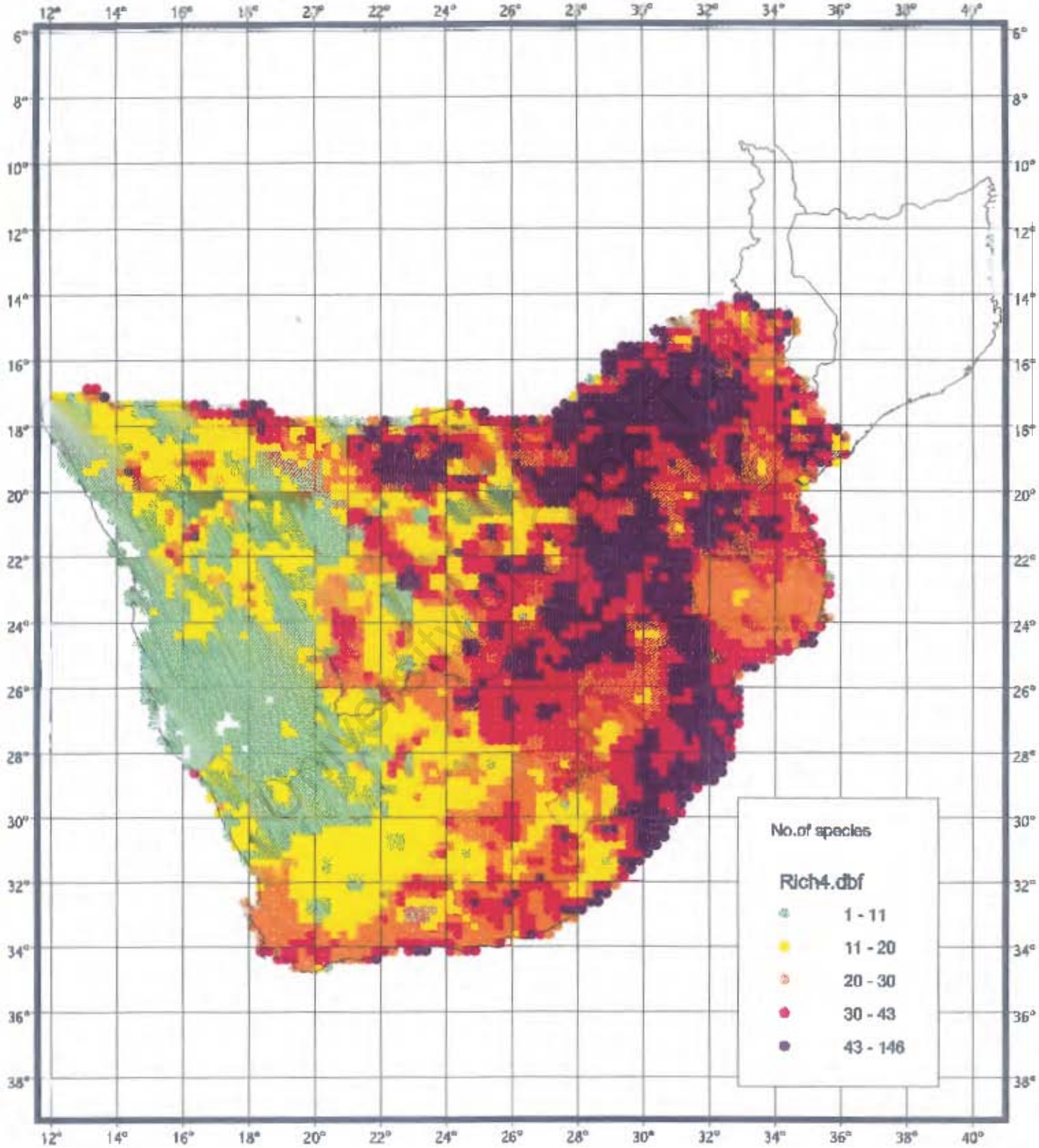


Figure 6: Species richness for all terrestrial species in the 50% core of their distributions derived from detection probability deciles (equation 9)

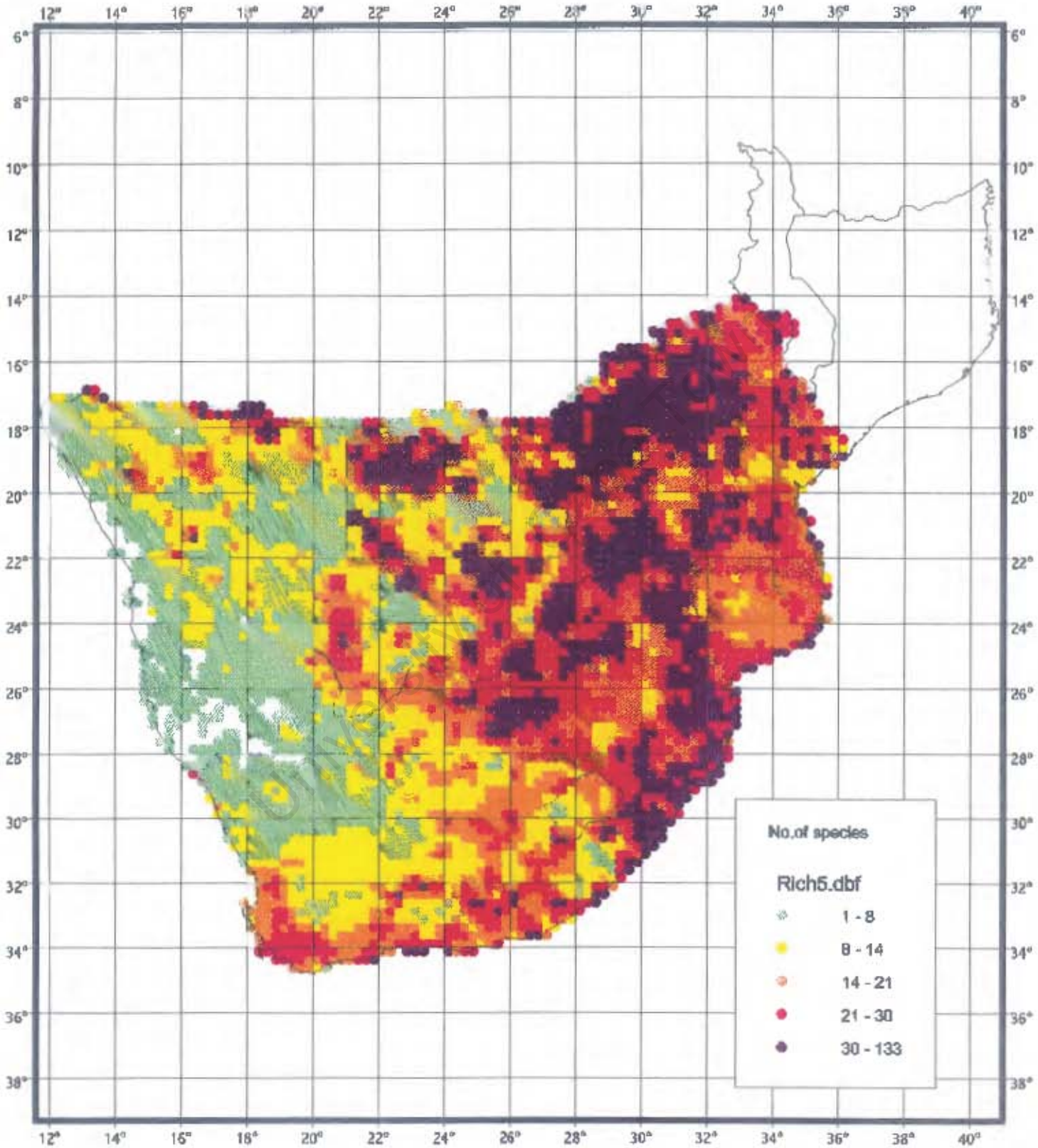


Figure 7: Species richness for all terrestrial species in the 60% core of their distributions derived from detection probability deciles (equation 9)

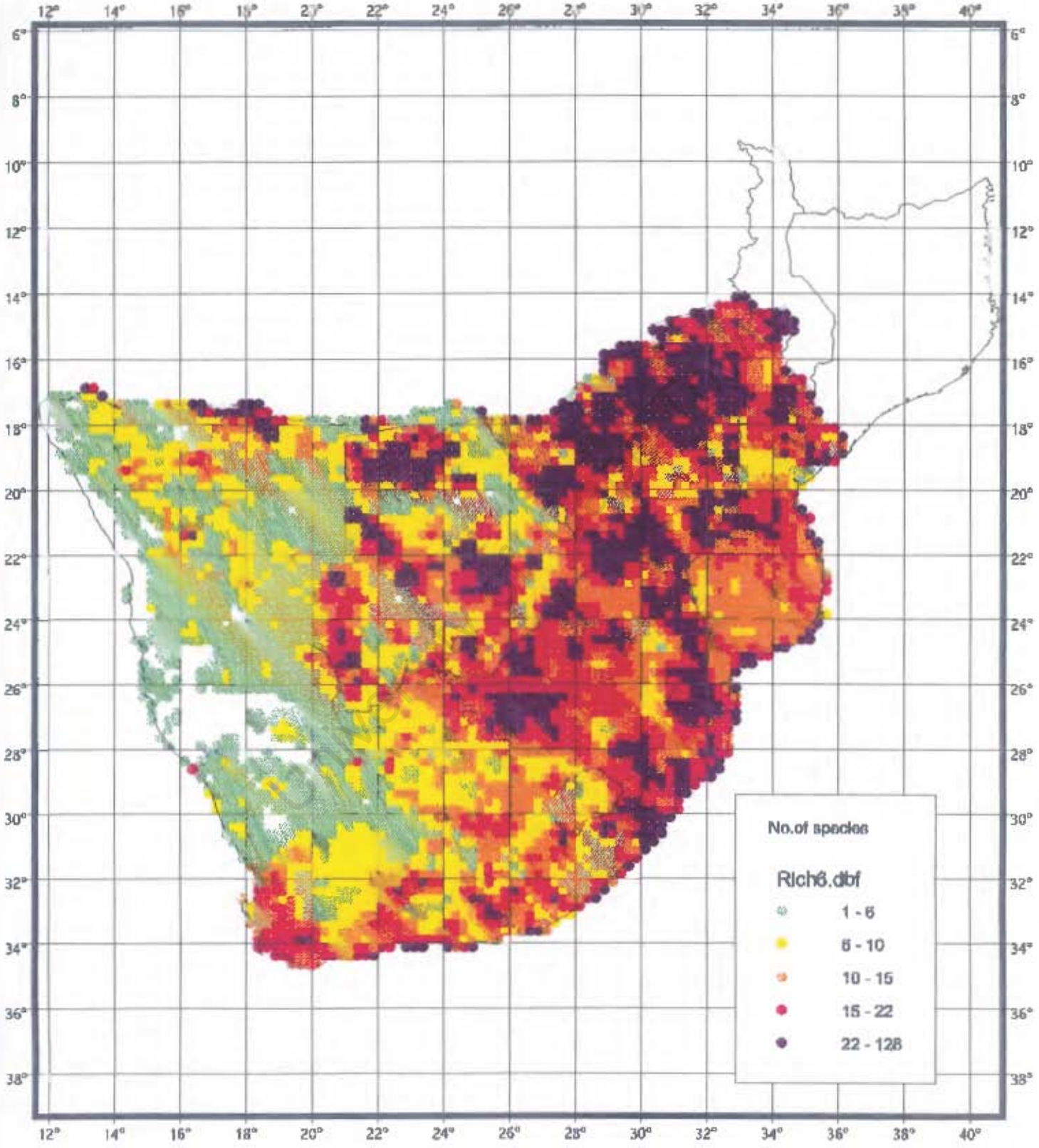


Figure 8: Species richness for all terrestrial species in the 80% core of their distributions derived from detection probability deciles (equation 9)

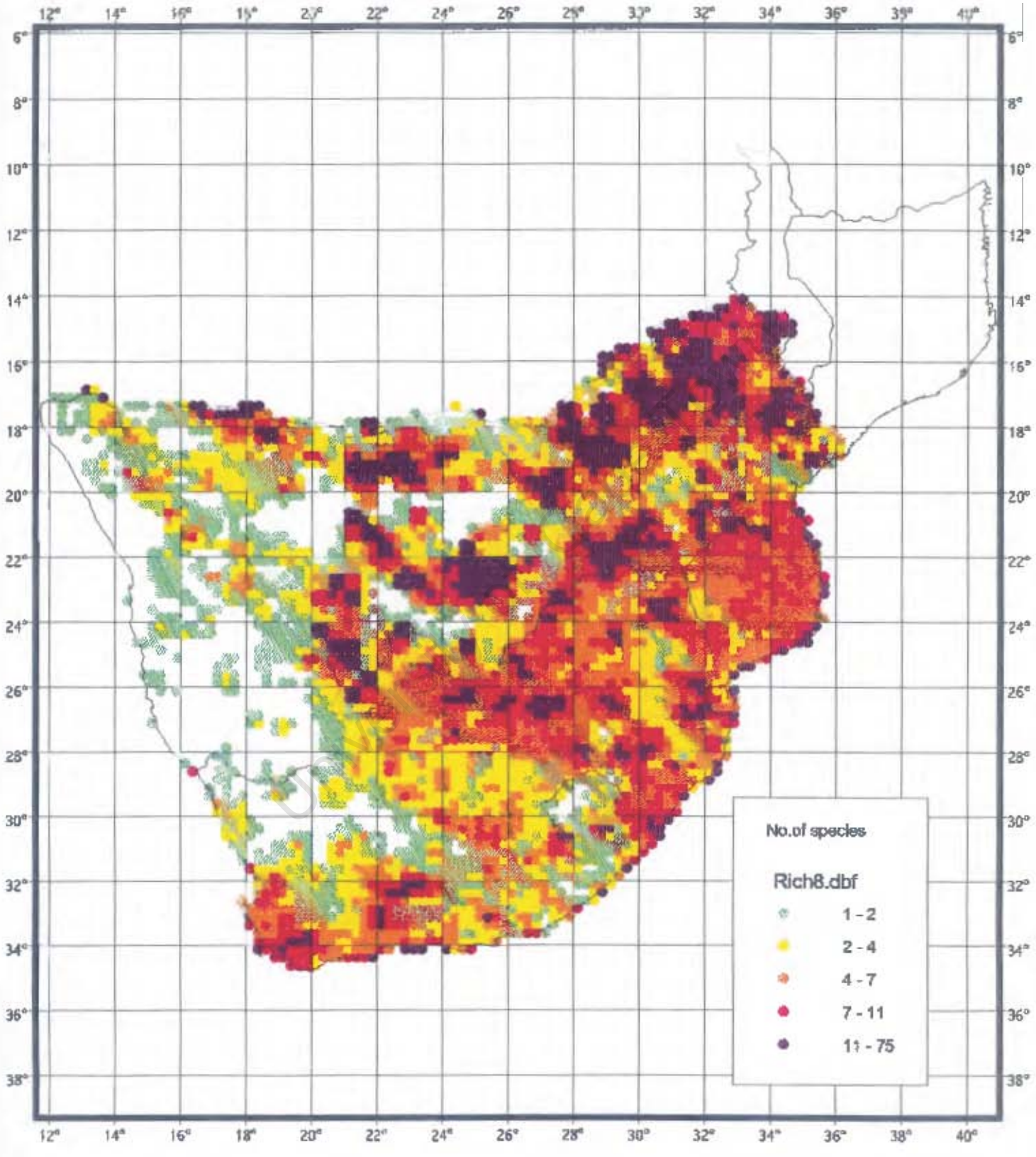


Figure 9: Distribution of apparent narrow endemism for terrestrial species using detection probability deciles (equation 14)

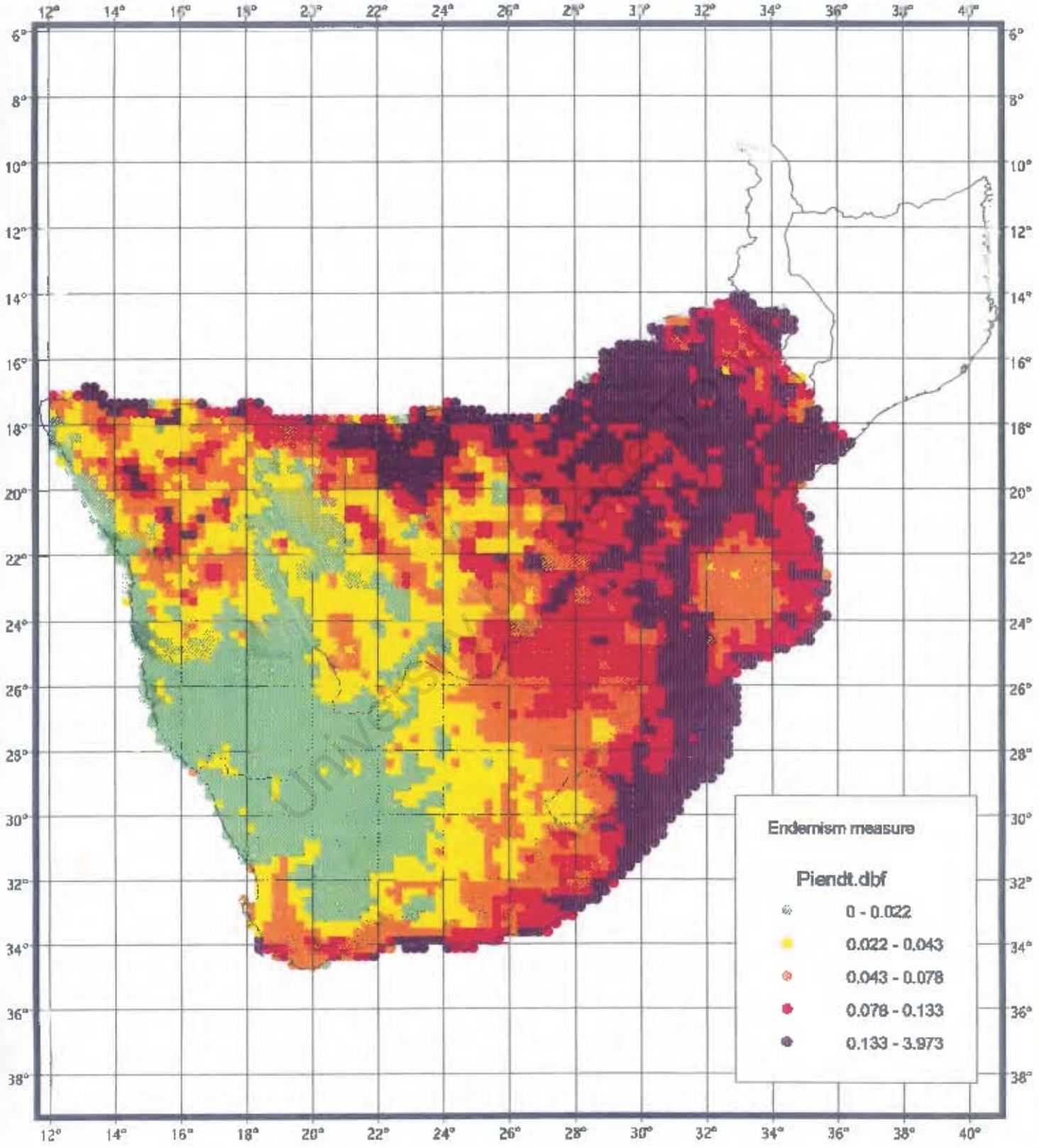


Figure 10: Comparing overall species richness to apparent narrow endemism for terrestrial species using detection probability deciles (equation 14 versus equation 7)

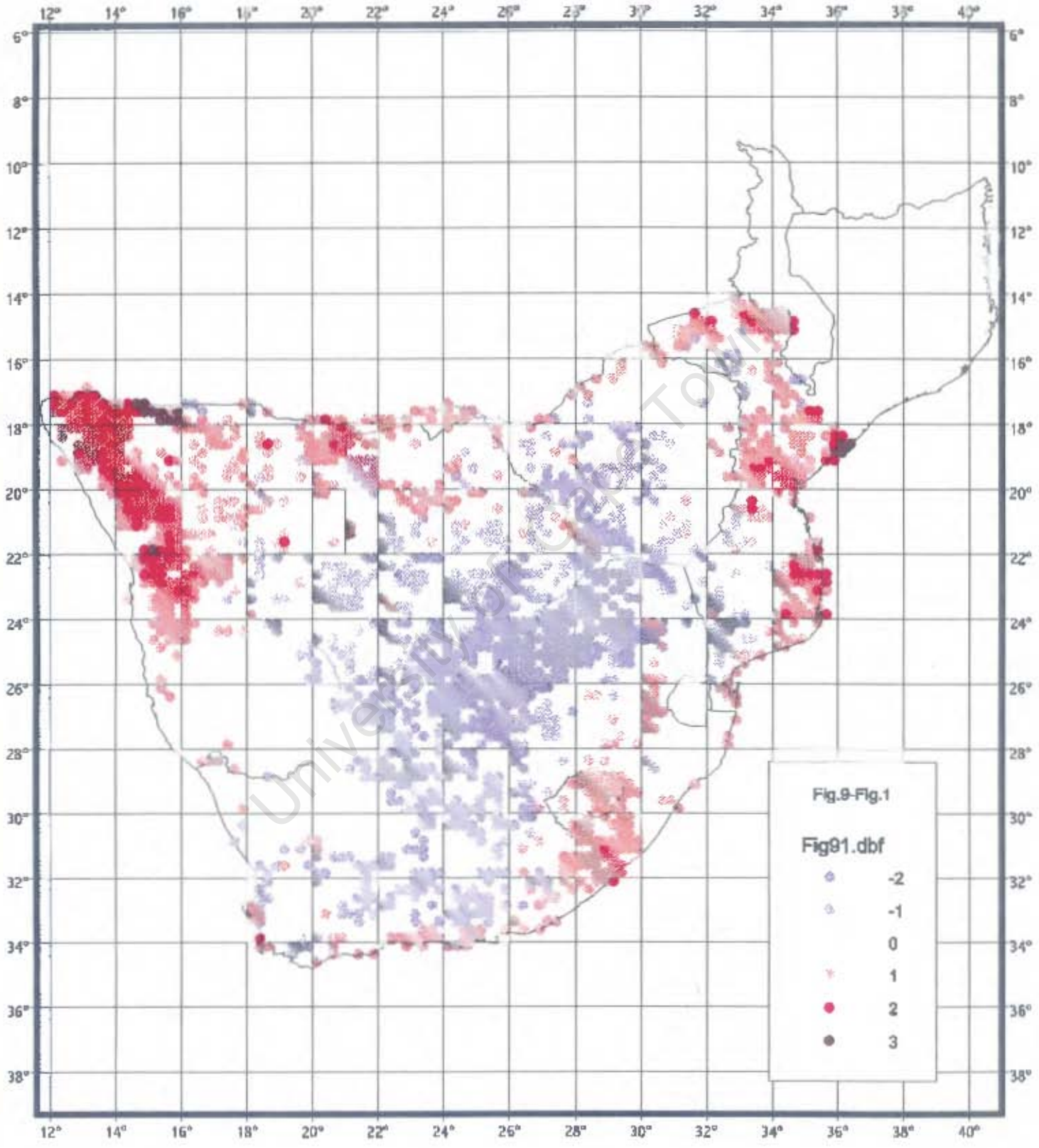


Figure 11: Top 20% of distribution of apparent narrow endemism for terrestrial species using detection probability deciles (equation 14)

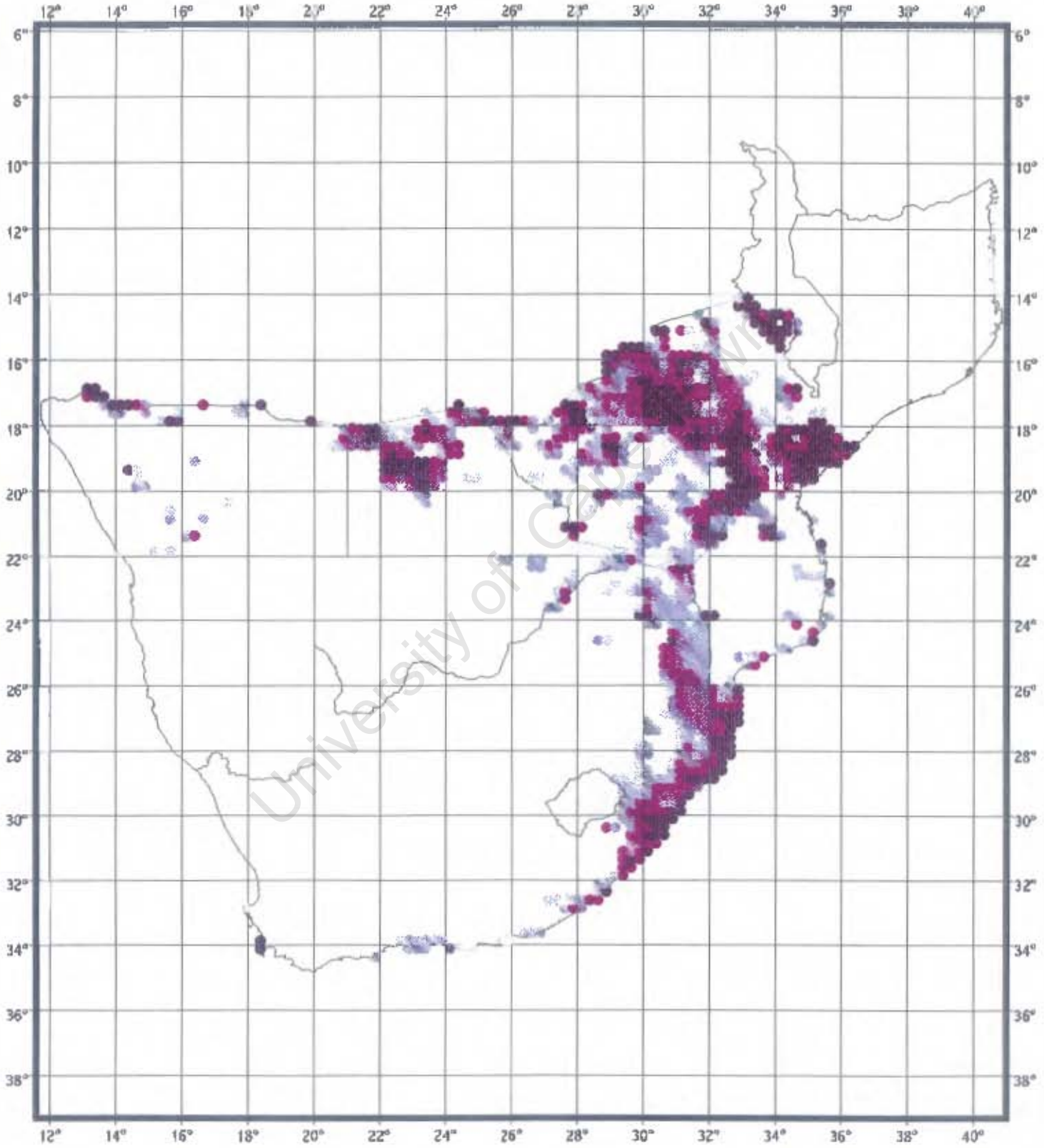


Figure 12: Species richness for southern African endemic species using detection probability deciles (equation 7)

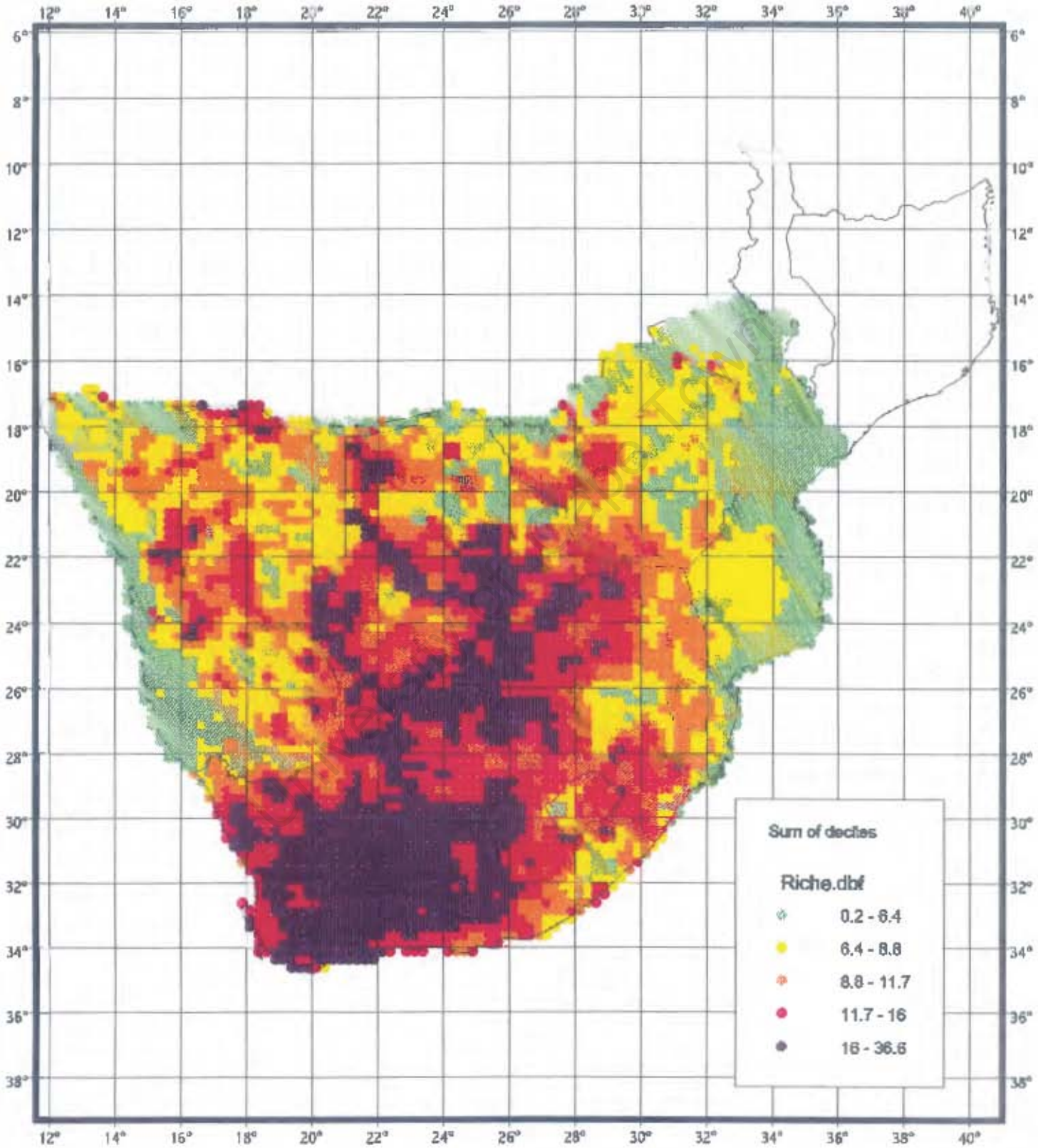


Figure 13: Species richness for southern African endemic species using presence-absence data (equation 1)

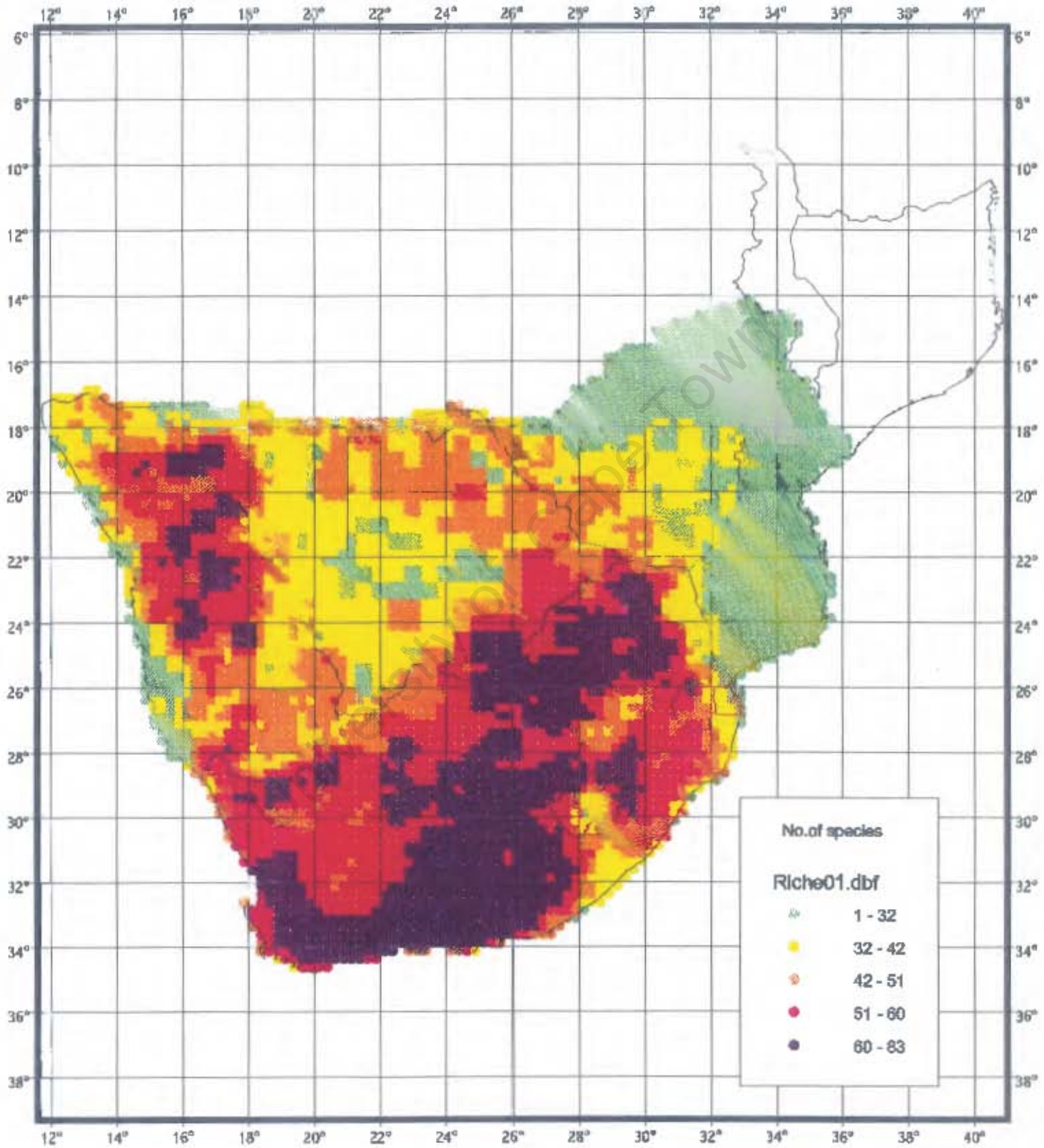


Figure 14: Comparing species richness for southern African endemic species using detection probability deciles and using presence-absence data (equation 7 versus equation 1)

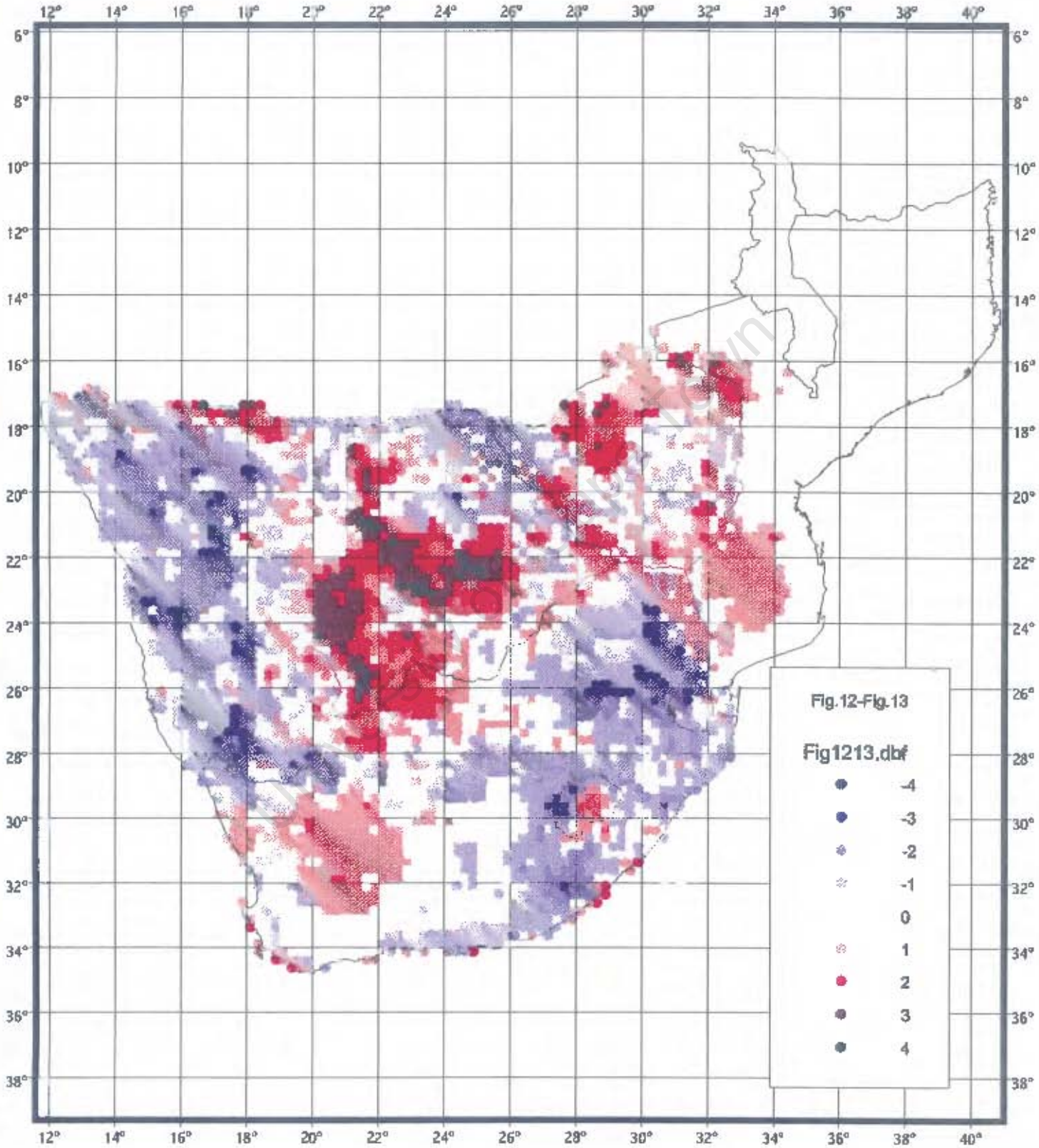


Figure 15: Centres of endemism for southern African endemic species using detection probability deciles (equation 14)

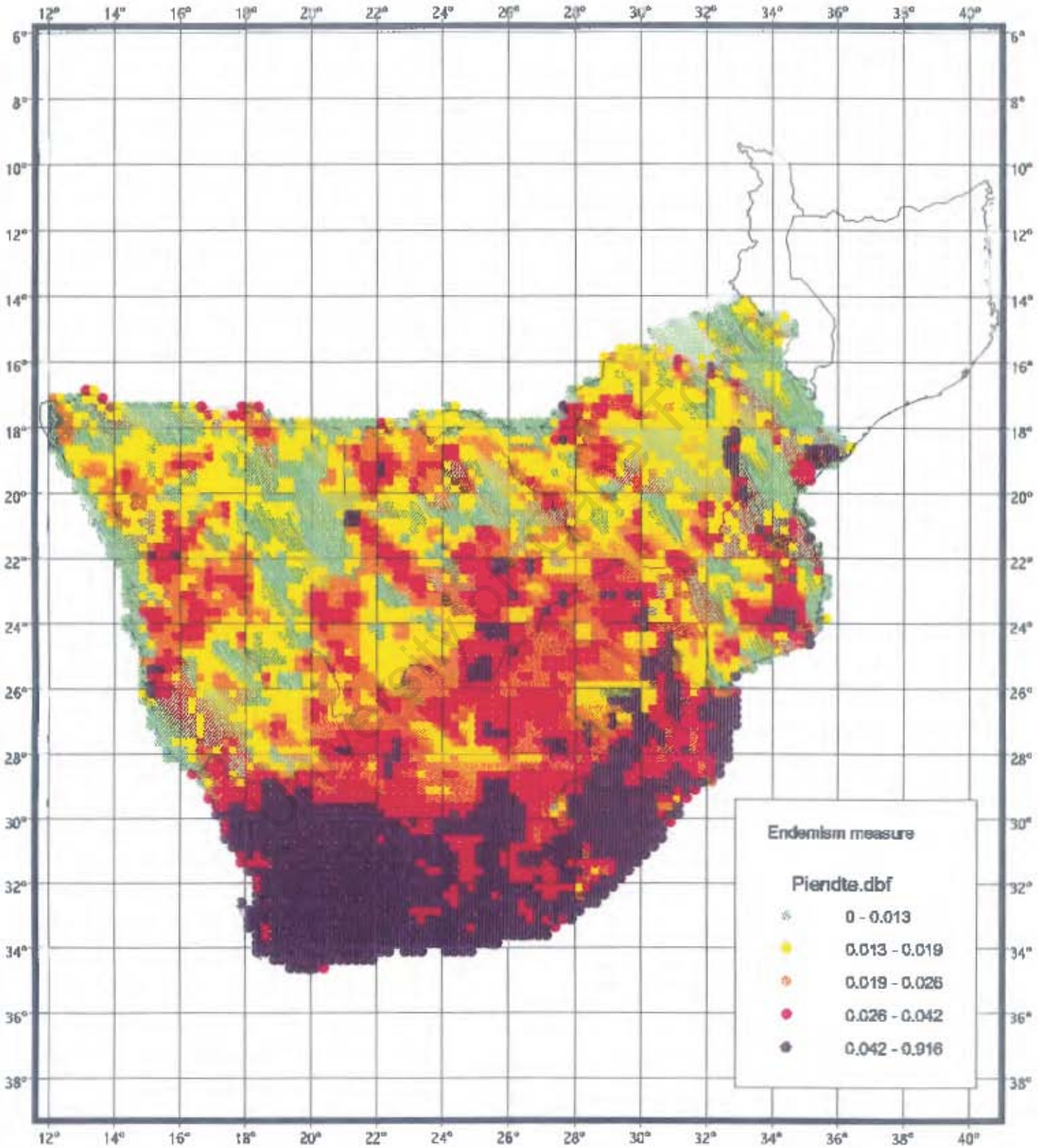


Figure 16: Comparing species richness to centres of endemism for southern African endemic species using detection probability deciles (equation 14 versus equation 7)

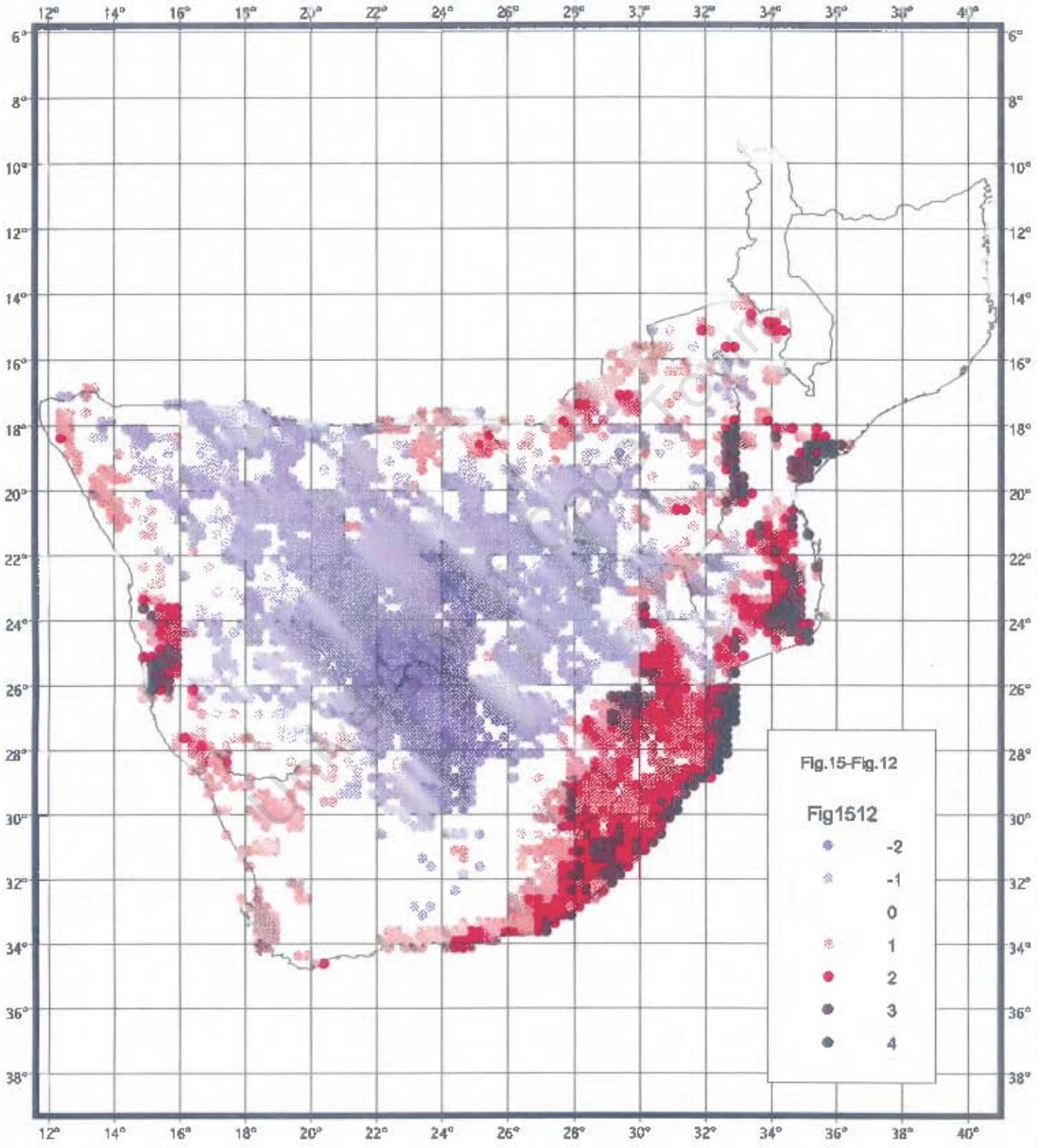


Figure 17: Centres of endemism for southern African endemic species using presence-absence data (equation 4)

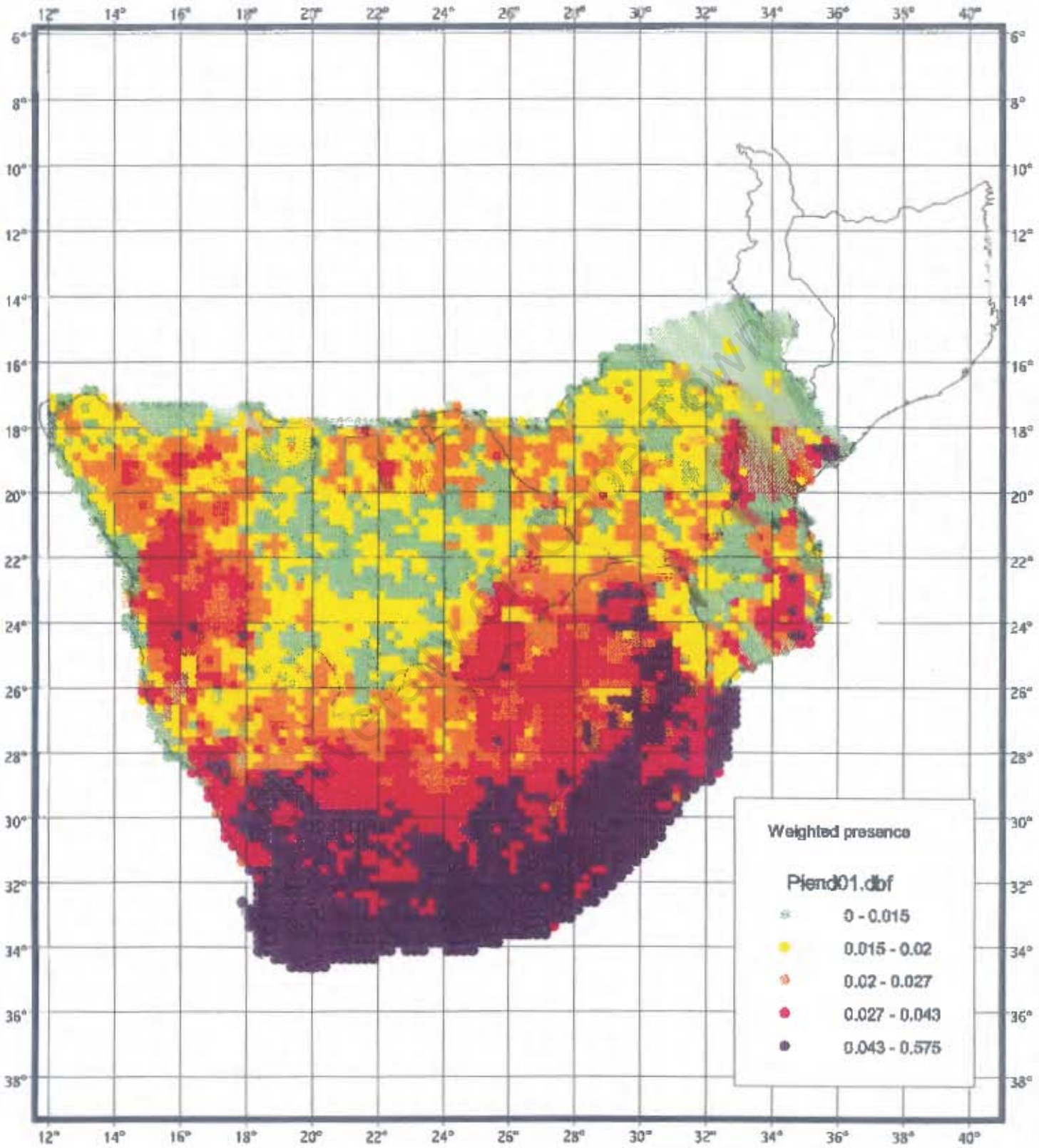


Figure 18: Comparing centres of endemism for southern African endemic species using detection probability deciles and using presence-absence data (equation 14 versus equation 4)

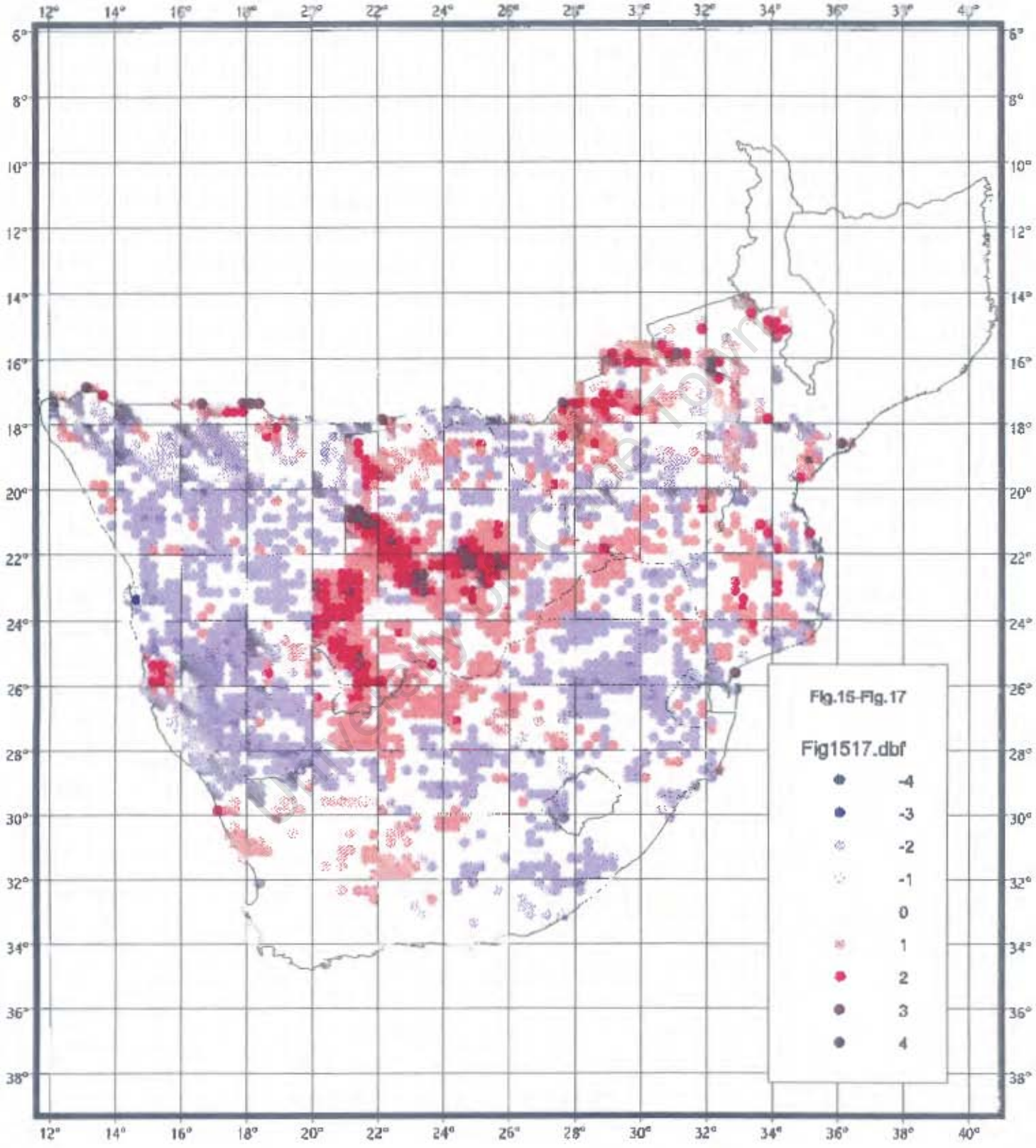


Figure 19: Top 20% of centres of endemism for southern African endemic species using detection probability deciles (equation 14)

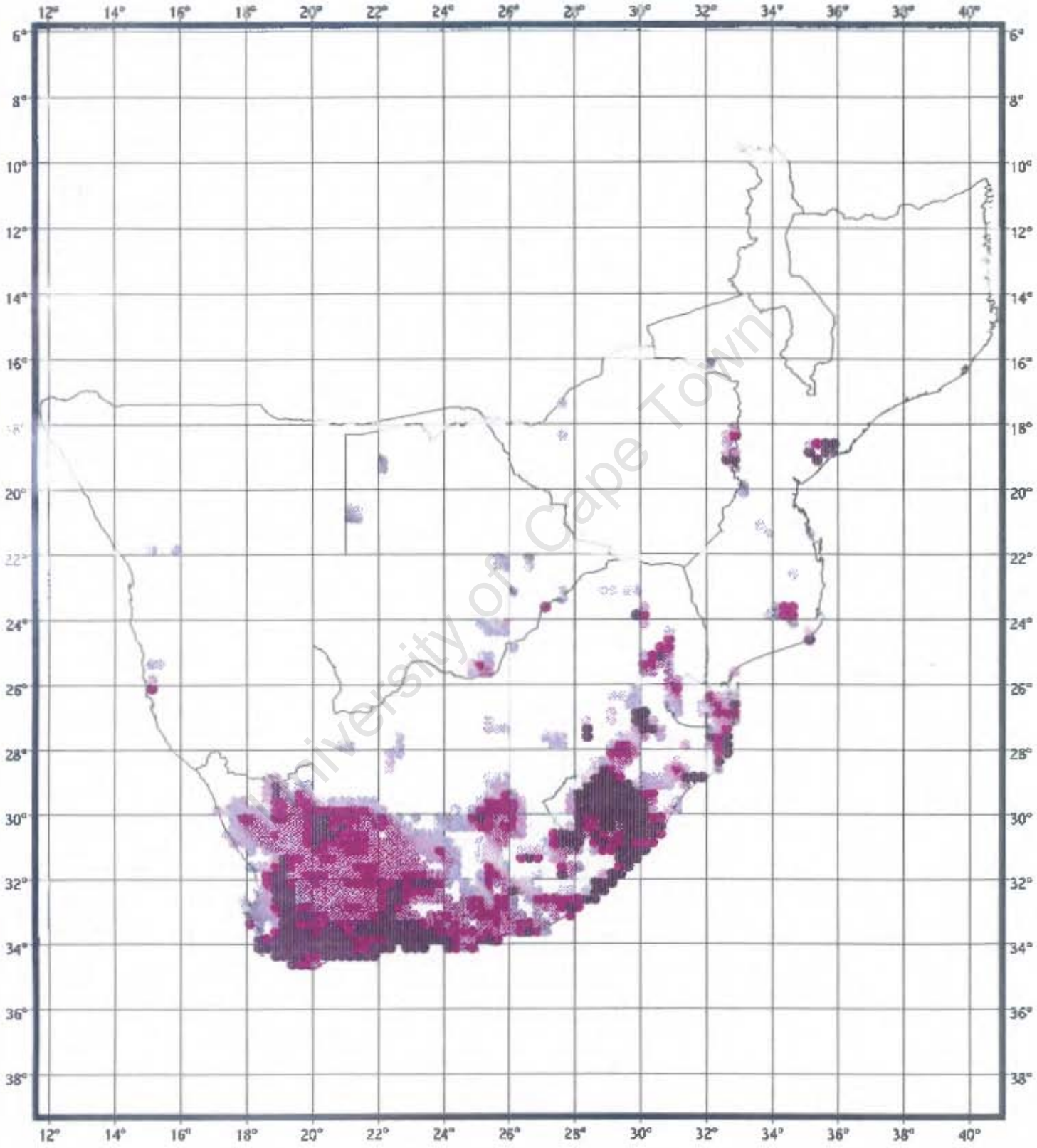


Figure 20: Top 20% of centres of endemism for southern African endemic species using presence-absence data (equation 4)

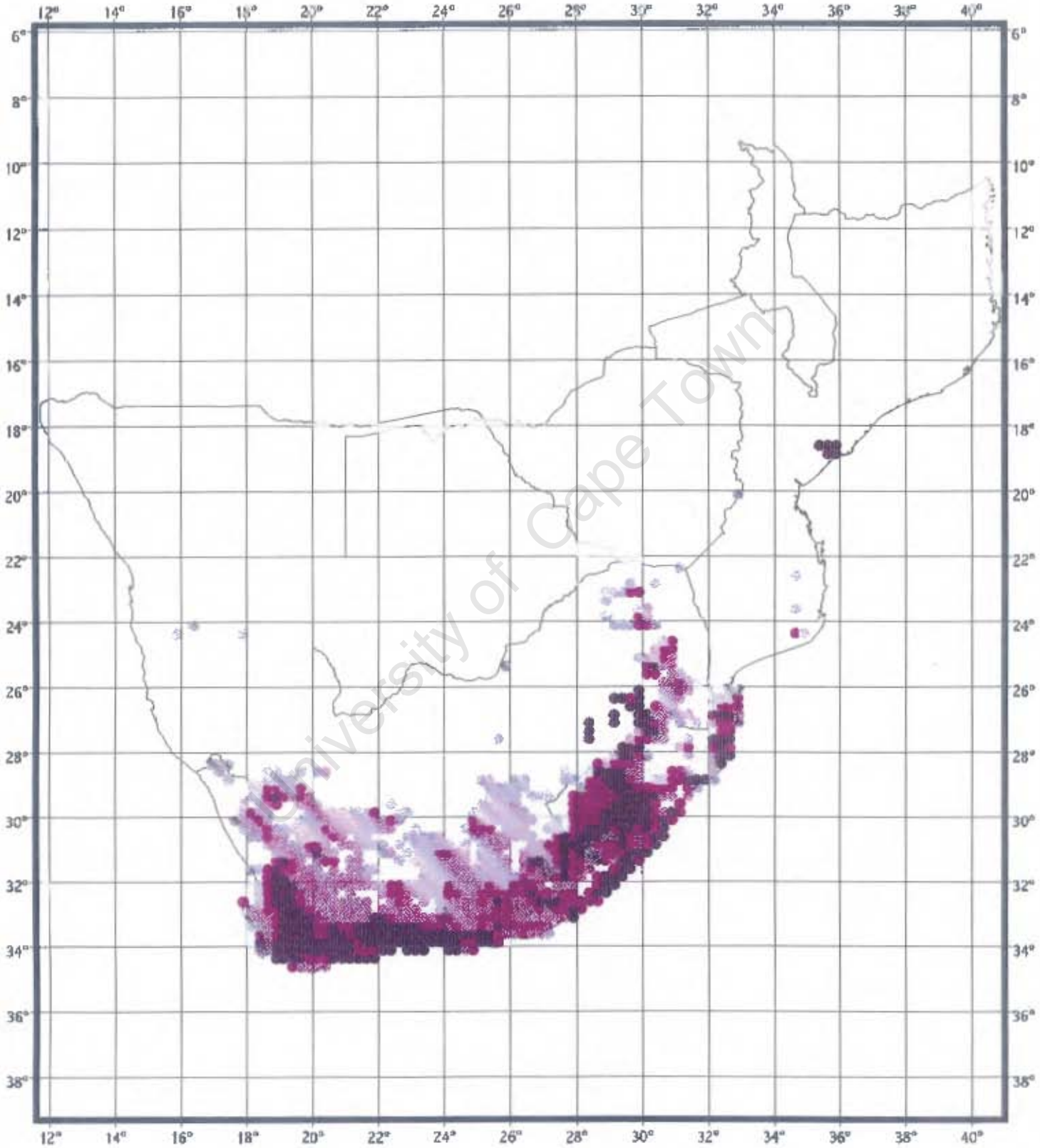


Figure 21: Comparing top 20% of centres of endemism for southern African endemic species using detection probability deciles and using presence-absence data (equation 14 versus equation 4)

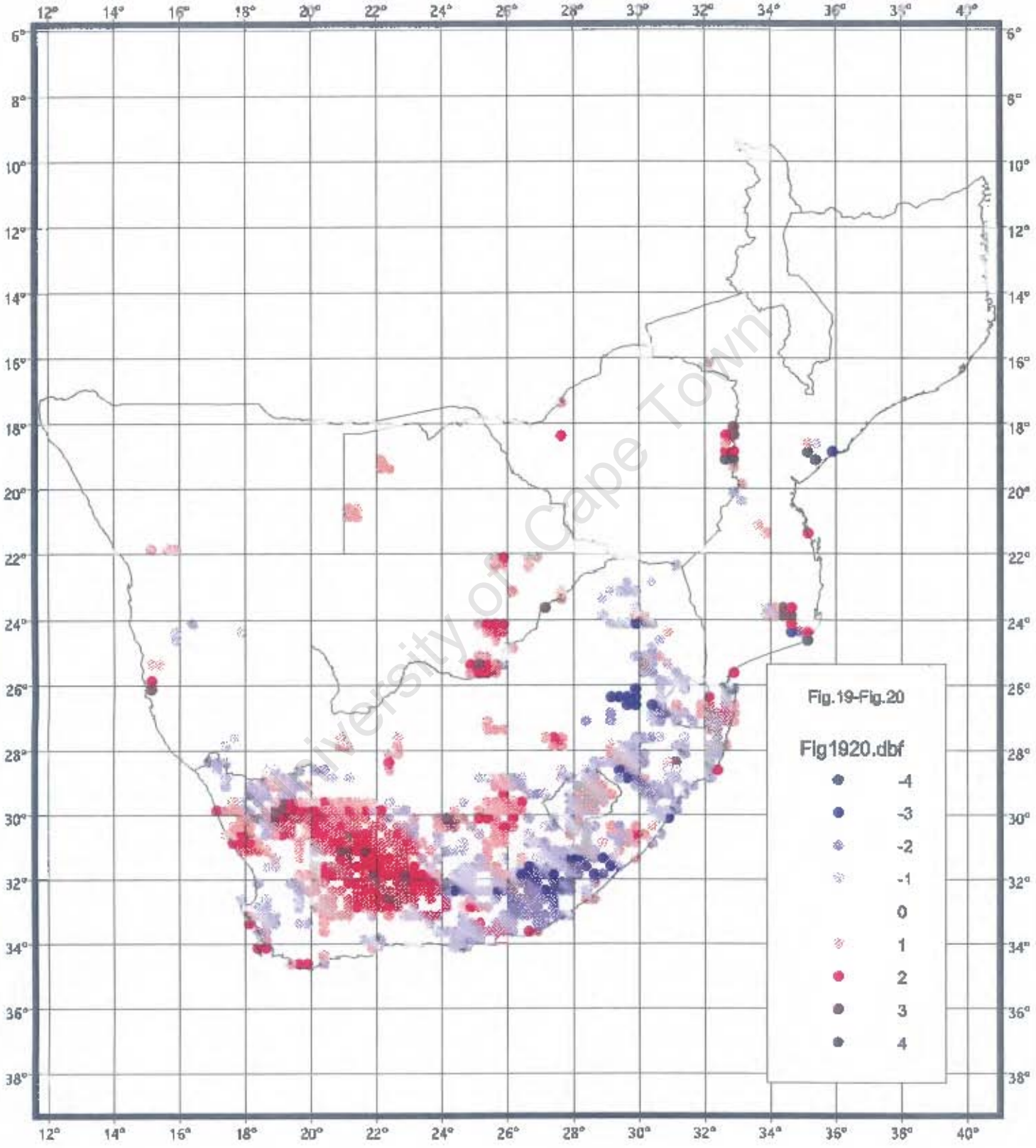


Figure 22: Distributions of core ranges for southern African endemic species

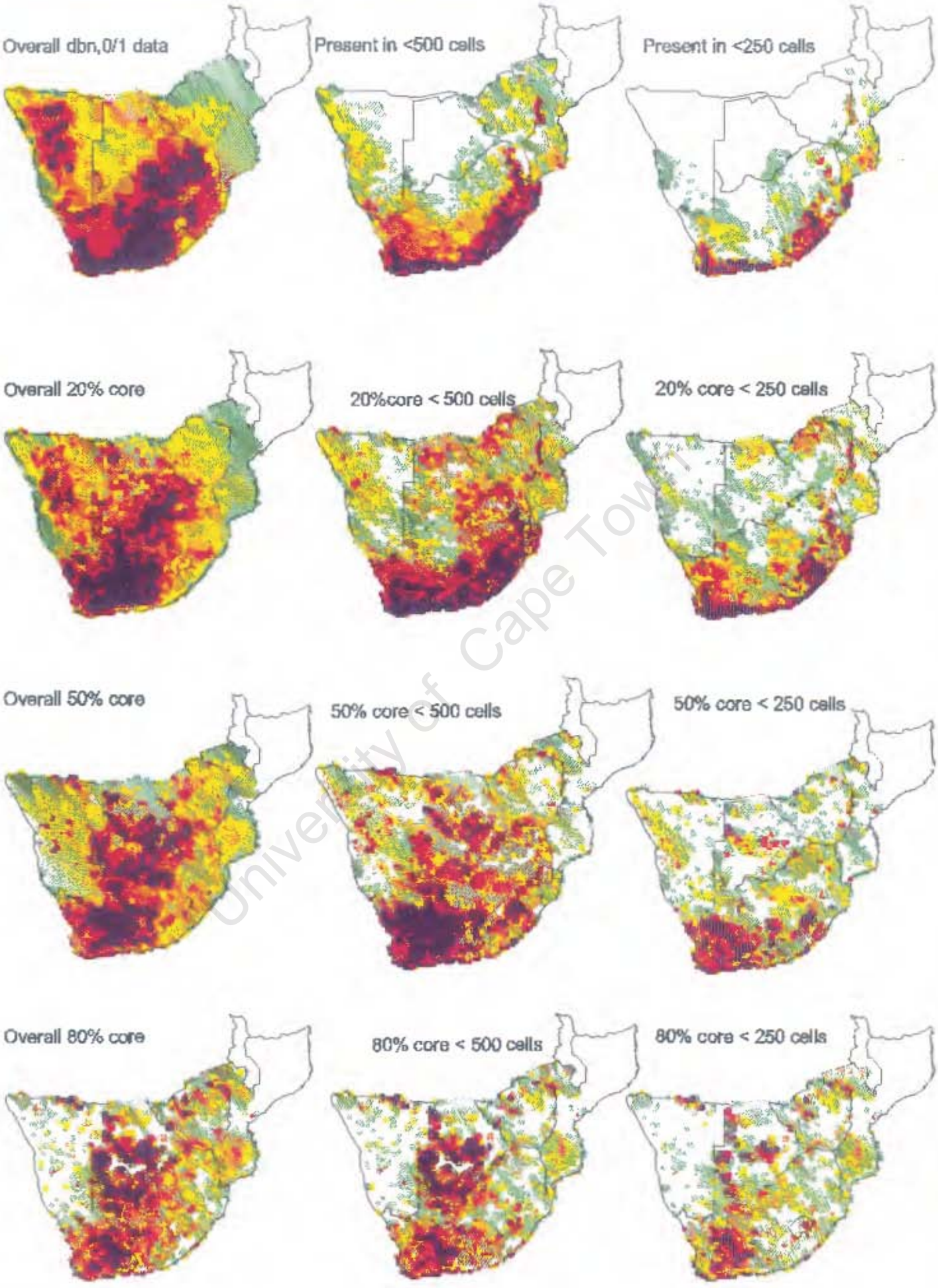


Figure 23: Species richness for southern African endemic species in the 20% core of their distributions with a restricted distribution of fewer than 500 grid cells (equation 12)

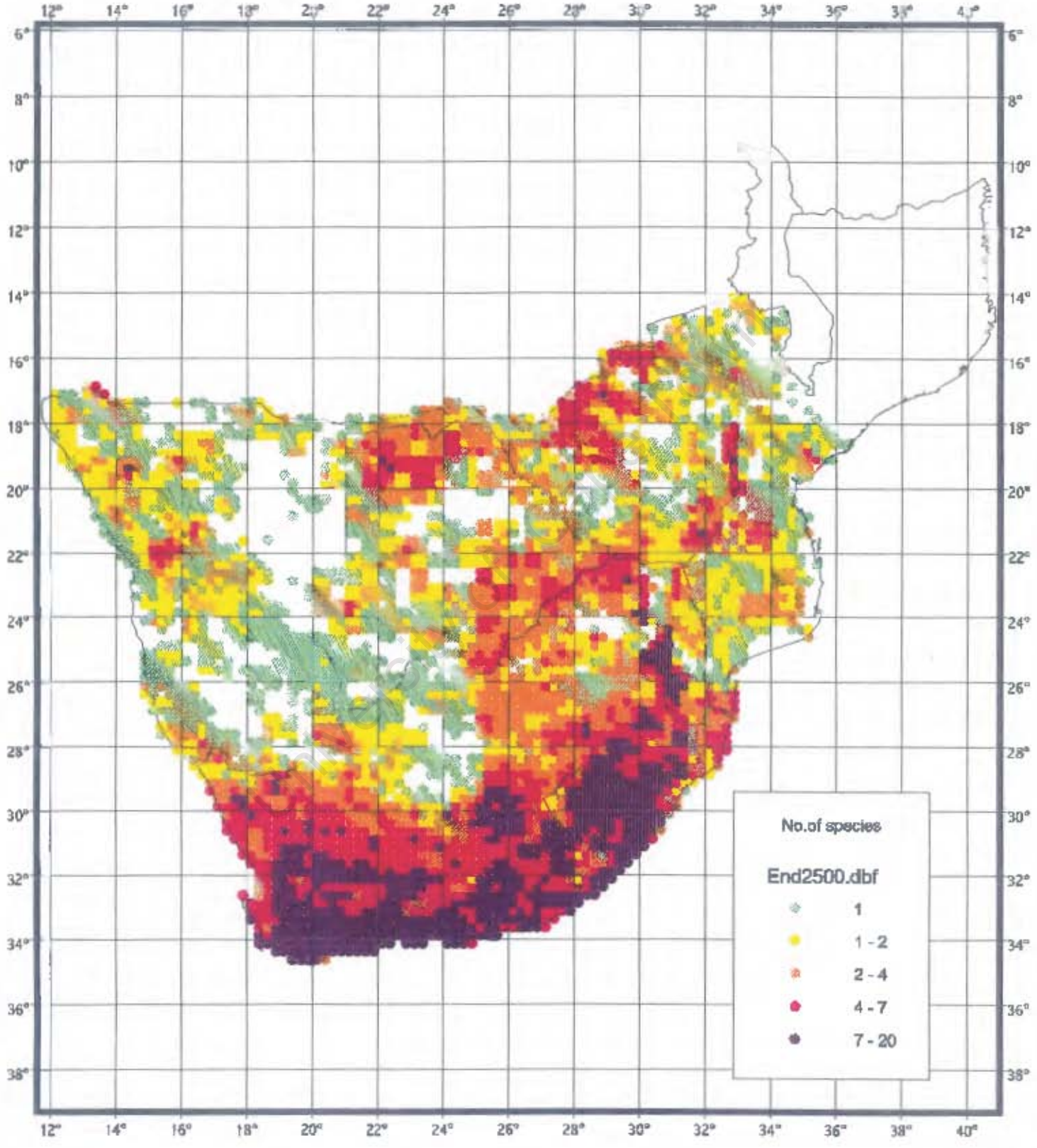


Figure 24: Ratio of species with restricted distributions of fewer than 500 grid cells to overall species richness for southern African endemic species in the 20% core of their distributions (equation 15)

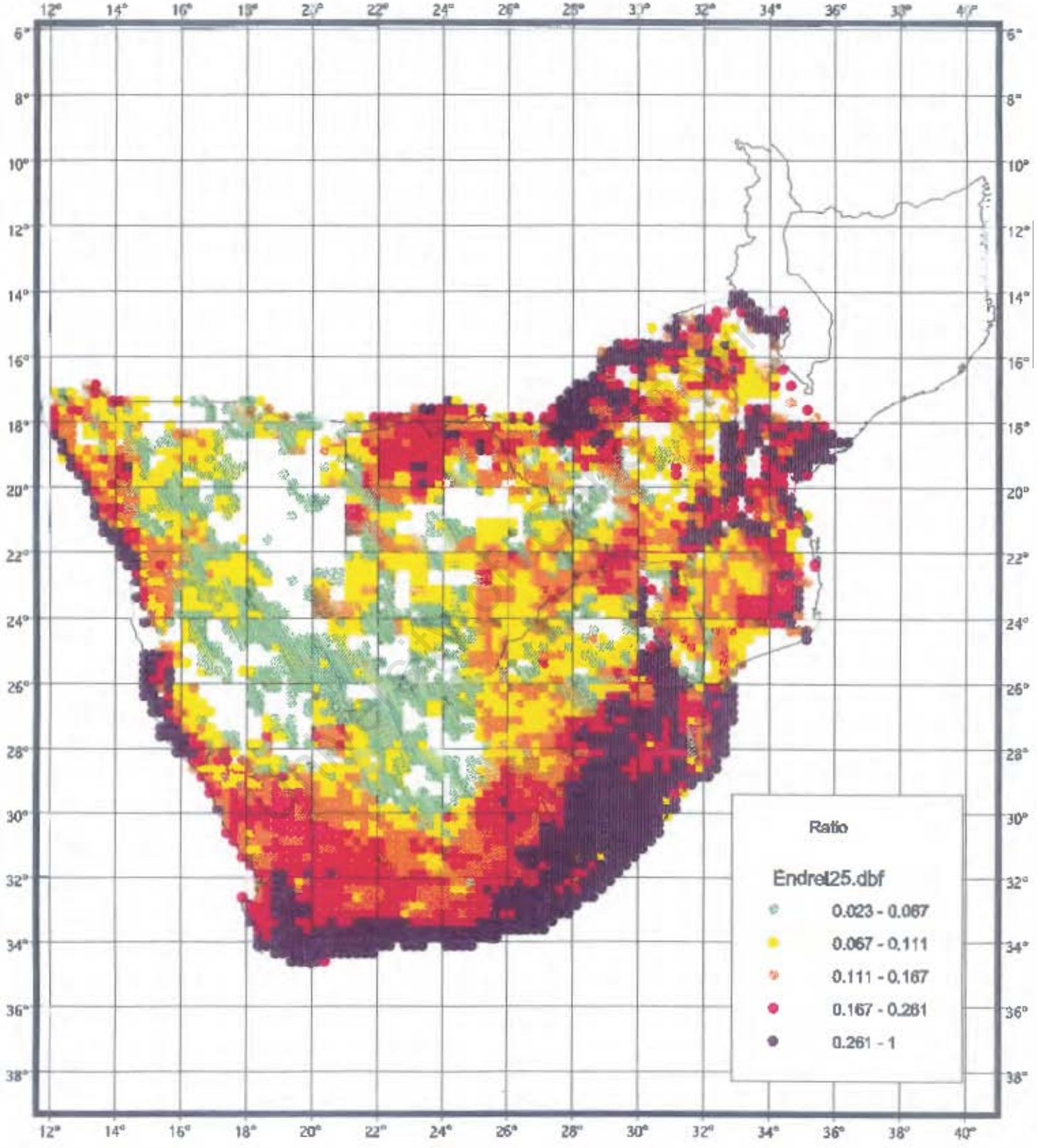
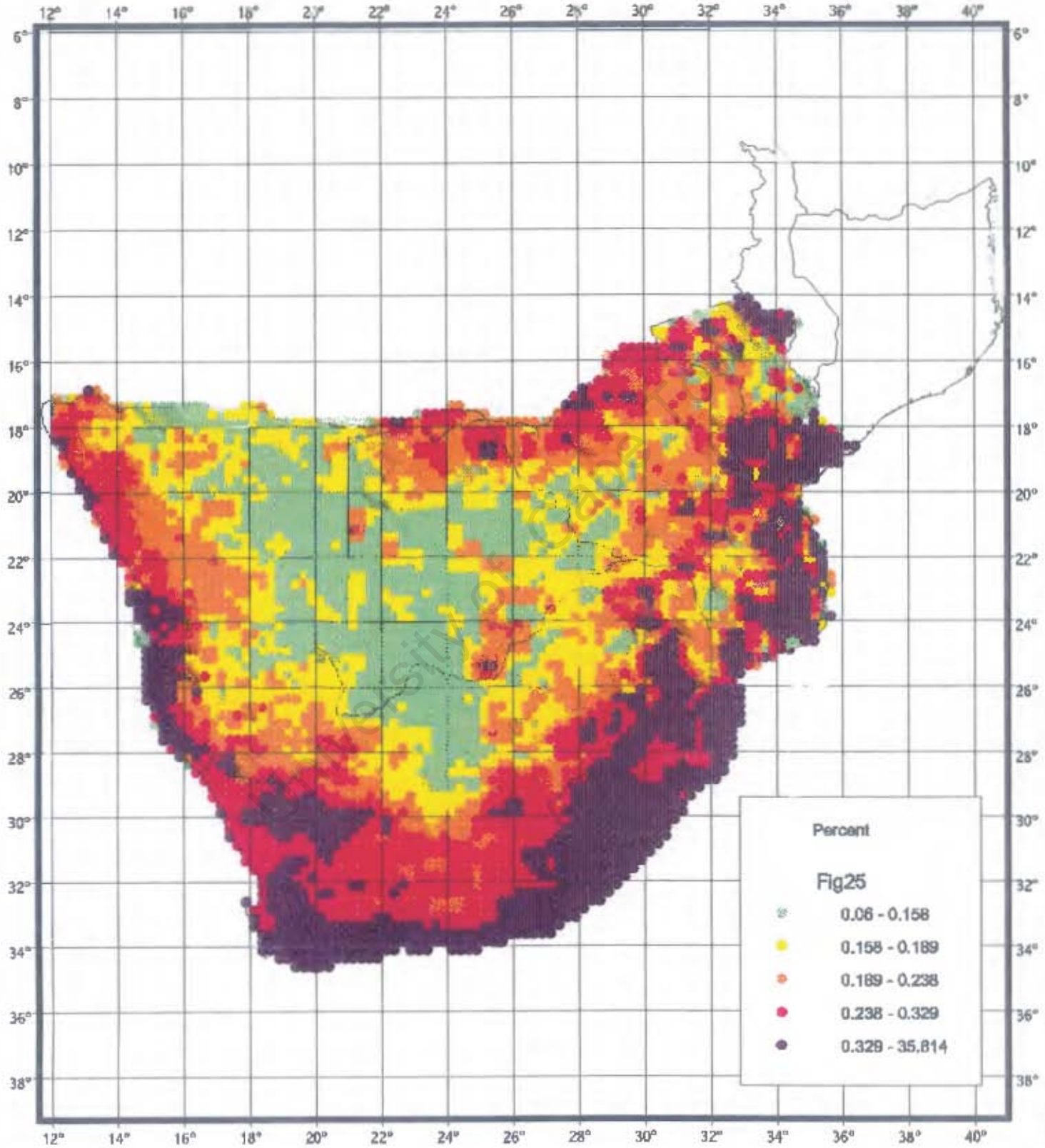


Figure 25: Ratio of narrow endemism to species richness for southern African endemic species using detection probability deciles (equation 16)



Chapter 5

The Gradients of Reporting Rate Surfaces for Bird Atlas Data in Southern Africa.

F. Little and L.G. Underhill

Avian Demography Unit, Department of Statistical Sciences, University of Cape Town, Rondebosch, 7701 South Africa

Introduction

We have concentrated on the distribution of birds, more specifically on the detection probabilities of species occurrence in southern Africa. In Chapter 2 we presented a model for generating smoothed distributions. Chapter 4 looked at combining the individual species distribution information so as to get an illustration of overall species richness or species endemism. Instead of focusing on the actual species composition in given geographical areas, it may be of interest to look at the rate of change of detection probabilities, both for single species and for groups of species. Areas where such rates of change in species composition are small will define regions of relative stability and generally homogenous species distributions. They will thus define biotic provinces (Linder 2001). On the other hand, areas of large change in species composition will define zonal boundaries. Such areas or borders may contain information regarding habitat, topography, vegetation and the effect of human intervention on the overall distributions of species.

Identifying areas of relative stability in species composition, as opposed to high species turnover and change, is also important from a conservation point of view. It is vital for the long-term effectiveness of conservation areas, that these areas be placed where conditions remain relatively constant and species populations are sustainable (Rodrigues *et al.* 2000b). A currently popular method for reserve selection is based on complementary set analysis that looks for the minimum area to represent the maximum number of species at least once (Rodrigues *et al.* 2000a). However, there is concern that this approach tends to select areas "in the transition between ecological regions" because these transition zones appear to be species rich as a result of habitat diversity (Gaston *et al.* 2001; Harrison & Martinez 1995).

Measuring the rate of species turnover in different geographical areas is thus important. Williams (1999) derived three indices to measure overall spatial turnover, gradients in species richness and replacements among species. These measures are all based on analysing the structure of spatial turnover among species within blocks of nine grid cells. Overall spatial turnover is based on the number of differences between species presence or absence for the central grid cell and the neighbouring grid cells. Gradients in species richness focus on consistent trends in species richness along opposite directions within the blocks of nine grid cells. Species replacement is based on Rapoport's (1982) index of neighbourhood segregation which counts the number of unique pairwise species where both species occur in the neighbourhood but which do not share a single grid cell within the neighbourhood. Particular disadvantages of these indices are that they do not work well for highly fragmented species distributions and they are based on presence-absence data which makes them insensitive to changes in species abundance. De Klerk (1998), in her analysis of the biogeography of terrestrial Afrotropical birds, measured neigbourhood segregation using Rapoport's index to indicate the degree of species replacement. She determined species richness gradients by measuring neigbourhood heterogeneity, as the sum of squares of deviations from the mean richness within a neigbourhood. Gaston

et al. (2001) calculated a beta-diversity measure that looks at patterns of species transition along different directional gradients and chooses the maximum value of these directional summaries. Once again these patterns are based on presence-absence data.

With presence-absence data, it is impossible to define the concept of a gradient for any single species; the gradient is effectively zero for both the area of presence and the area of absence, with a discontinuity along the edge of the range. The existence of maps showing reporting rates opens up the possibility of discussing the concepts of gradients within the range of a single species. Intuitively, the gradient for a species is large when adjacent grid cells have divergent reporting rates. Our use of a mathematical model to generate smoothed detection probabilities for species creates the opportunity to calculate rates of change among these probabilities for a species directly from the model. In this chapter we explore this further. We first introduce the general mathematical theory of directional derivatives and gradients on which our approach is based. Then we apply this theory to our smoothing models. We apply it first to single species, and then combine gradients for groups of species.

We test our method on a subset of 12 bird species that were all demonstrated to have an edge at 26°S (Harrison *et al.* 1997a, b). The discontinuity in the distributions of many southern African species at 26°S is due to the fact that the woodland biome and the grassland biome abut along this line of latitude (see Allan *et al.* 1997). Allan *et al.* (1997) stated, "Numerous species of both biomes have remarkably linear boundaries at this latitude." We use the results for one of these species to explore the effect of different models on the gradient distributions and to validate the gradient calculations. We then illustrate the smoothed distributions of detection probabilities and the associated gradient distributions to show how the gradients identify areas of change within the species distributions. We devise a method for cumulating the gradients for these 12 species and we map these gradient "summaries". Finally we extend the cumulation to all terrestrial species in southern Africa to identify areas of relatively small and large changes in species composition within the study area.

Material and Methods

Theory of Directional Derivatives and Gradients

To measure the change in species distributions within each grid square, we make use of the mathematical models used to determine the predicted response rates and their directional derivatives and gradients. The smoothing models base the prediction on data from grid cells surrounding the target grid cell for which the prediction is being made. In Chapter 2, where full details are given, we introduced the following model for a prediction based on nine cells,

$$\ln \frac{\pi_i}{1-\pi_i} = f(x,y) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy, \tag{1}$$

where π_i = the predicted probability in grid cell i and the x and y values are north-south and east-west co-ordinates relative to the target grid cell, which has $x = 0, y = 0$. When we based our predictions on 25 cells, we had the choice of using the above four-parameter model, or of adding two further terms to the model, viz.,

$$\ln \frac{\pi_i}{1-\pi_i} = f(x,y) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy + \beta_4 x^2 + \beta_5 y^2. \tag{2}$$

Both equations can be interpreted as defining a surface in (π, x, y) space. In Chapter 2 we fitted equations (1) and (2) to estimate the reporting rate at the target grid cell of a block of nine or 25 grid cells. Each

target grid cell would have had a different neighbourhood of nine or 25 grid cells and thus each estimated reporting rate would have been generated by a unique model. Each of these models generated a fitted surface for the blocks of nine or 25 grid cells. In this chapter we consider the gradient of the fitted surface at the target grid cell.

We review the mathematical theory for deriving directional derivatives as described in Thomas & Finney (1979) and in Lang (1976). For any two points, $P_0 = (x_0, y_0)$ and $P_1 = (x_1, y_1)$, and a bivariate function $\pi = f(x, y)$ in the xy-plane, the change in π in going from π_0 at P_0 to π_1 at P_1 is

$$\Delta\pi = \pi_1 - \pi_0 = f(x_1, y_1) - f(x_0, y_0). \tag{3}$$

We can fix P_0 and let P_1 approach it along some specific smooth curve in the xy-plane. For example, the approach can be along a straight line making an angle θ with the x-axis. We can measure the change in π as the distance $\Delta l = P_1 - P_0$ between the two points approaches zero, as

$$\frac{d\pi}{dl} = \lim_{\Delta l \to 0} \frac{\Delta\pi}{\Delta l} = \lim_{\Delta l \to 0} \frac{f(x_1, y_1) - f(x_0, y_0)}{\sqrt{\Delta x^2 + \Delta y^2}}, \tag{4}$$

where $\Delta x = x_1 - x_0$ and $\Delta y = y_1 - y_0$. This measures the instantaneous rate of change in π at P_0 in the direction of P_1 and is called the *directional derivative* of $\pi = f(x, y)$ at (x_0, y_0) in the direction of P_1.

We can fix $y = y_1 = y_0$ and let P_1 lie on the line $y = y_0$, parallel to the x-axis, in which case the directional derivative simplifies to the partial derivative of $\pi = f(x, y)$ with respect to x at $P_0 = (x_0, y_0)$,

$$f_x(x_0, y_0) = \left(\frac{\delta\pi}{\delta x}\right)_{x_0, y_0}. \tag{5}$$

This will measure the rate of change at P_0 in the direction of the x-axis. Similarly, we can fix $x = x_0$ and let P_1 lie on this line so that $\Delta x = 0$ and the directional derivative simplifies to the partial derivative of $\pi = f(x, y)$ with respect to y at $P_0 = (x_0, y_0)$,

$$f_y(x_0, y_0) = \left(\frac{\delta\pi}{\delta y}\right)_{x_0, y_0}, \tag{6}$$

the rate of change at P_0 in the direction of the y-axis. The directional derivative of π at P_0 in the more general direction that does not restrict P_1 to lie on lines parallel to either the x or y axes, can be calculated from the partial derivatives of $f(x, y)$ using the relationship

$$\frac{d\pi}{dl} = f_x(x_0, y_0)\cos\theta + f_y(x_0, y_0)\sin\theta, \tag{7}$$

where θ is the angle between the line $\overrightarrow{P_0P_1}$ and the x-axis.

If i is a unit vector from $(0, 0)$ to $(1, 0)$ and j is a unit vector from $(0, 0)$ to $(0, 1)$, then the line from P_0 to P_1 can be defined in terms of these unit vectors as

$$\overrightarrow{P_0P_1} = i\Delta x + j\Delta y. \tag{8}$$

The length of this line is given by $\Delta l = \sqrt{\Delta x^2 + \Delta y^2}$. The direction is given by

$$u = \frac{\overrightarrow{P_0P_1}}{\Delta l} = i\frac{\Delta x}{\Delta l} + j\frac{\Delta y}{\Delta l} = i\cos\alpha + j\cos\beta \tag{9}$$

where α is the angle between the tangent line $f_x(x_0, y_0)$ and the x-axis and β is the angle between the tangent line $f_y(x_0, y_0)$ and the y-axis. If $f(x, y)$, $f_x(x, y)$ and $f_y(x, y)$ are all continuous functions of x and y in some neighbourhood of the point $P_0(x_0, y_0)$, we find that the directional derivative of $\pi = f(x, y)$ at P_0 in the direction of $u = i\cos\alpha + j\cos\beta$ is

$$\frac{d\pi}{dl} = f_x(x_0, y_0)\cos\alpha + f_y(x_0, y_0)\cos\beta. \tag{10}$$

This can be expressed as the dot product of the vector u in equation (9) and the vector

$$v = if_x(x_0, y_0) + jf_y(x_0, y_0). \tag{11}$$

That is

$$\frac{d\pi}{dl} = u \cdot v. \tag{12}$$

Equation (12) separates the directional derivative into two parts: a part u which depends upon the direction and a part v which depends upon the function defining the surface at the point P_0. The vector v is called the *gradient* of f at P_0.

We denote the gradient by

$$\text{grad}\,\pi = \nabla\pi = i\frac{\delta\pi}{\delta x} + j\frac{\delta\pi}{\delta y}, \tag{13}$$

and express the directional derivative in terms of the gradient as

$$\begin{aligned}\left(\frac{d\pi}{dl}\right)_{P_0} &= (\nabla\pi)_{P_0} \cdot u \\ &= \mid (\nabla\pi)_{P_0} \mid\mid u \mid \cos\theta \\ &= \mid (\nabla\pi)_{P_0} \mid \cos\theta \end{aligned} \tag{14}$$

where θ is the angle between the vector $(\nabla\pi)_{P_0}$ and the unit vector u. From equation (14) we see that $\left(\frac{d\pi}{dl}\right)_{P_0}$ will be largest when $\cos\theta = 1$, that is when u and $\nabla\pi$ have the same direction. Thus we can say that the function $\pi = f(x, y)$ changes most rapidly in the direction given by the vector grad π. Moreover, the directional derivative in this direction is equal to the magnitude of the gradient. Thus to find the maximal rate of change in species density at a given point, we need to calculate the *magnitude* of the gradient of the function used for predicting the smoothed detection probabilities at that point.

Derivation of the Gradient for the Logit Models used to Calculate the Smoothed Detection Probabilities

From the logistic regression model (1), we obtain the following expression for the smoothed detection probability,

$$\pi = \frac{\exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy)}{1 + \exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy)}, \tag{15}$$

where the estimation of the coefficients is described in Chapter 2. We wish to determine the gradient of π at a given point P, defined as

$$\text{grad}\,\pi(P) = \left(\frac{\delta\pi}{\delta x}, \frac{\delta\pi}{\delta y}\right)_P. \tag{16}$$

The partial derivatives of π in (15) with respect to x and y are

$$\frac{\delta\pi}{\delta x} = \frac{(\beta_1 + \beta_3 y)\exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy)}{(1 + \exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy))^2} \tag{17}$$

and

$$\frac{\delta\pi}{\delta y} = \frac{(\beta_2 + \beta_3 x)\exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy)}{(1 + \exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy))^2} \tag{18}$$

We wish to calculate the gradient for the central grid square with coordinates $(0, 0)$. By substituting $P_0 = (0, 0)$ in equations (17) and (18) we get

$$\text{grad}\,\pi(0,0) = \left(\frac{\beta_1 \exp\beta_0}{(1+\exp\beta_0)^2}, \frac{\beta_2 \exp\beta_0}{(1+\exp\beta_0)^2}\right) \tag{19}$$

We are specifically interested in the magnitude of the gradient which is given by

$$\sqrt{\left(\frac{\beta_1 \exp\beta_0}{(1+\exp\beta_0)^2}\right)^2 + \left(\frac{\beta_2 \exp\beta_0}{(1+\exp\beta_0)^2}\right)^2}. \tag{20}$$

For the model (2) that includes the two extra terms, x^2 and y^2, the magnitude of the gradient simplifies to the same equation at the point $P_0 = (0, 0)$. The estimates of β_0, β_1 and β_2 will of course be different from those obtained for the four-parameter model (1).

The Direction of the Gradient

The magnitude of the gradient provides a measure of the degree of change of reporting rates in a given area. The direction of the gradient vector gives us an indication of the direction of this change. The direction of the maximal change relative to the x-axis, i.e., relative to the west-east direction, can be obtained from the partial derivatives as

$$\phi = \arccos\left(\frac{\frac{\delta\pi}{\delta x}}{\sqrt{\frac{\delta\pi}{\delta x}^2 + \frac{\delta\pi}{\delta y}^2}}\right) \tag{21}$$

Since the arccos function gives the same value for the two halves of the 360° rotation, we have to use the following adjustment:

$$\text{If } \tfrac{\delta\pi}{\delta y} < 0 \text{ then } \phi = 360° - \phi.$$

These mathematics generate 0° in an easterly direction and rotates anti-clockwise so that 90° is north, 180° is west and 270° is south.

Combining the Gradients for More than One Species

Where the gradients for individual species are steep, we identify the effective edges of the distributions for the species. Of more interest is the identification of areas where many species are at the effective edges of their ranges. To determine such areas, we need to combine the gradient information for a number of species. We form two different summations of the individual species gradients. Firstly, we sum the magnitudes of the gradients, while ignoring any directional information,

$$G_j = \sum_{i=1}^{n} g_{ij}, \tag{22}$$

where g_{ij} = the gradient for species i in grid cell j. The resulting gradient sum will be large if grid cell j has many large species gradients.

Secondly, we view the gradients as vectors and thus we wish to take into consideration their associated directions in the summation. We cannot use vector summation because we do not wish gradients in opposite directions to cancel each other. We overcome this problem by summing the magnitudes of the species gradients in each of 16 directions for each grid cell. We determine these 16 discrete directions by dividing the 360° rotation by 22.5° to give the set of values,

$$d_k \in [0, 22.5, 45, 67.5, 90, 112.5, 135, 157.5, 180, 202.5, 225, 247.5, 270, 292.5, 315, 337.5].$$

We replace each gradient direction, ϕ_{ij}, by the discrete direction just smaller than its actual value, viz.,

$$d_{ij} = (\text{int}(\phi_{ij}/22.5)) * 22.5. \tag{23}$$

Thus directional gradients from 0° to 22.5° get transformed to 0°, from 22.5° to 45°, to 45°, and so on. We then sum the gradients in each of the 16 directions for each grid cell,

$$D_{jk} = \sum_{i=1}^{n} g_{ij} * s_i, \tag{24}$$

where $s_i = 1$ if $d_{ij} = d_k$, and $s_i = 0$ otherwise, for $k = 1$ to 16. Thus each grid cell will have 16 summed gradient totals corresponding to the 16 directions, 22.5° apart.

The magnitude of the above summations of gradients will depend both on the magnitude of the individual species gradients and on the number of species detected in a given grid cell. Grid cells rich in species will tend to have larger gradient sums than grid cells with fewer species. To remove from the gradient sums this possibly confounding effect of species richness , we calculate the median gradient for each grid cell. For each grid cell, 50% of the individual species gradients will be larger than this median value. It thus gives a measure of the overall magnitude of individual species detection probability gradients in a given area.

Mapping the Gradients

For individual species, we can calculate a rate and direction of maximal change for each grid cell in southern Africa. When mapping these, we wish to illustrate both the degree of change, as well as the direction of change. We use shaded arrows for mapping purposes. We divide the range of gradient magnitudes into four quartiles and use four different shades to illustrate these quartiles. The distribution of the gradients is very skewed with long tails to the right, hence the top quartile has a relatively wide range compared to the other three quartiles. We rotate the arrow head to point in the direction of maximal change, where the directions have been rounded to one of the 16 directions, each 22.5° apart, described above.

Maps of the directional gradient sums can potentially have a "star" with differently shaded arrows pointing in 16 directions for each grid cell. In practice, this requires large scale maps for a clear presentation. To overcome this problem of scale, we choose to display only the gradients that fall in the top percentiles of the gradient distribution. The areas that contain these large changes will be the areas where the turnover in species is large. The directions in which these large changes occur will in addition provide insight into overall patterns of change in species composition. To map the gradient sums that are independent of the directions, we divide the total range of these sums into five quintiles and use shaded dots to illustrate the relative magnitudes.

Results

Gradient Distribution for the Blue Waxbill Generated by the Three Different Models

Figs 1 to 3 show the smoothed distributions of detection probabilities for the Blue Waxbill *Uraeginthus angolensis* based on nine grid cells and based on 25 grid cells using both the four- and six-parameter models. The Blue Waxbill occurs in northern Namibia, northern Botswana, throughout Zimbabwe and south and central Mozambique. In South Africa the distribution follows a horseshoe shape along the north-eastern border. Detection probabilities decrease sharply to zero just north of 26°S between 28°E and 31°E. Areas with large detection probabilities include northern Zimbabwe and central Mozambique, southern Mozambique and along the border between Botswana and the Limpopo Province of South Africa. Isolated areas with small detection probabilities are scattered in central Namibia and southern Botswana. The above distributional features are apparent from all three versions of the smoothed maps (Figs 1-3). However, the map based on blocks of nine grid cells (Fig. 1) and the map based on blocks of 25 grid cells using the six-parameter model (Fig. 3) are both more patchy within the overall range of the species. The six-parameter model has also expanded and magnified the isolated areas of low detection probabilities in central Namibia and southern Botswana.

The gradient distributions generated by the three models (Figs 4-6) illustrate the major edges of the distributions of detection probabilities. However, the gradient distribution based on the nine grid cell model (Fig. 4) appears too sensitive to small changes between neighbouring grid cells and thus masks the more important areas of substantial changes. By contrast, the gradient distribution that uses the same four-parameter model on blocks of 25 grid cells (Fig. 5) is simpler and clearly highlights the areas where the detection probabilities of the Blue Waxbill show large changes. It illustrates the horseshoe shape of the edge along 28°E, 26°S and 31°E. It highlights the areas with relatively large detection probabilities compared to the detection probabilities in their surrounding areas. The gradients are conspicuous along the edges of the area with large detection probabilities. This is clear for example, in northern Namibia, in the area with large detection probabilities in northern Botswana and in the area with large detection probabilities in southern Mozambique that extends southwards in a narrow band east of the Great Escarpment.

Because of the way that the maps of smoothed detection probabilities are constructed, with equal numbers of grid cells of each shade, the cutpoints forming the boundaries between shades are not equidistant. For example, for the Blue Waxbill (Fig. 2), the green shade has width 6.5%, the yellow shade 29.8%, the orange shade 26.5%, the red shade 12.0% and the purple shade 25.2% (see legend in Figure 2). Most of the area shaded purple will have detection probabilities in the lower end of the range for this shade. This implies that steep gradients are unlikely to be found in areas with green, red or purple shades. The gradient map (Fig. 5) reflects this because the gradients in Zimbabwe fall in the second quartile of the gradient range showing the small difference among the detection probabilities for the areas shaded in purple and red in Fig. 2. However, we have dark gradients, falling in the top quartile of the gradient range, in the grid cells centred on 27°S, 27°E, for example. The gradients point towards the relatively larger detection probabilities for these cells compared to the surrounding area even though the detection probabilities for this area are shaded yellow and orange. This illustrates that the gradients measure the change in detection probabilities, irrespective of the magnitude of the individual probabilities.

The directions in which the increases in the reporting rates (as estimated by the detection probabilities) occur are represented by the directions in which the arrows point in Fig. 5. Areas of large change are marked

by arrows pointing consistently in the same direction, for example the northwards increase in detection probabilities along latitude 26°S. By comparison, the directions of the smaller gradients appear to be more random.

Fig. 6 illustrates the gradient distribution generated by the six-parameter model based on blocks of 25 grid cells. It illustrates the same major edges as discussed for the previous map. However, it seems to magnify changes in areas in which the species occurs at scattered grid cells. For example, we notice large gradients surrounding a few grid cells in central Namibia and southern Botswana. This corresponds to the magnification of the detection probabilities for these cells (Fig. 3) when using the six-parameter model. We therefore have reservations about the usefulness of this model, and do not consider it further.

Validating the Gradient Calculations

To validate the gradient calculations and to further examine the differences between the three models, we choose four grid cells at which we look in detail at the estimated surfaces and the magnitude of the gradients in relation to the slopes of these surfaces. Figs 7 to 10 illustrate these surfaces for the four grid cells. The chosen grid cells are 1531AC in central Mozambique, which has a relatively large gradient for the Blue Waxbill; 2230CD that has a relatively small gradient; 2319DA, one of the isolated cells in central Namibia for which the two models based on blocks of 25 grid cells generate very different gradients; and 2726CB in an area with relatively low detection probabilities. Comparing calculations for grid cells 1531AC and 2230CD will validate large versus small gradients. Examining the calculations for grid cell 2319DA will facilitate a comparison of the two models for the blocks of 25 grid cells. Comparing grid cell 2726CB to grid cell 1531AC will investigate the independence of the gradients from the overall level of the range of detection probabilities for the surrounding neighbourhood.

The atlas coverage reported for the neighbourhood surrounding grid cell 1531AC in Fig. 7 shows that there were no checklists for this target cell or for some of its neighbouring grid cells and very low coverage for all but two of the other grid cells in the blocks of nine or 25 grid cells. This sparse coverage generated many zero observed reporting rates, especially in the south-eastern corner of the nine grid cell block, while most of the reporting rates in the north-west were larger than 0.89. When adding the grid cells to enlarge the block to 25 grid cells, we add some large observed rates in the south. The surfaces generated by the three models are substantially different from one another. The slope of the surface based on the model for the block of nine grid cells is steep and increases towards the north-west. This corresponds to the large gradient of 1.292 in a north-westerly direction. The four-parameter model based on 25 grid squares generates a much flatter surface because it takes into account the larger observed reporting rates to the north and south of the row of zero rates that were a result of poor coverage. This model is more successful in estimating reporting rates for the empty cells based on the information in the surrounding neighbourhood. The flatter surface corresponds to the smaller gradient of 0.1316. The six-parameter model adds too much curvature to the model because of the extra weight of the grid cells at the extremes of the five-by-five block, through the terms x^2 and y^2 in the model. As a result the gradient for this surface is larger at 0.2802.

Fig. 8 illustrates much flatter surfaces surrounding grid cell 2230CD compared to the results for grid cell 1531AC in Fig. 7. The correspondingly smaller gradient values for all three models illustrate how the relative magnitude of the gradients accurately reflect the differential slopes of the reporting rate surfaces. Fig. 9 shows that grid cell 2319DA is one of two grid cells with non zero observed rates in its neighbourhood of nine or 25 grid cells. Since both these non zero rates fall within the nine grid cell block, the surface generated

for this block is not quite as flat as the surface for the 25 grid cell block using the four-parameter model. Once again the six-parameter model adds a lot of curvature to the surface in its attempt to capture both the two non zero reporting rates at its centre and the zero rates surrounding it. Hence the relatively large gradient of 0.1331 for this model compared to the gradients of 0.0457 and 0.0019 for the surfaces generated by the four-parameter model for the blocks of nine and 25 grid cells, respectively. We feel that the four-parameter model is more successful in sharing out the non zero observed rates among the grid cells in the neighbourhood.

Finally, Fig. 10 illustrates the surfaces around grid cell 2726CB. When we compare the surface generated by the four-parameter model for the block of 25 grid cells to the corresponding surface for grid cell 1531AC (Fig. 7), we notice how these surfaces have similar slopes, though they lie at different levels with respect to the probability axis. This corresponds to the gradients of 0.1316 for grid cell 1531AC and 0.1174 for grid cell 2726CB that are not too different in magnitude and illustrates the independence between gradient magnitudes and the overall level of the range of the detection probabilities for the grid cells in the neighbourhood. The directions for the gradients of all the surfaces illustrated in Figs 7 to 10 clearly indicate the directions in which the slopes of these surfaces increase most steeply.

In conclusion, the model based on nine grid cells is too sensitive to small changes in detection probabilities and the six-parameter model gives too much weight to isolated large observed rates. Our chosen model for the gradients analysis that will highlight edges and areas of change most clearly is thus the four-parameter model applied to blocks of 25 grid cells. We do not consider the other two models further in our analysis of species gradients.

Individual Species Gradients

We chose a subset of 12 species that were shown to have an edge at the 26°S line to show how gradients identify areas of change within species distributions and to evaluate our proposed methods of combining the gradient information for several species. These 12 species are listed in Table 1. We have already discussed the smoothed probability distribution (Fig. 2) and the corresponding gradient distribution (Fig. 5) for the Blue Waxbill. Fig. 11 illustrates the smoothed probability distribution for the Longtailed Widow *Euplectes progne*. It has a core of large detection probabilities which decrease sharply towards the north at latitude 26°S and which decrease more gradually towards the southern and western edges of its range. Fig. 12 shows the gradient distribution generated by the four-parameter model based on data from 25 neighbouring cells. This map illustrates the edges of the distribution. In addition, it highlights the change in relative magnitude of detection probabilities south of latitude 28° along the northern border of Lesotho.

Figs 13 and 14 show the smoothed distribution of detection probabilities and the distribution of gradients for the Rattling Cisticola *Cisticola chiniana*, respectively. The overall distribution of the Rattling Cisticola (Fig. 13) is similar to that of the Blue Waxbill (Fig. 2) but the Rattling Cisticola has a more fragmented distribution which has a complex pattern of patches with large detection probabilities, rather than simply a single "core" area. The increase in detection probabilities north of 26°S is also more gradual than is the case for the Blue Waxbill. The gradient distribution (Fig. 14) illustrates the changes in detection probabilities within the overall range. The edge immediately to the north of 26°S is again highlighted, and it indicates the more gradual change in rates for this species than for the Blue Waxbill.

Fig. 15 illustrates the smoothed distribution of the Cloud Cisticola *Cisticola textrix*. For this species, detection probabilities are small overall, as is illustrated by the shorter range in the legend; the darkest shade

commences at a detection probability of 10.3%. A change from red to purple on the distribution map for the Cloud Cisticola reflects variations in detection probabilities between 4.9% and 27.5%, an interval of 22.6%, while the two top quintiles for the Blue Waxbill covered a range of detection probabilities between 62.8% and 100%. All of the bottom quintile and most of the second quintile of detection probabilities for the Cloud Cisticola are less than 2% and are thus illustrated by a cross. The gradient map that divides the gradient range of the Cloud Cisticola into four species-specific quartiles (Fig. 16) again highlights the areas of relatively large changes in detection probabilities. However, it may give the false impression of large changes. When we use a different scale to subdivide the gradient range that divides the combined gradient range for all 12 selected species into four quartiles (Fig. 17), we see that the changes among the detection probabilities for the Cloud Cisticola are relatively small compared to the changes for the other species.

Figs 18 to 33 show the detection probability distribution and gradient distribution maps for the other eight species in the selected subset (see Table 1). The gradient maps highlight individual borders and areas of change apparent from the individual species probability maps. Apart from the unique features of each species, they all share a common edge or border at 26°S. The distributions of several of these species approach 26°S from the north, and extend southwards in a narrow strip along the east coast of South Africa: Arrowmarked Babbler *Turdoides jardineii* (Fig. 18), Southern Black Tit *Parus niger* (Fig. 20), Blackheaded Oriole *Oriolus larvatus* (Fig. 22) and Lesser Striped Swallow *Hirundo abyssinica* (Fig. 26). Other species share a distribution south of 26°S: South African Cliff Swallow *Hirundo spilodera* (Fig. 24), Spikeheeled Lark *Chersomanes albofasciata* (Fig. 28), Pied Starling *Spreo bicolor* (Fig. 30) and Bokmakierie *Telophorus zeylonus* (Fig. 32).

In addition to the 12 species discussed above, we looked at the gradient distribution for the Bateleur Eagle *Terathopius ecaudatus*. This species is known to occur in some of the larger conservation parks but has virtually disappeared from areas outside these parks. Figs 33a and b confirm this pattern of distribution and the large gradients (illustrated by the red arrows in Fig. 33b) demarcate many of the larger game parks within the species distribution range.

Summing the Gradients for a Subset of 12 Species

The distributions of individual species gradients clearly identify the edges of the species distributions and highlight changes in detection probabilities within the overall distribution ranges. To a large extent this information can be obtained by a careful scrutiny of the detection probability distribution maps. Of more interest is combining the gradient information for groups of species to detect areas that mark both the changes in species richness, as well as changes in species composition. To this end we sum the gradients for all species in a given grid cell using equation (22). Fig. 34 shows the results of this summation. The range of the gradient sums has been divided into five quintiles for plotting purposes. The area immediately north and south of 26°S is shaded purple, identifying the area with the largest gradient sums. The 12 selected species had an edge of their distributions at the 26°S. In addition many had a border inland from the east coast of South Africa and others had a west-north-east curved edge at the northern limit of their distribution south of 26°S. These areas are thus correctly identified as the areas with steep gradients reflecting coincident changes in species composition for several of the 12 species. In addition, Fig. 34 shows large gradient sums around the southern edge of the Okavango delta and in Mozambique. Several species, including Blue Waxbill (Fig. 2), Rattling Cisticola (Fig. 13) and Southern Black Tit (Fig. 20) have two distinct cores of large detection probabilities in central and southern Mozambique.

We also sum the gradients for each grid cell taking into consideration their direction using equation (24). Each grid cell can thus potentially have a gradient sum in each of the sixteen directions, resulting in a "star" effect. Grid cells that identify ecological boundaries will have species moving in single directions and this will be illustrated by a graphical image with arrows in possibly only two opposite directions. However, since each grid cell now has to accommodate up to 16 arrows, the scale of the maps becomes a problem and lacks clarity. Because we are particularly interested in the direction of the gradient in areas of major changes, i.e., in areas with large gradient sums, we can reduce the problem of scale and improve the visual image of the directional maps by plotting only the gradient sums that fall in the upper sub-range of the distribution. Fig. 35 illustrates the distribution of the top 10% of the directional gradient sums for our 12 selected species by using arrows that point in the direction of these maximal changes. It clearly shows the north-south turnover of species along 26°S latitude and the westward turnover of species away from the east coast.

Fig. 36 illustrates the distribution of the median gradients for the 12 selected species. The large gradients for the species to the north or south of 26°S are highlighted. The common edge of many of the species along the Drakensberg escarpment north of Lesotho and south of Swaziland is illustrated. Many of the 12 species with distributions to the north of 26°S have the edges of their distributions stretching from northern Namibia, eastwards along the southern edge of the Okavango delta, northern Zimbabwe and the northern part of central Mozambique, and extending southwards along the east coasts of Mozambique and KwaZulu-Natal. These edges correspond to large median gradients in these areas. The Bokmakierie (Fig. 32), Pied Starling (Fig. 30) and Spikeheeled Lark (Fig. 28) all have major changes in detection probabilities within their ranges along 30°S in the Northern Cape. The Rattling Cisticola (Fig. 13) and Spikeheeled Lark (Fig. 28) have relatively large detection probabilities in the Kalahari Game Reserve compared with surrounding areas which give rise to the large median gradients in this area. All of these areas with large median gradients are visible in the distribution of the sum of the gradients (Fig. 34) but are not as clearly highlighted unless the edges are shared by most of the 12 species. The areas with high median gradients also correspond to the distribution of the top 10% of directional gradient sums (Fig. 35).

Summing the Gradients for all Terrestrial Species

Fig. 37 shows the distribution of the gradient sums for all terrestrial species. The range has again been divided into five quintiles for plotting purposes. The areas of maximal change in southern Africa include the area north of 26°S, the east coast of South Africa, the Escarpment, the Limpopo Province, eastern Zimbabwe, the Zambezi valley in Mozambique and the southern border of the Okavango delta in northern Botswana. Zimbabwe and the South African region north of latitude 26°S have large gradient sums falling into the top 40% of the distribution. The area shaded in green and to a lesser extent the adjoining yellow areas, in Fig. 37 represent the areas where the gradient sums are very small. These areas cover the Kalahari and parts of Namibia, the Northern Cape and the Karoo. However, this distribution is confounded by species richness in that the magnitude of the gradient sums depends not only on the magnitude of the individual species gradients but also on the number of species that can contribute individual gradients to this sum. When we compare the species richness distribution for all terrestrial species (Fig. 1 of Chapter 4) to the distribution of the sum of the gradient deciles (Fig. 37), we notice how areas rich in species will also tend to have large gradient sums. This is true for northern Botswana, Zimbabwe, and the Northern and Limpopo provinces of South Africa, extending southwards along the Escarpment to KwaZulu-Natal. By contrast, central and southern Botswana is characterised by low relative species richness, except for some patches where species have large detection probabilities (Fig. 1, Chapter 4). The sparse distribution of species in this area is very clearly illustrated in Fig. 2 of Chapter 4 that shows a simple count of the number of species present. Grid

cells in this area had between 1 and 82 species with non zero detection probabilities, compared with between 174 and 291 species per grid cell in the darkly shaded areas. The distribution of the sum of the gradients (Fig. 37) does illustrate the gradients towards the areas with high detection probabilities in central and south-western Botswana, but the gradient sums are not as large as in the areas with many species.

Fig. 38 illustrates the top 10% of the directional gradient sums. Each grid cell can potentially have 16 arrows representing sums in 16 different directions. If the directional sums for a given area point in one direction, they depict species richness gradients. If the directional sums in the individual grid cells are predominantly in two opposite directions, they depict areas of species replacement. Large directional sums in diverging directions can be indicators of species fragmentation or can be the result of summing many small gradients in areas with many species.

In this way we identify the north-south transition of species along 26°S, the predominantly west-east transition along the Escarpment and in KwaZulu-Natal and the north-south transition of species between forested areas along the south coast and the more arid Karoo region. Species richness gradients are along the border of the Fynbos Biome in the Western Cape, in an easterly direction across the border from Namibia to south-western Botswana, between Mozambique and the Eastern Highlands of Zimbabwe, from the west coast of northern Namibia to the highlands of the interior and along the southern border of the Okavango delta. Areas where many species have random fluctuations include large parts of Zimbabwe, especially in the south-east.

The gradient sums inevitably tend to be largest in areas which have large numbers of species, because there are then many contributions to the sum. To reduce the effect of species richness, we therefore need a concept of "average" gradient regardless of the number of species present. We achieve this by calculating the median gradient for each grid cell. Fig. 39 illustrates the resulting distribution. This map should highlight areas where species have large changes in the distribution of their detection probabilities. The entire Okavango delta area is characterised by large median gradients, as opposed to the sum of the gradients that highlights the edges of this region. Large median gradients occur along the Zambezi river that form the northern borders of Zimbabwe and central Mozambique and extend southwards to the Zambezi valley. Within Zimbabwe, areas of high median gradients probably reflect changes in detection probabilities between urban and rural areas. The edge along the Eastern Highlands between Zimbabwe and Mozambique is still highlighted although it is narrower in Fig. 39 compared to Fig. 37. The gradients along 26°S are not as large as before. This indicates that though many species do have edges here, these edges are on average not as steep as in other areas. The border between the Kalahari basin and the grasslands of the higher lying areas to the south-east is characterised by large median gradients. This border extends westwards along the Orange river to the west coast of South Africa. The Fynbos Biome in the Western Cape is clearly demarcated by large median gradients. The high median gradients south of 26°S between 22°E and 24°E probably illustrate the transition in species composition between the Succulent-Karoo and Nama-Karoo and between the Nama Karoo and mixed grasslands further to the east.

The most dominant and surprising feature of Fig. 39 is, however, the purple shading of the Kalahari region in central and south-western Botswana. This is an arid area with few species. Some of these high median gradients correspond to the areas of species transition between the southern and central Kalahari and between the central and northern Kalahari, and the species richness gradient between Namibia and Botswana (see Fig. 38). However, most of the high gradients reflect changes in detection probabilities within wider species ranges due to fragmentation. A possible cause for this fragmentation is the different species composition of

conservation areas in this region compared to rangeland where over grazing has lead to bush encroachment (see Alan *et al.* 1997). This is also an area that was very poorly and unevenly covered and so we cannot rule out the possibility that this may be the cause of the large degree of apparent fragmentation in species distributions here. The data in Botswana were collected for half degree grid cells, while the smoothing algorithm is based on blocks of 25 quarter degree grid cells. As a result, our smoothing algorithm may have been less effective in generating interpolated values for grid cells not included in the coverage. Further fieldwork is required, collecting data for Botswana on a quarter degree grid scale, to investigate the possibility that the large gradients observed in south-western Botswana are not due to a mathematical artefact. This is beyond the scope of the present analysis.

Fig. 40 maps the top 10% of median directional gradients. The median directional gradients are the values above which 50% of the gradients in each of the sixteen directions fall. As such they give a summary measure of the magnitude of the gradients in each direction. We plot the top 10% of these summary values. Once again most of the larger median gradients are in central and south-western Botswana, though the other areas of species transition highlighted in Fig. 38 are also shown. Fig. 41 enlarges the map showing the distribution of median directional gradients for Botswana. The varying directions of these large median gradients confirms that the large degree of species transition in this area is due to fragmentation.

Discussion

In this chapter we have looked at deriving measures for the rate of change in detection probability distributions. In doing this we have focussed on two separate aspects. Firstly, we derived a method for calculating gradients for individual species distributions, something that was previously not possible when using presence-absence data. Secondly, we explored ways in which we could combine these gradients for groups of species.

We have developed and presented a measure for the changes in the detection probability distributions for species that derives directly from the regression model used for generating smoothed distributions of detection probabilities. This is a further advantage and motivation for our choice of smoothing method. The gradients are determined from the surfaces fitted for individual target grid cells. These surfaces are fitted for the surrounding neighbourhood of nine or 25 grid cells. These surfaces thus tend to have more curvature than the eventual smoothed surfaces of the smoothed probability distributions. For this reason we found it better to use the larger blocks of 25 grid cells for the gradient analysis. In fact, apart from validating the gradient calculations, Figs 7 to 10 provide a good visual comparative assessment of the different models.

We have illustrated how these gradients measure and reflect the degree of change among detection probabilities and how the distribution of these gradients thus highlights effective edges for individual species probability distributions, thus enabling us to define the concept of a gradient for individual species. The effective edge of a distribution, where reporting rates decrease more rapidly, frequently lies well inside the limits of the distribution, viewed from a presence-absence perspective. Several of the species for which we have illustrated the smoothed probability distributions, have more than one core of high detection probabilities within their wider ranges. These species thus have "multimodal patterns of abundance", rather than a single core that declines gradually towards the boundaries (Brown 1984). The identification of the parameters that change across the steeper gradients for a threatened species, will lead to a better understanding of what interventions, if any, could improve the status of the species.

Secondly, we have explored several ways in which we can combine the gradient information for many species. The distributions of gradient sums highlight areas with a large degree of change among the detection probabilities for species. This can be either due to large gradients for individual species or due to the presence of many species in the area. Median gradients remove the effect of species richness and focus on areas where large changes in detection probabilities may occur regardless of the number of species present. Directional gradient summations allow us to distinguish between areas of transition due to species enrichment, species replacement or random fluctuations. Looking at the median gradients and the directional gradient sums together enables us to ascribe the random fluctuations to either species fragmentation (in areas with sparse species distributions) or to smaller fluctuations among many species in areas rich in species. There is no mathematical guarantee that either the sum or the median of the detection probability gradients are equal to the sum or median of abundance gradients. However, Figs 34 to 37 illustrate that these tools work well for a group of species since known areas of transition, for example between the Kalahari and the Okavango and again along the Cunene River valley, are highlighted in all these distribution maps.

The magnitude of the individual gradients, the gradient sums and median gradients are measures of the "strength" of the transition zones (Williams *et al.* 1999). The "breadth" is illustrated by the longitudinal or latitudinal width of the consistent directional gradients. Together, these different gradient distributions point towards the possible demarcation of biotic provinces. However, there is overlap in the different types of transition in given areas, for example in Botswana where the large median gradients are a result of species enrichment, species replacement and species fragmentation. Furthermore, areas of transition are often quite wide as in the area along the Escarpment stretching southwards towards KwaZulu-Natal (see Figs 12, 23 and 33). This points to the notion that there are no real sharp edges or borders but rather a more gradual and continuous transition between different avian biomes. Our method is, however, able to identify sharp transitional borders where these are known to exist. For example, Fig. 33b illustrates the distribution of the gradients for the Bateleur and very clearly shows the sharp edges surrounding many of the game parks within its distribution range.

Ideally, we would have liked to be able to provide biogeographers with gradients based on actual abundance (or equivalently densities, i.e., birds per ha). Unfortunately, the data to achieve this are not available. Griffioen's (2001) finding, that (observed) reporting rates and densities have a "virtually linear" relationship for reporting rates below 50%, is likely to result in a lack of interest in ever determining absolute densities. So we offer the alternative, gradients of detection probabilities. This represents a major advance on the inability, up to now, to provide any form of within distribution gradient, a consequence of presence-absence distribution maps, where the within species gradient is zero apart from a discontinuity along the arbitrarily defined edge of the range.

References

Allan DG, Harrison JA, Herremans M, Navarro RA, Underhill LG 1997. Southern African geography. In: Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) The Atlas of Southern African Birds. Vol 1: Non-passerines, lxv-ci. BirdLife South Africa, Johannesburg.

Brown JH 1984. On the relationship between abundance and distribution of species. American Naturalist 124: 255-279.

De Klerk HM 1998. Biogeography and Conservation of Terrestrial Afrotropical Birds. Unpublished Ph.D. thesis, University of Cape Town.

Gaston KJ, Rodrigues ASL, Van Rensberg BJ, Koleff P, Chown SL 2001. Complementary representation and zones of ecological transition. Ecology Letters 4: 4-9.

Griffioen P 2001. Temporal Changes in the Distributions of Bird Species in Eastern Australia. Unpublished Ph.D thesis, La Trobe University, Bundoora, Victoria, Australia.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997a. The Atlas of Southern African Birds. Vol 1: Non-passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997b. The Atlas of Southern African Birds. Vol 2: Passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Martinez P 1995. Measurement and mapping of avian diversity in southern Africa: implications for conservation planning. Ibis 117: 410-417.

Lang S 1976. A First Course in Calculus. Third Edition. Addison-Wesley, Reading Massachusetts.

Linder HP 2001. On areas of endemism, with an example from the African Restionaceae. Systematic Biology 50: 892-912.

Rapoport EH 1982. Areography, geographical strategies of species. Pergamon Press, Oxford.

Rodrigues ASL, Cerdeira JO & Gaston KJ 2000a. Flexibility, efficiency, and accountability: adapting reserve selection algorithms to more complex conservation problems. Ecography 23: 565-574.

Rodrigues ASL, Gaston KJ, Gregory RD 2000b. Using presence-absence data to establish procedures that are robust to temporal species turnover. Proceedings of the Royal Society London B 267: 897-902.

Thomas GB, Finney RL 1979. Calculus and Analytical Geometry. Fifth Edition. Addison-Wesley, Reading Massachusetts.

Williams PH, De Klerk HM, Crowe TM 1999. Interpreting biogeographical boundaries among Afrotropical birds: spatial patterns in richness gradients and species replacement. Journal of Biogeography 26: 459-474.

Table 1. The subset of twelve species with an edge to their distributions at latitude 26°.

Common Name	Scientific Name
Blue Waxbill	*Uraeginthus angolensis*
Longtailed Widow	*Euplectes progne*
Rattling Cisticola	*Cisticola chiniana*
Cloud Cisticola	*Cisticola textrix*
Arrowmarked Babbler	*Turdoides jardineii*
Southern Black Tit	*Parus niger*
Blackheaded Oriole	*Oriolus larvatus*
South African Cliff Swallow	*Hirundo spilodera*
Lesser Striped Swallow	*Hirundo abyssinica*
Spikeheeled Lark	*Chersomanes albofasciata*
Pied Starling	*Spreo bicolor*
Bokmakierie	*Telephonus zeylonus*

Figure 1: Blue Waxbill - Smoothed distribution based on four-parameter model for blocks of 9 grid cells

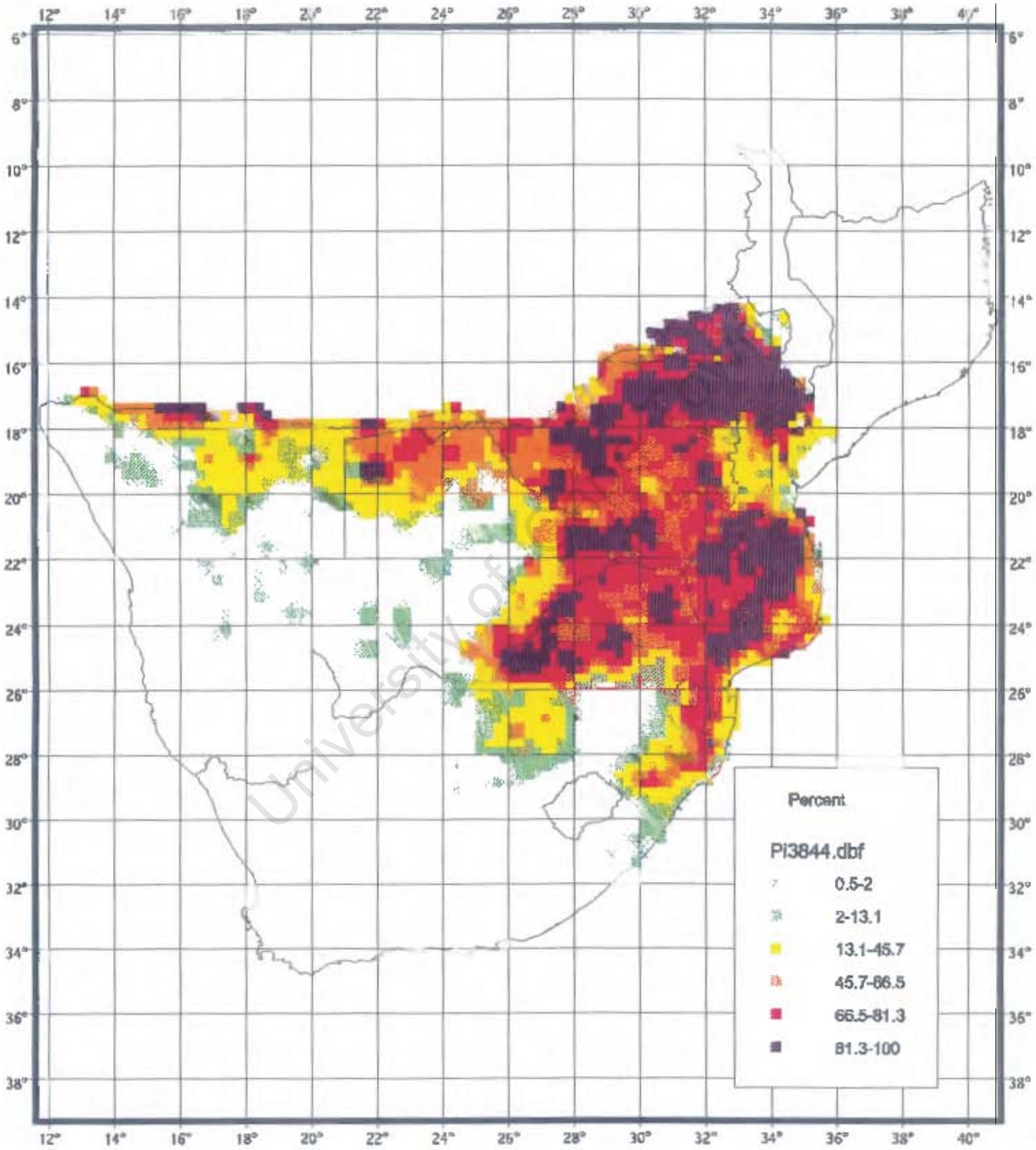


Figure 2: Blue Waxbill - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

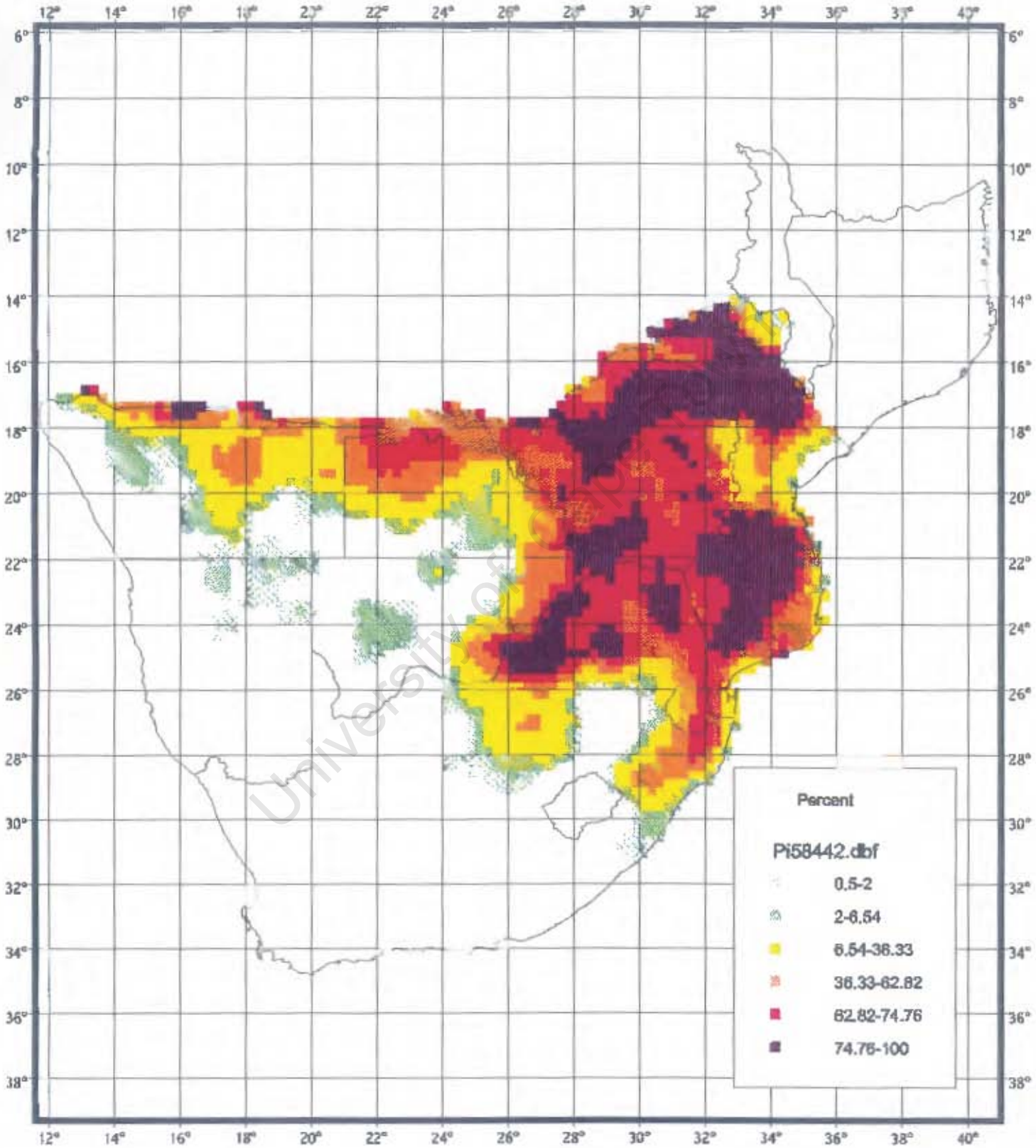


Figure 3: Blue Waxbill - Smoothed distribution based on six-parameter model for blocks of 25 grid cells

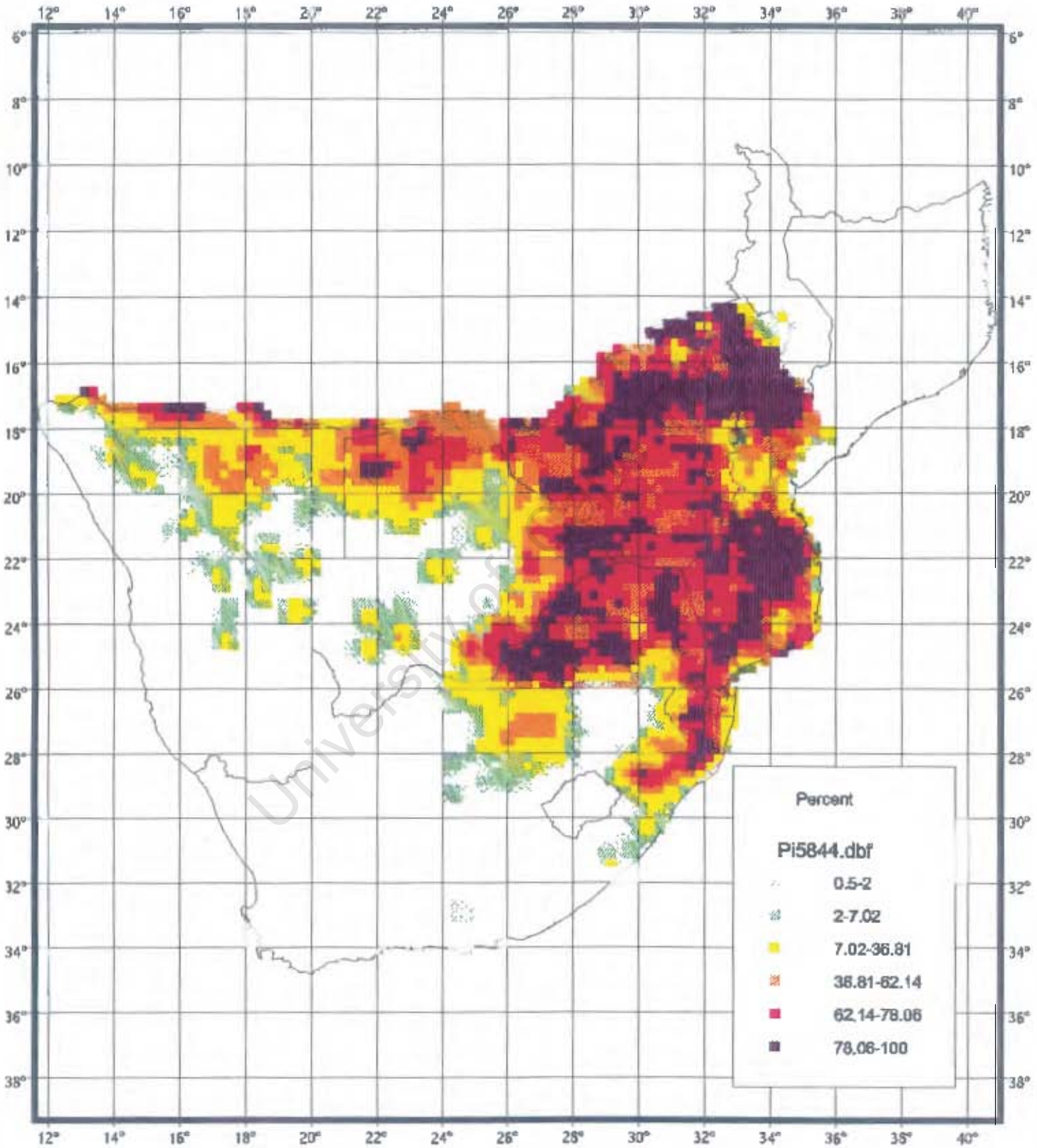


Figure 4 : Blue Waxbill - Gradient distribution based on 4-parameter model for blocks of nine grid cells

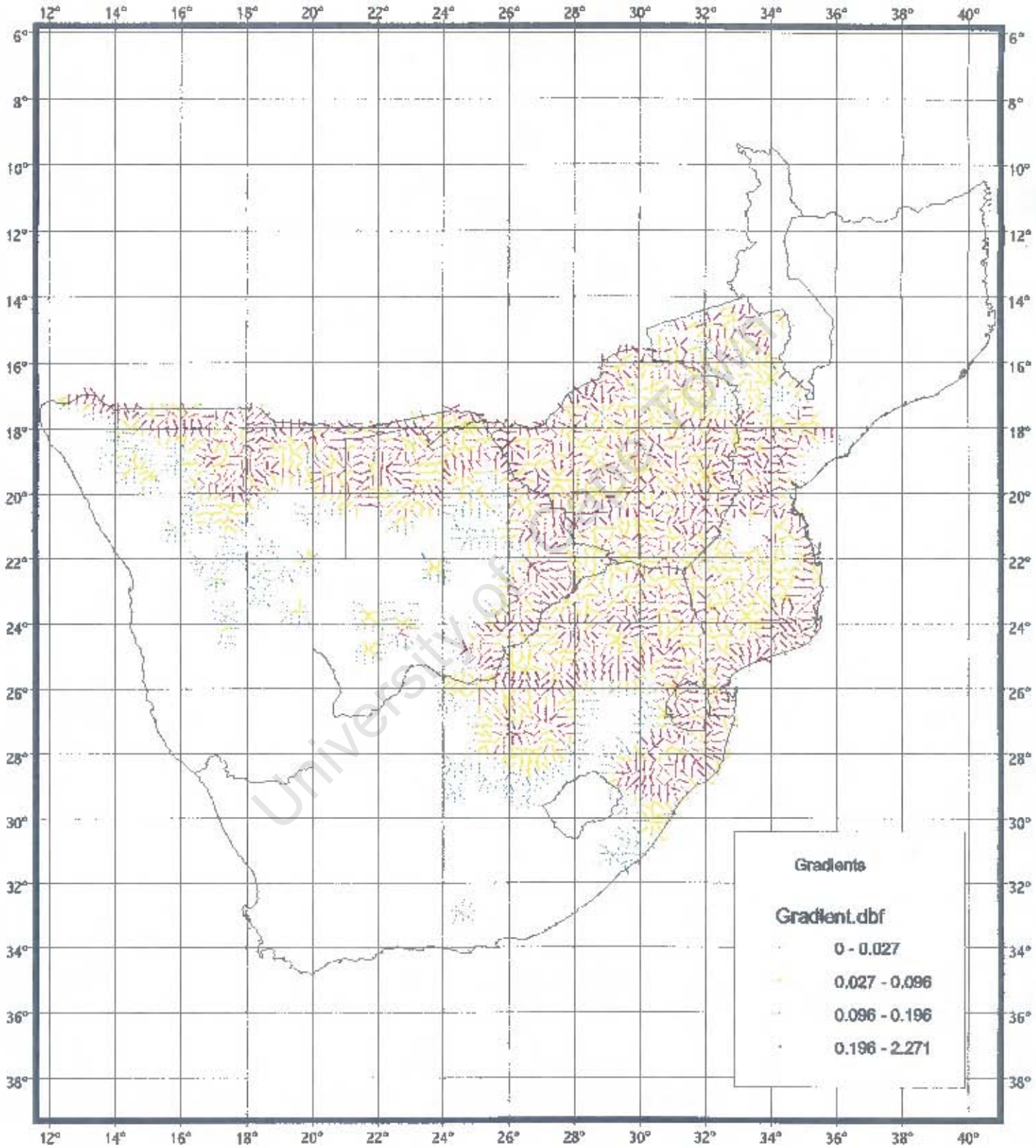


Figure 5: Blue Waxbill - Gradient distribution based on four-parameter model for blocks of 25 grid cells

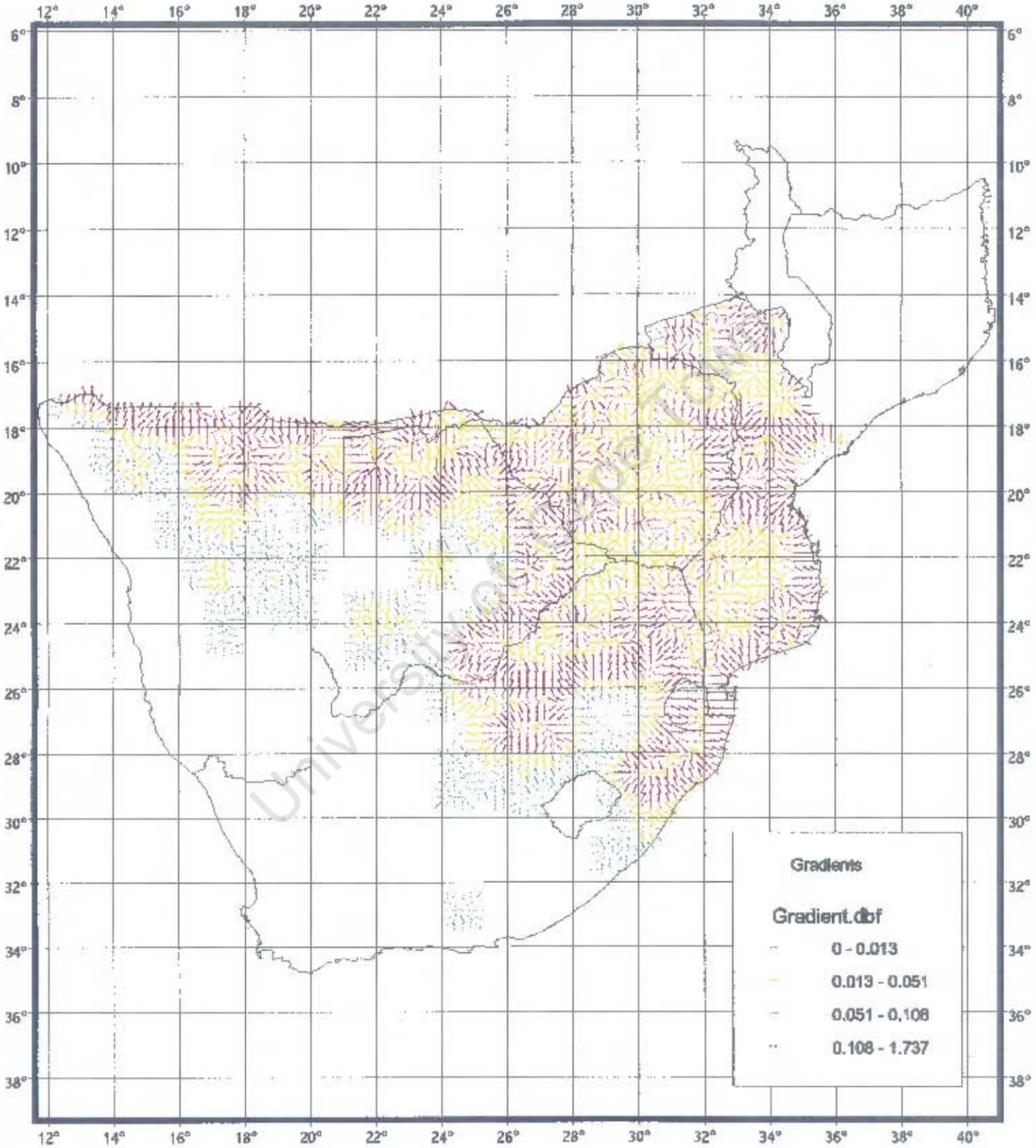


Figure 6: Blue Waxbill - Gradient distribution based on six-parameter model for blocks of 25 grid cells

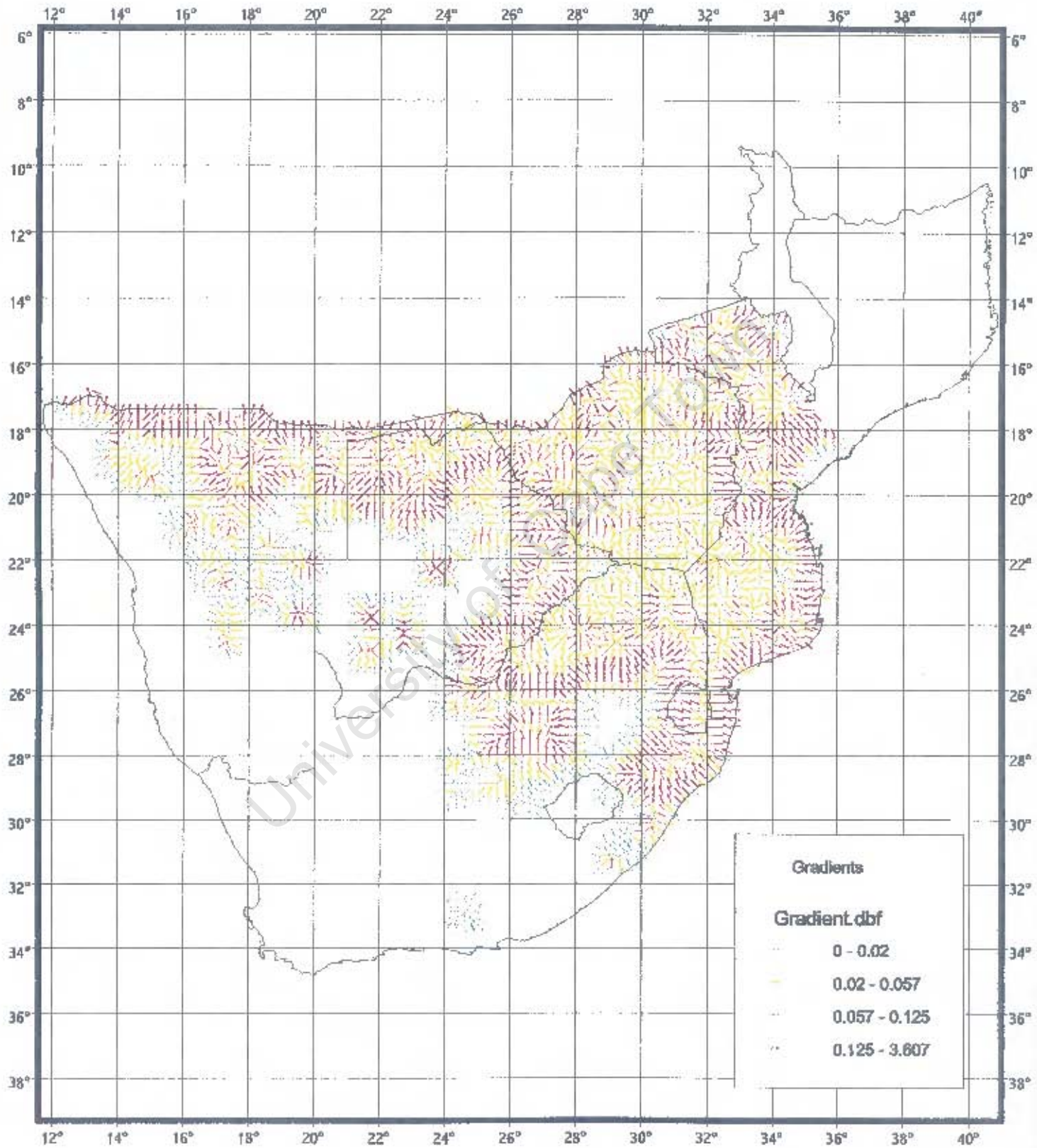
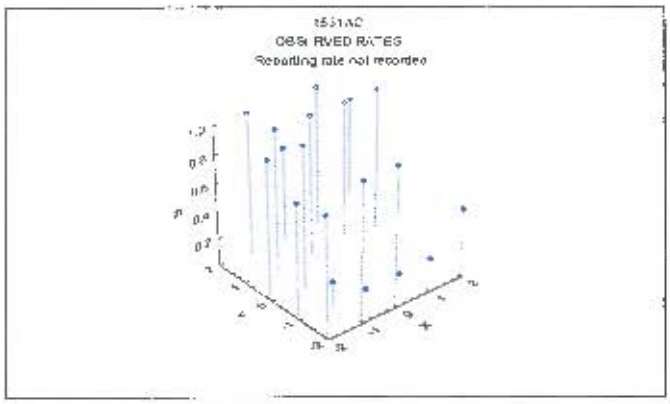
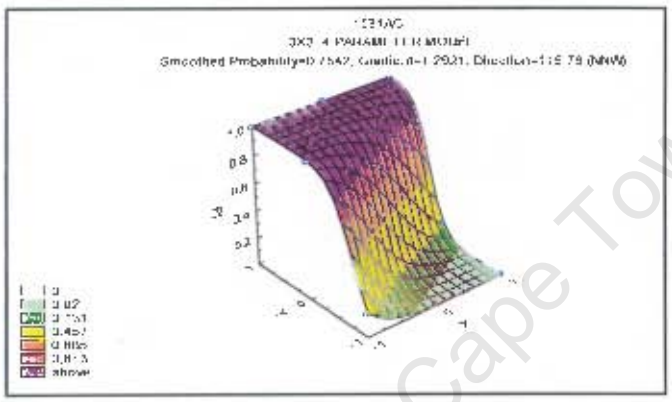


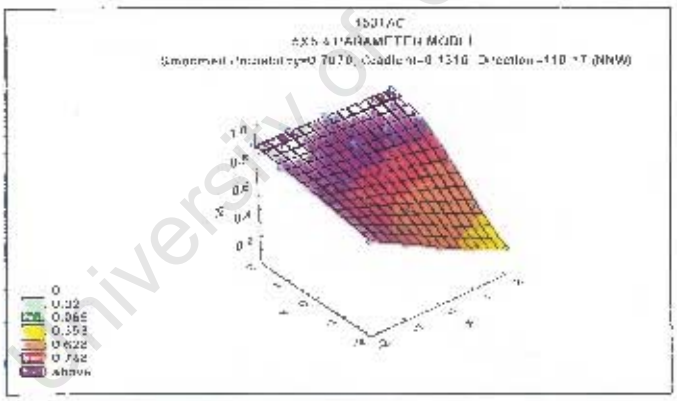
Figure 7: Comparing Three Models for Blue Waxbill
Grid Cell 1631AC



Atlas Coverage
No. of checklists

0	1	3	5	6
0	4	4	1	3
1	2	0	0	0
41	17	3	1	1
8	1	1	0	4





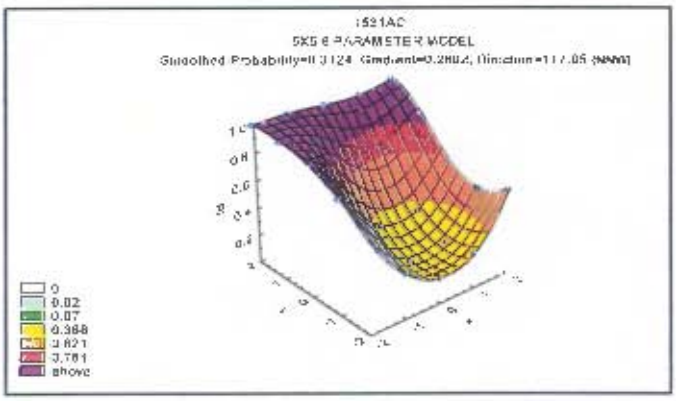
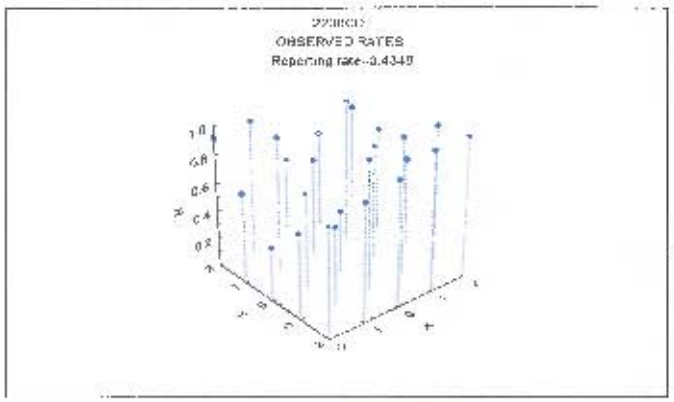
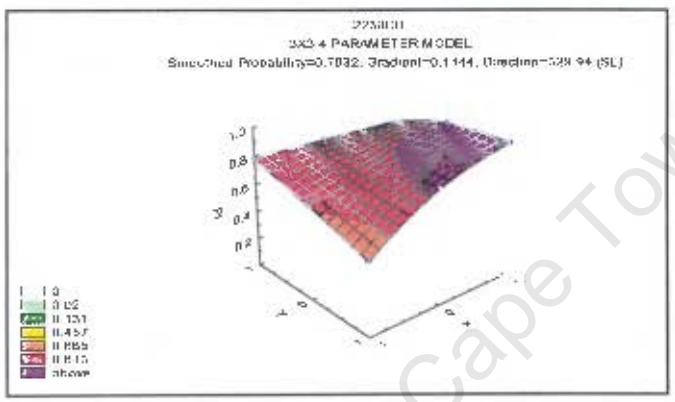


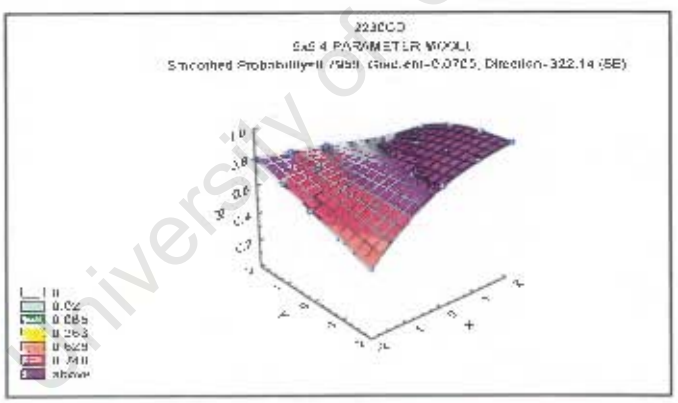
Figure 8: Comparing Three Models for Blue Waxbill Grid Cell 2230CD



Atlas Coverage
No. of checklists

13	55	33	18	22
14	16	16	9	42
23	15	23	15	18
113	140	60	37	28
39	13	31	117	14





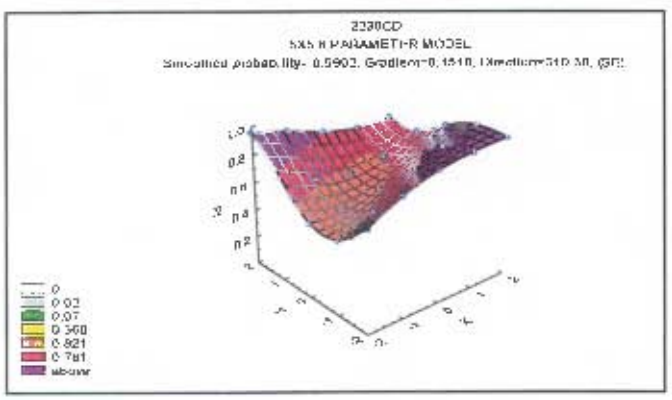
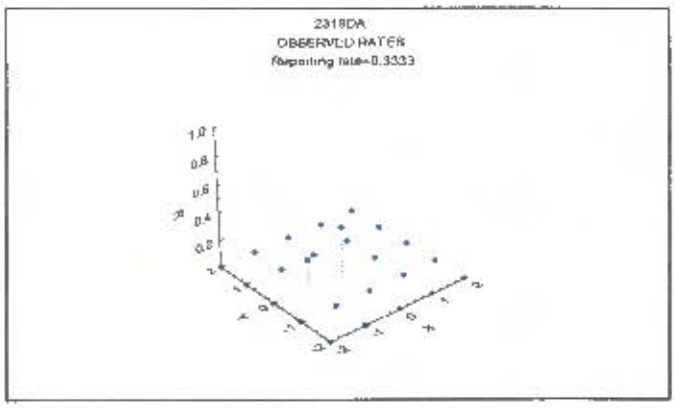
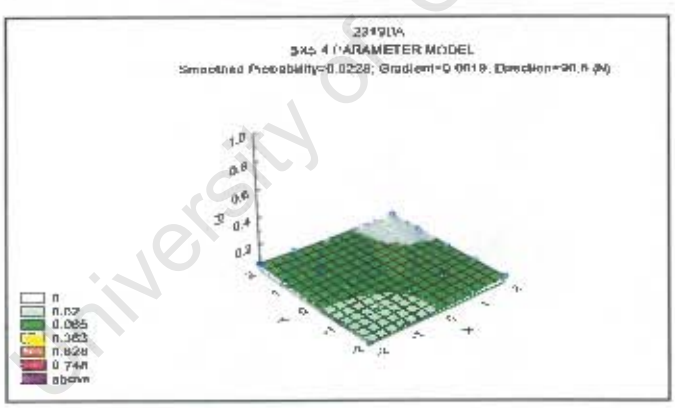


Figure 9 : Comparing Three Models for Blue Waxbill
Grid Cell 2319DA



Atlas Coverage
No. of checklists

1	4	5	2	2
1	3	2	2	2
4	5	3	2	3
2	2	4	4	3
41	2	3	1	1



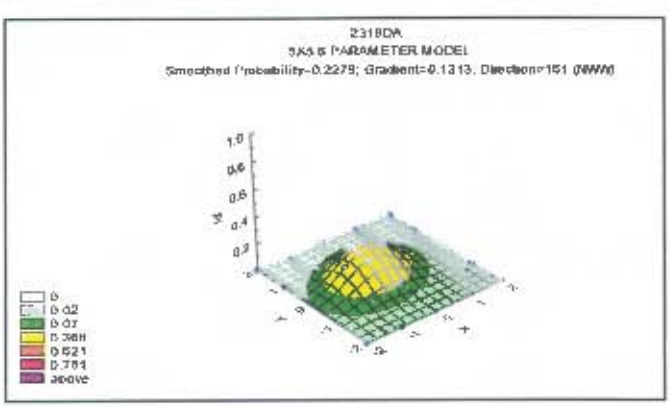
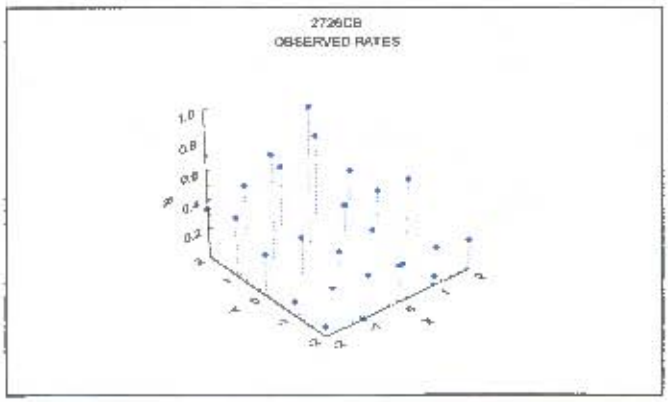
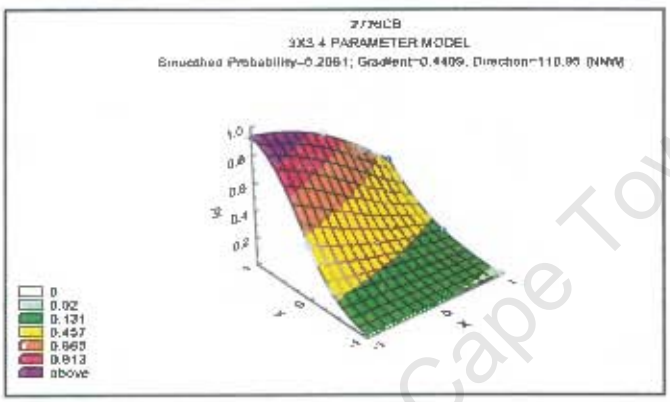


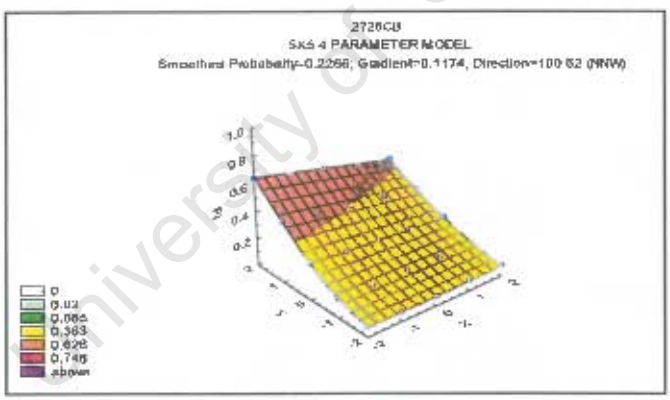
Figure 10 : Comparing Three Models for Blue Waxbill
Grid Cell 2726CB



Atlas Coverage
No. of checklists

36	40	37	51	49
32	30	77	34	22
21	10	12	32	13
60	14	21	165	43
14	11	16	65	72





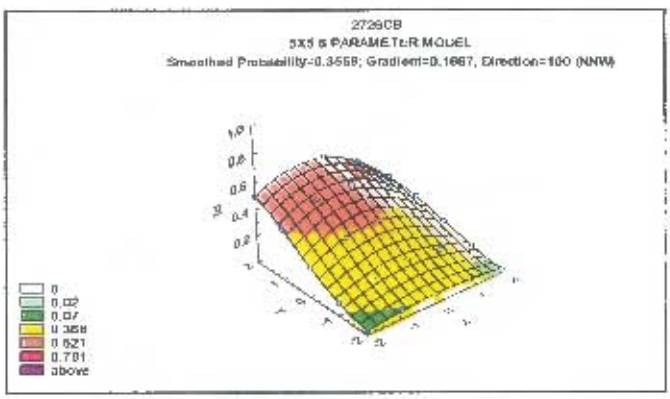


Figure 11: Longtailed Widow - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

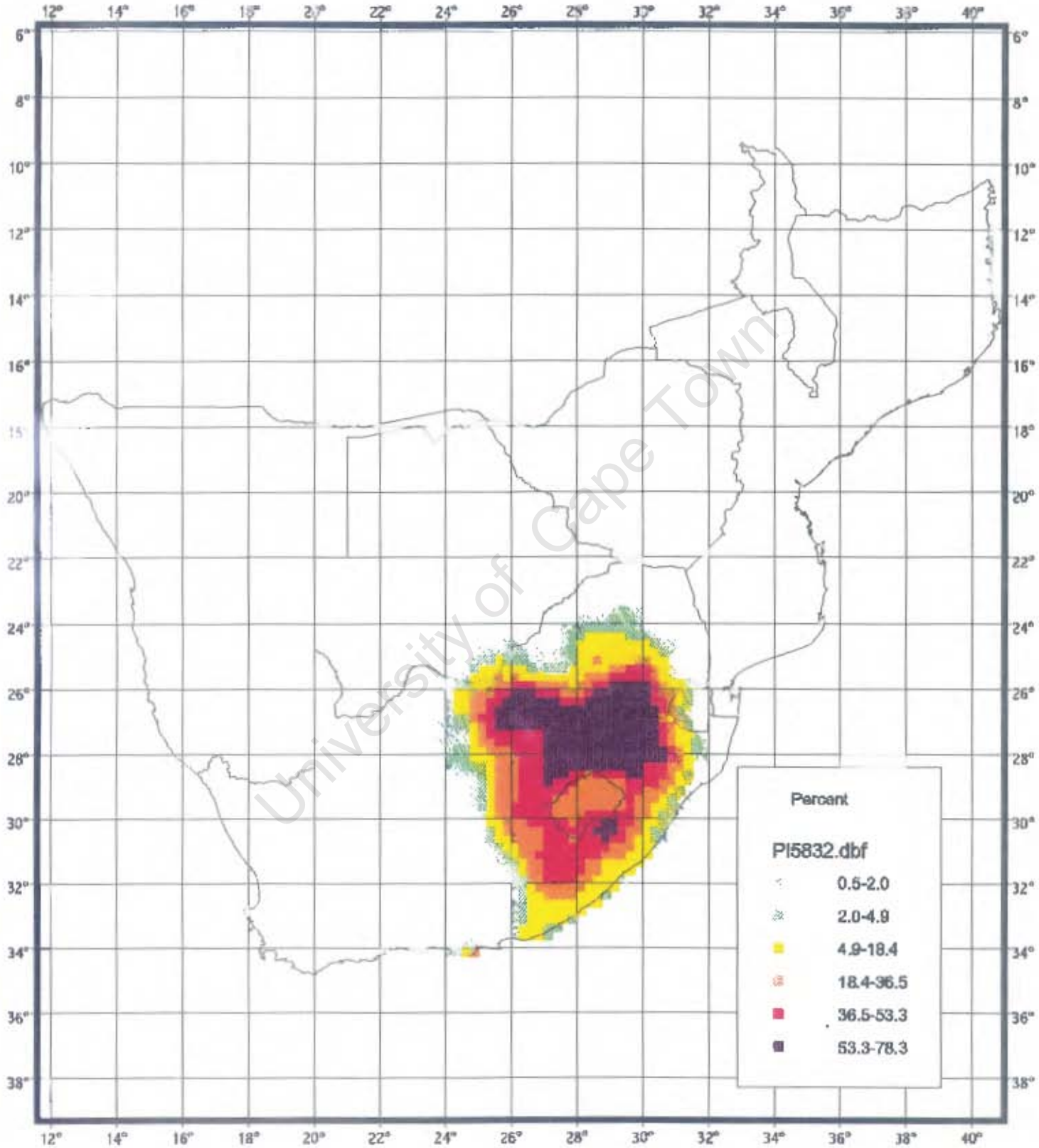


Figure 12: Longtailed Widow - Gradient distribution based on four-parameter model for blocks of 25 grid cells

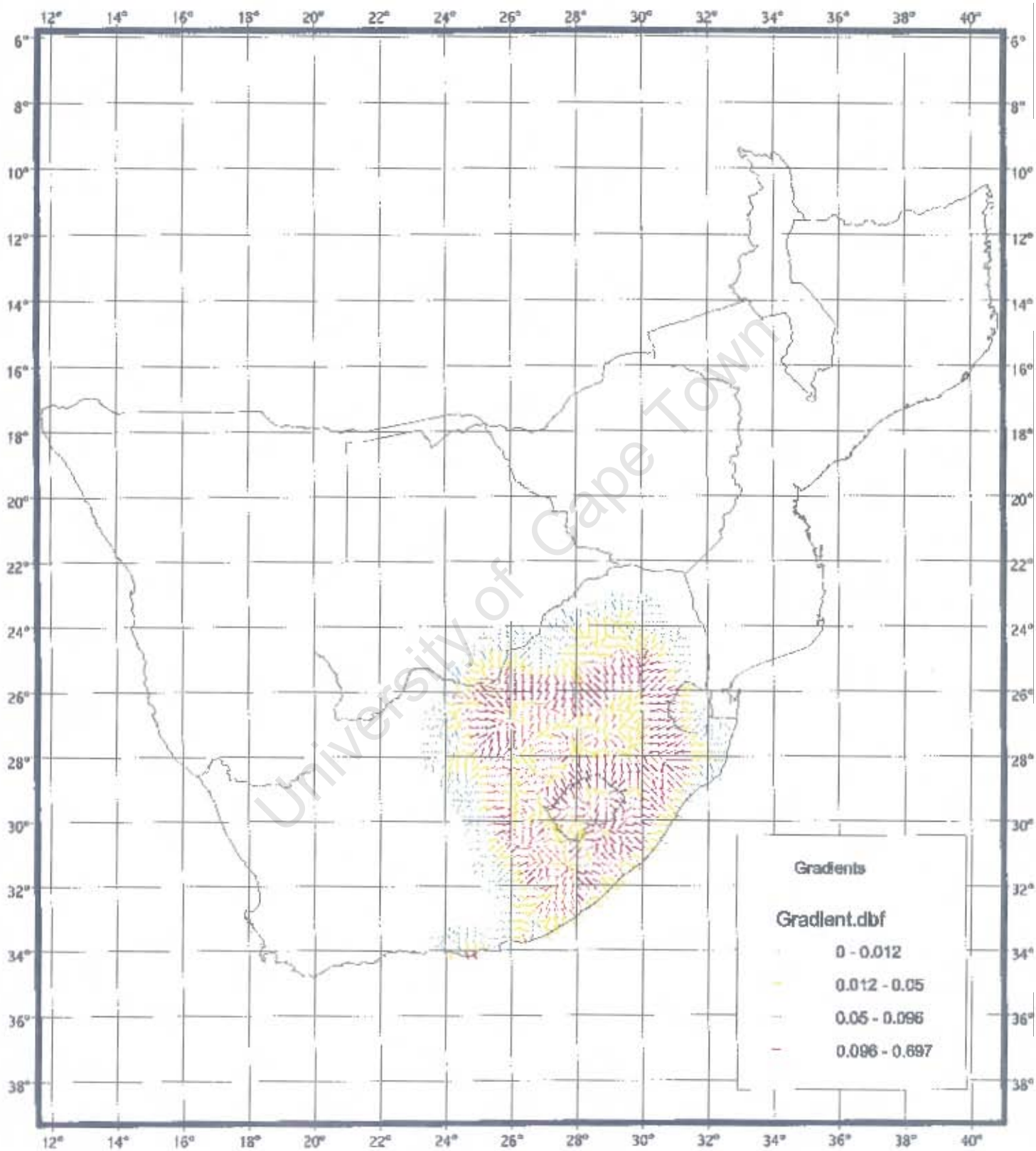


Figure 13: Rattling Cisticola - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

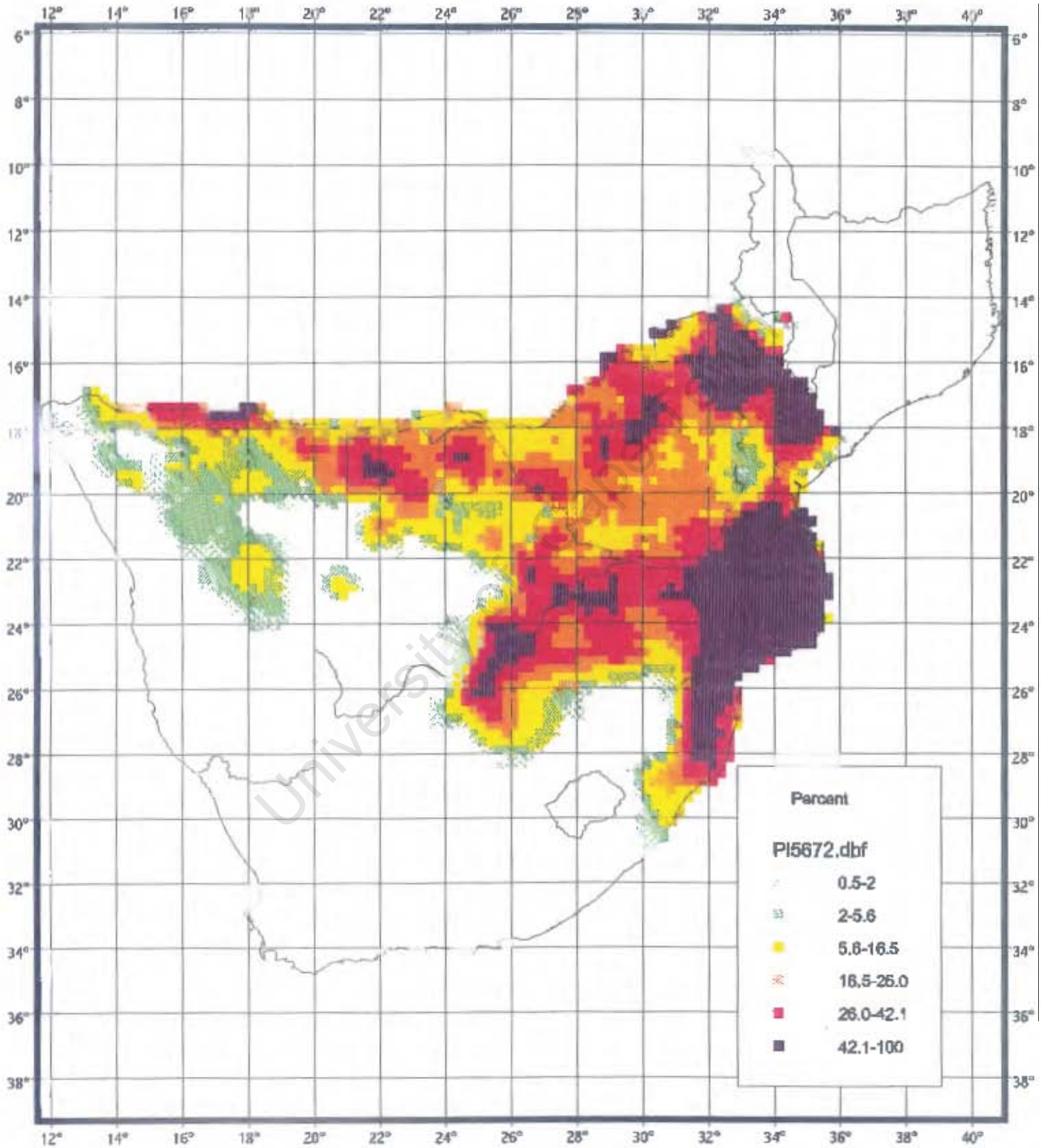


Figure 14: Rattling Cisticola - Gradient distribution based on four-parameter model for blocks of 25 grid cells

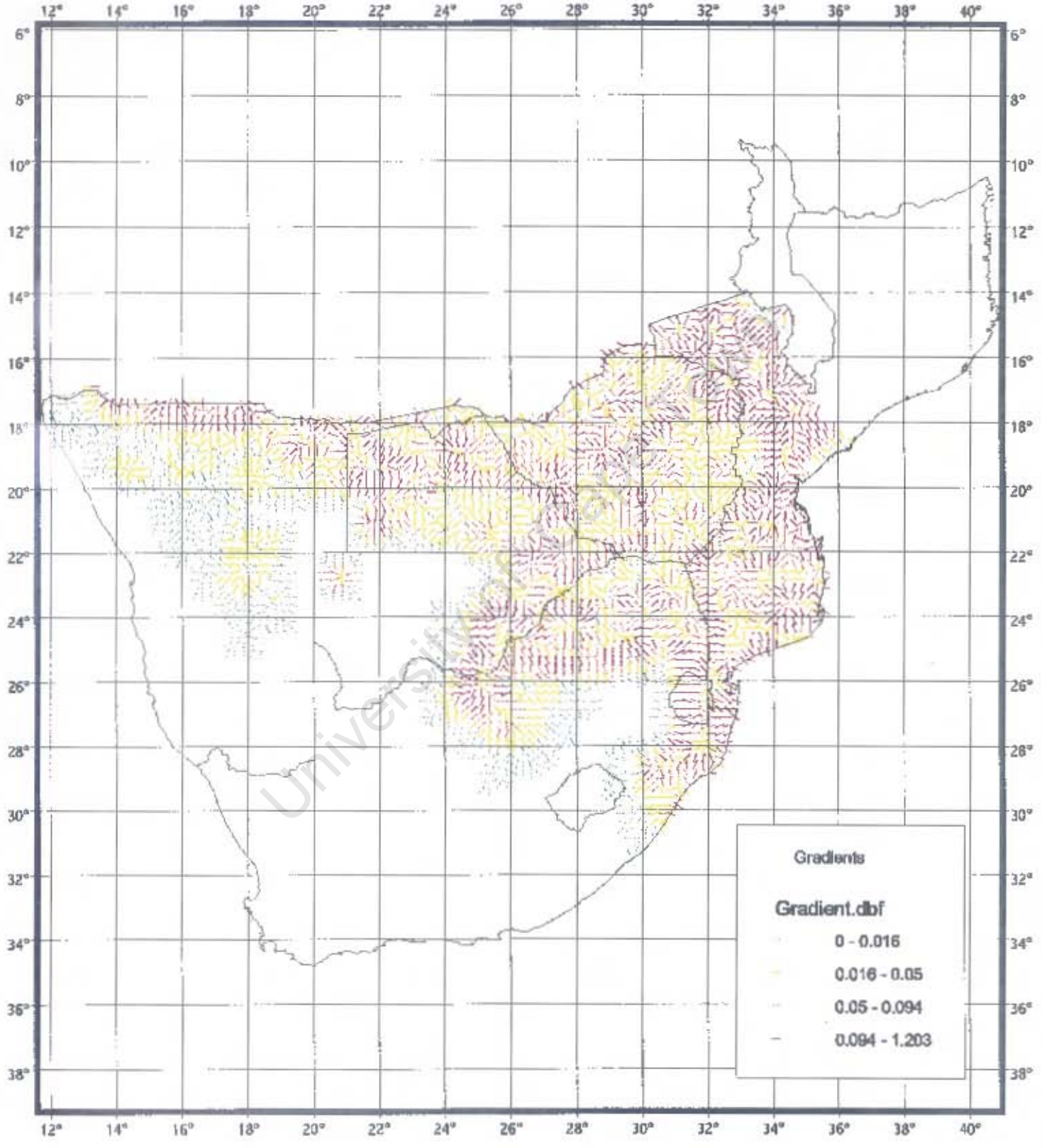


Figure 15: Cloud Cisticola - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

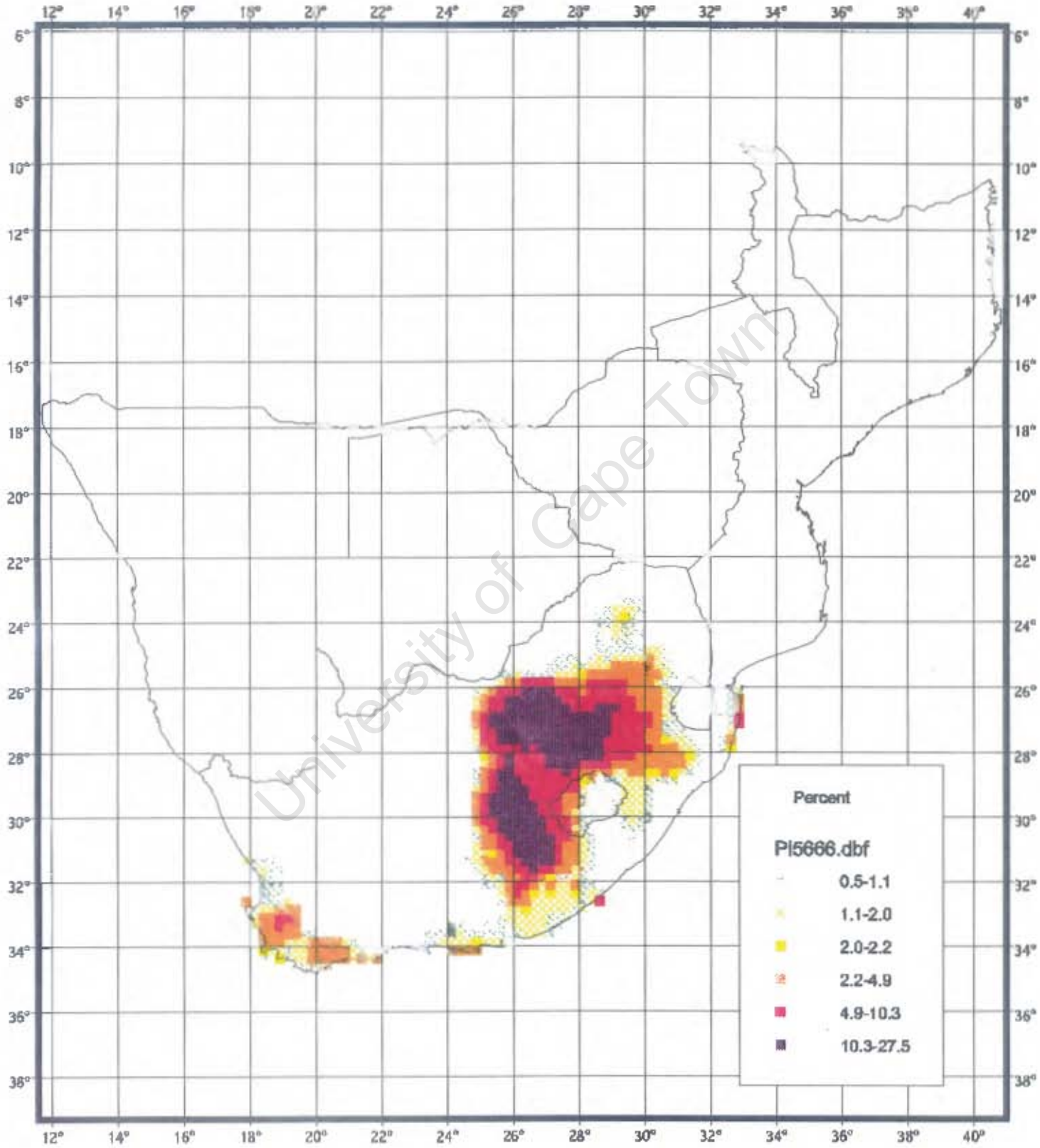


Figure 16: Cloud Cisticola - Gradient distribution based on four-parameter model for blocks of 25 grid cells using species-specific colour coding

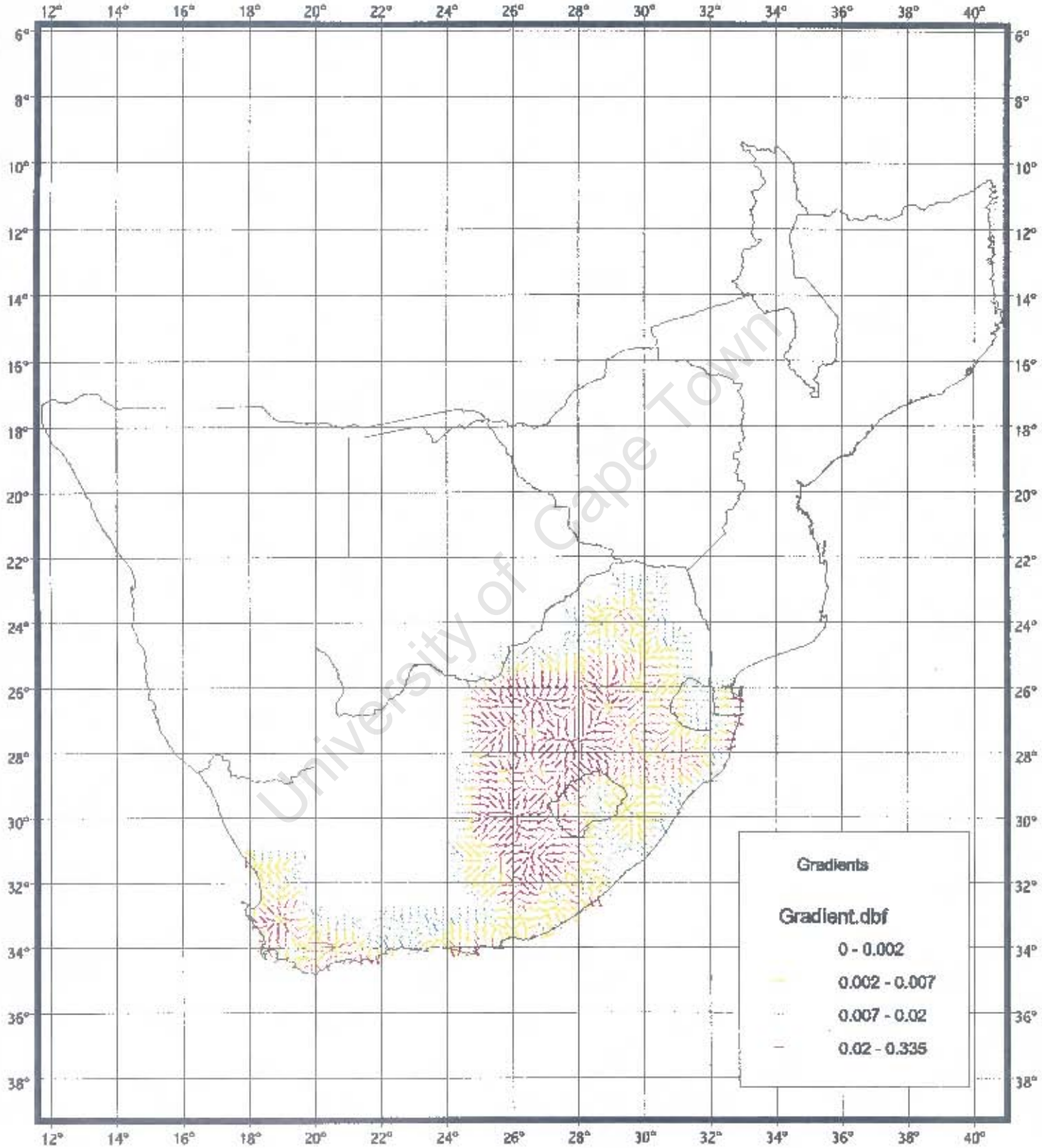


Figure 17: Cloud Cisticola - Gradient distribution based on four-parameter model for blocks of 25 grid cells using absolute colour coding based on gradient ranges for all 12 species

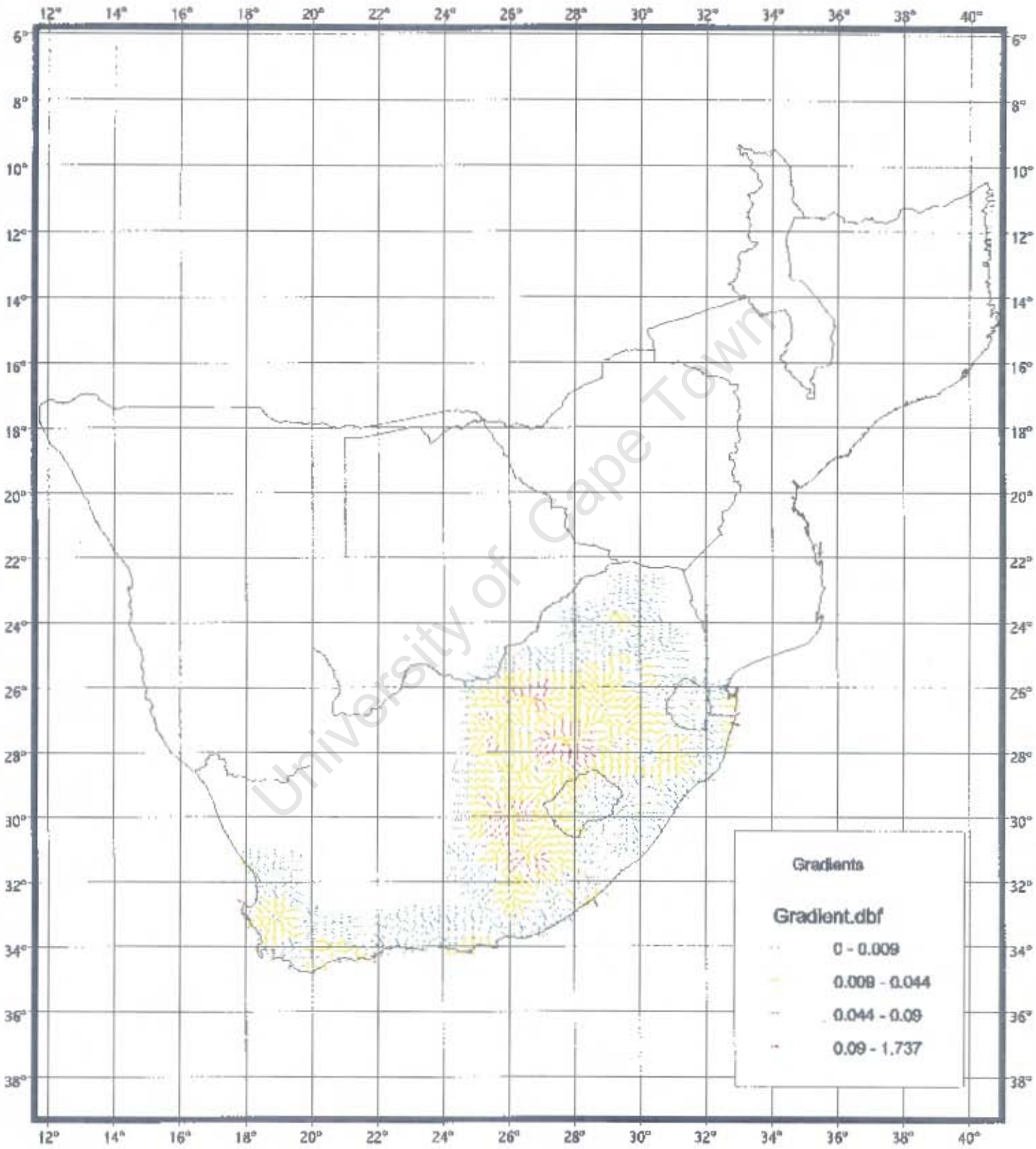


Figure 18: Arrowmarked Babbler - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

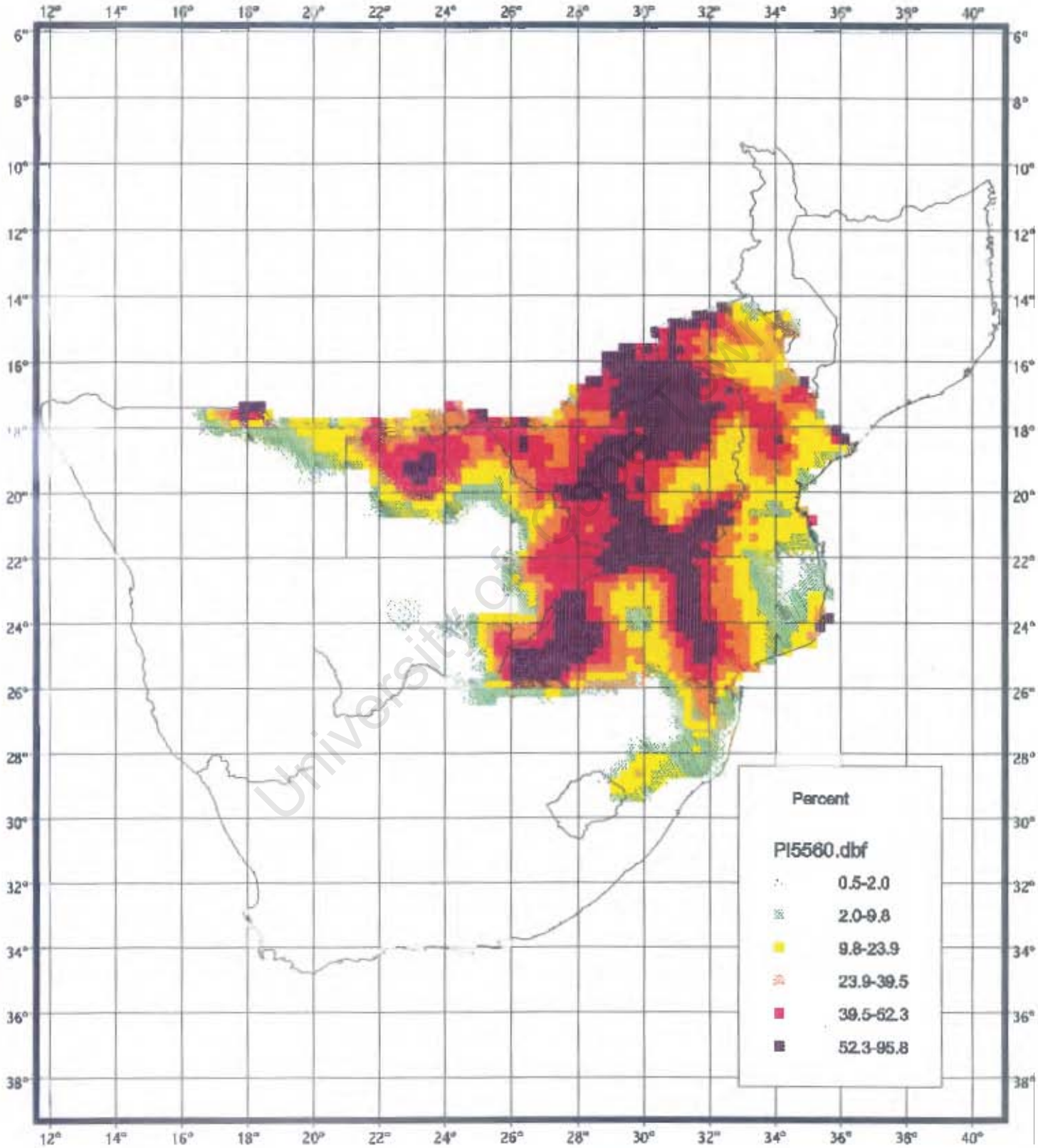


Figure 19: Arrowmarked Babbler - Gradient distribution based on four-parameter model for blocks of 25 grid cells

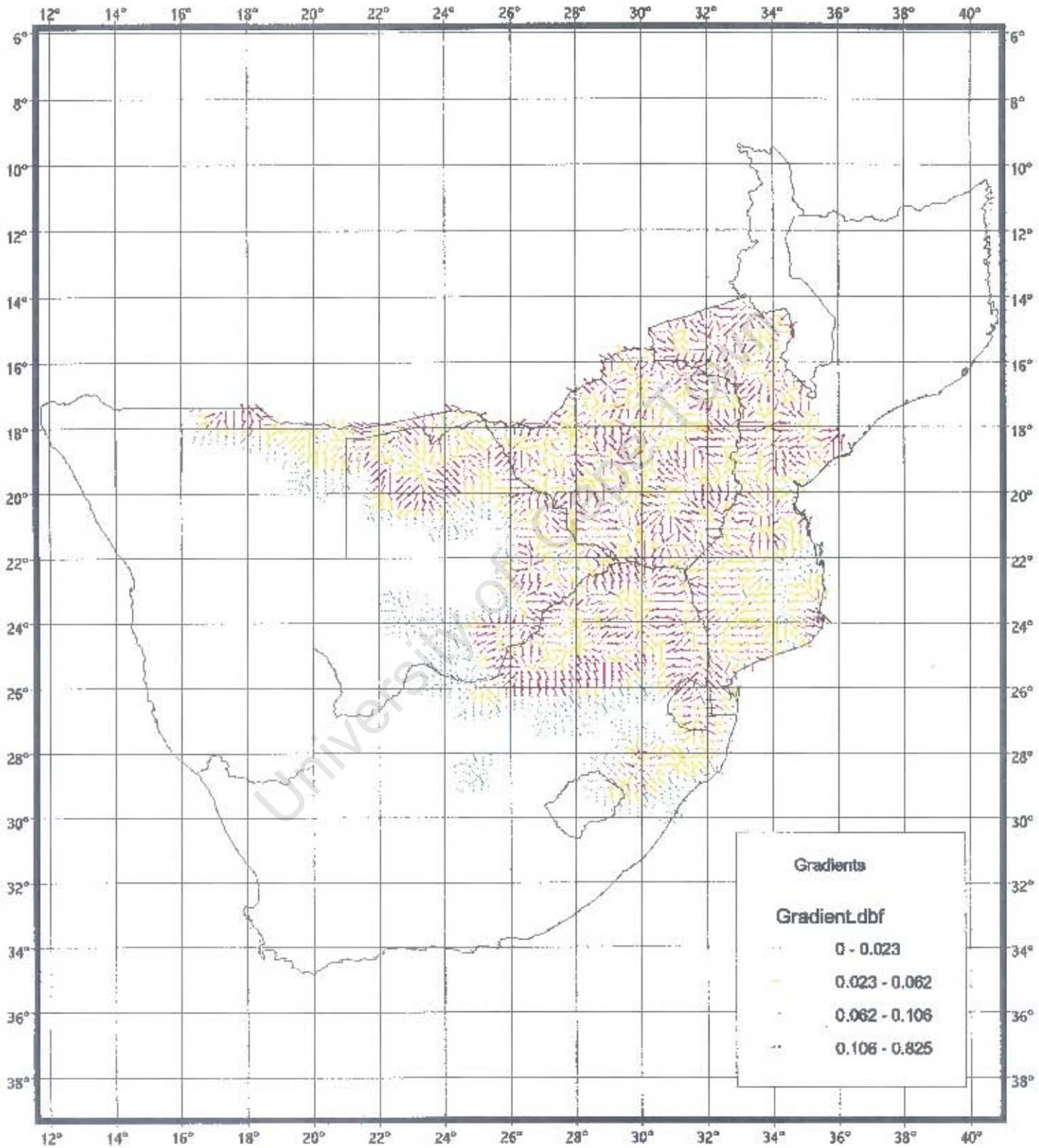


Figure 20: Southern Black Tit - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

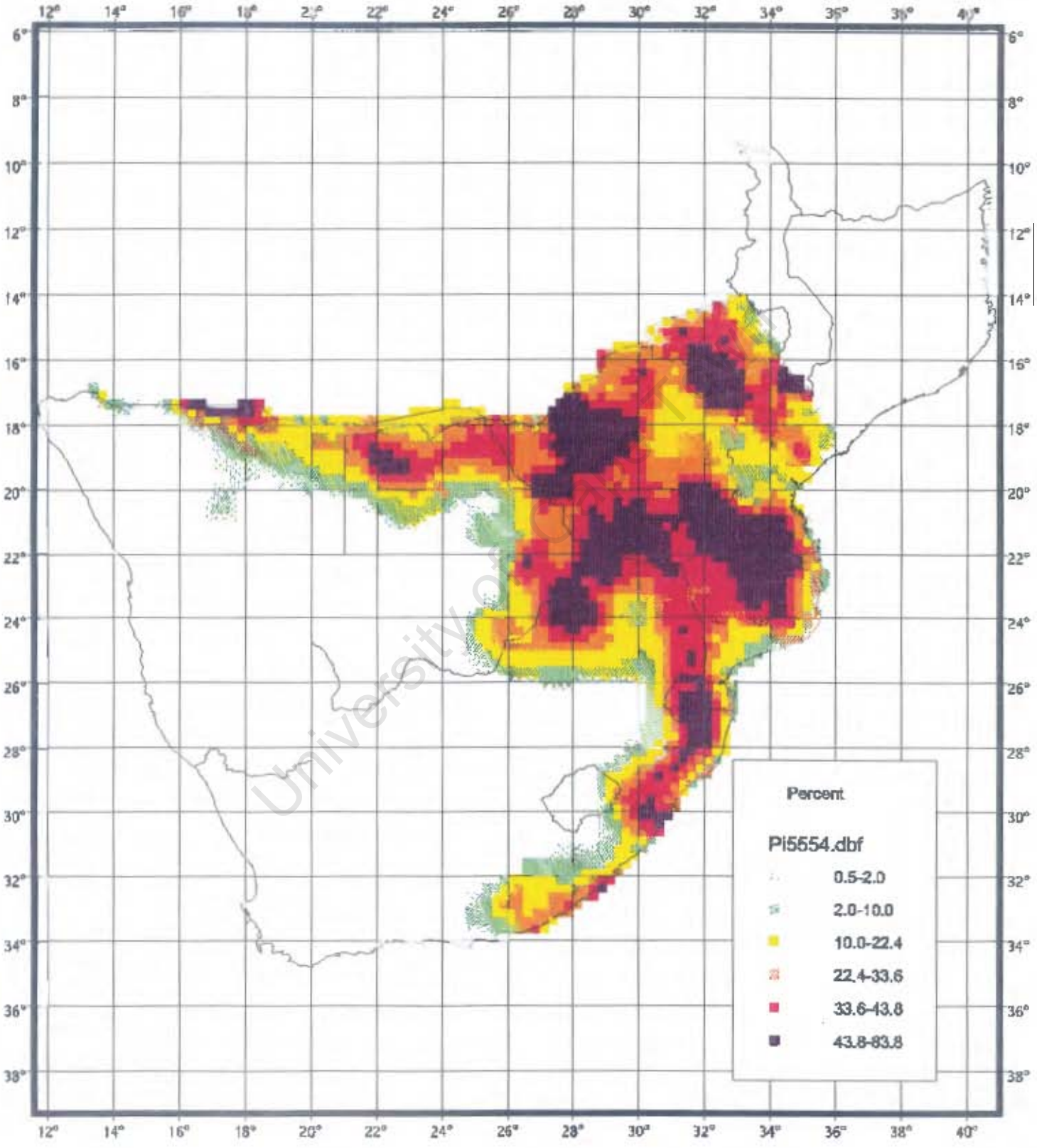


Figure 21: Southern Black Tit - Gradient distribution based on four-parameter model for blocks of 25 grid cells

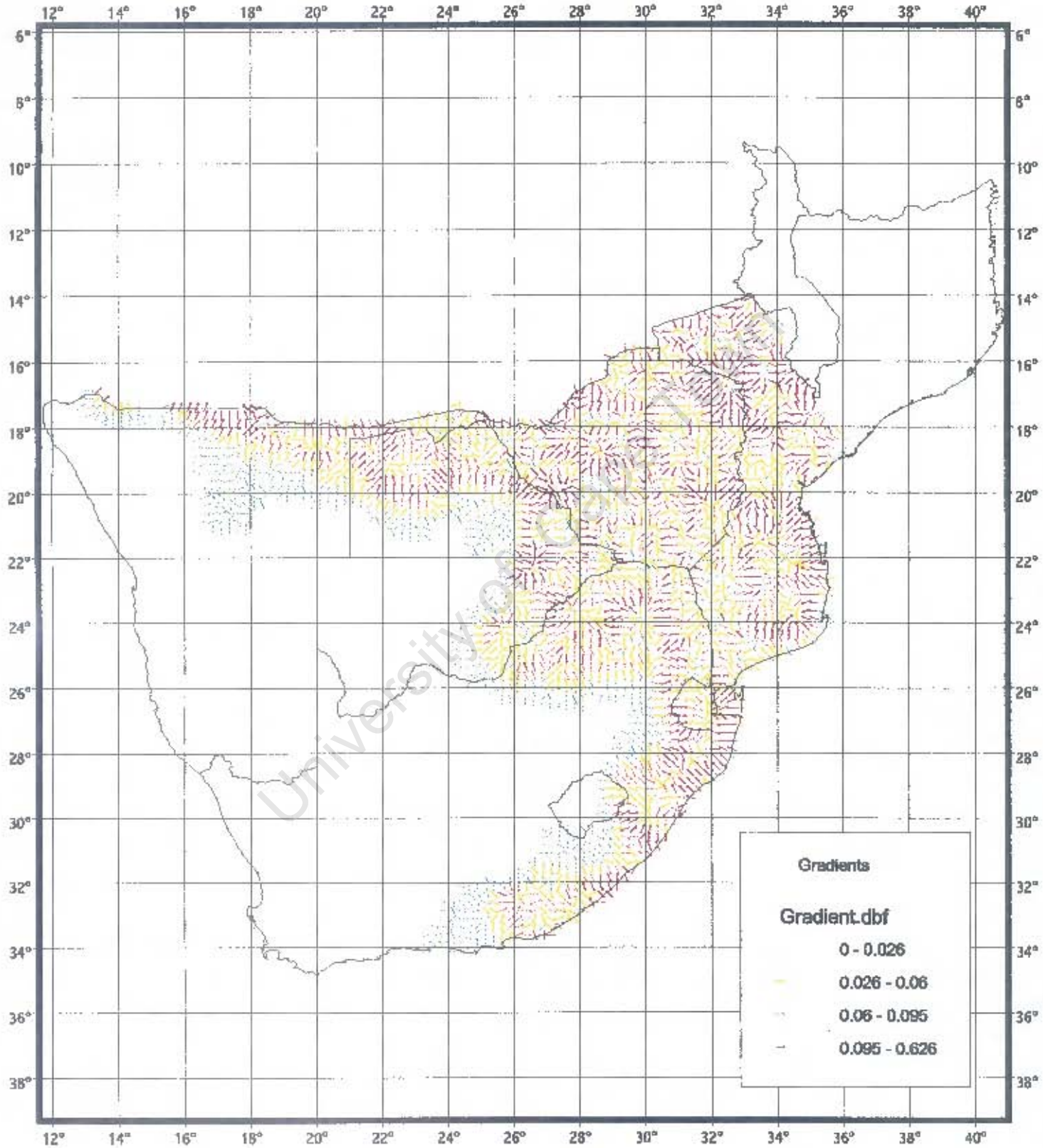


Figure 22: Blackheaded Oriole - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

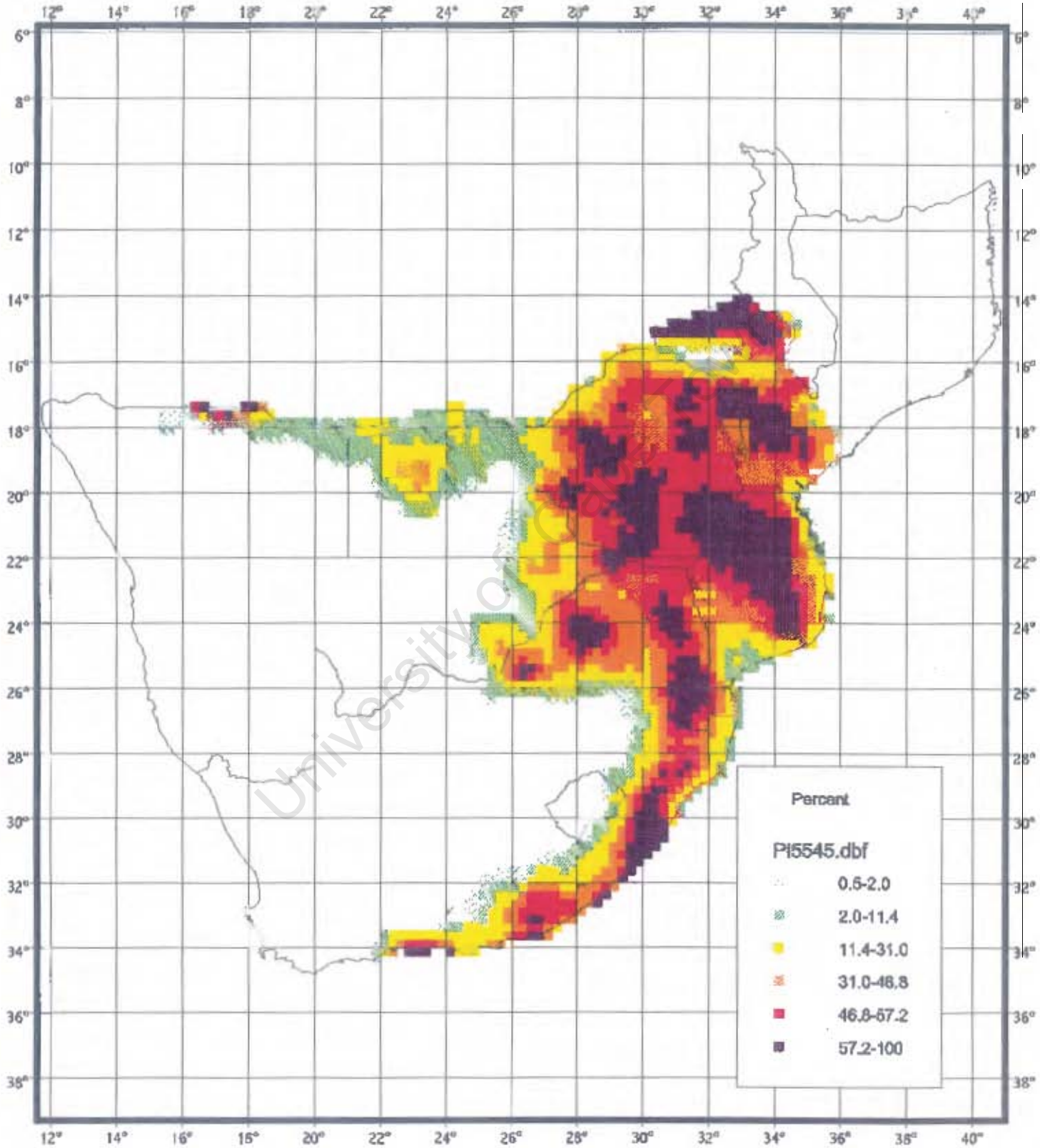


Figure 23: Blackheaded Oriole - Gradient distribution based on four-parameter model for blocks of 25 grid cells

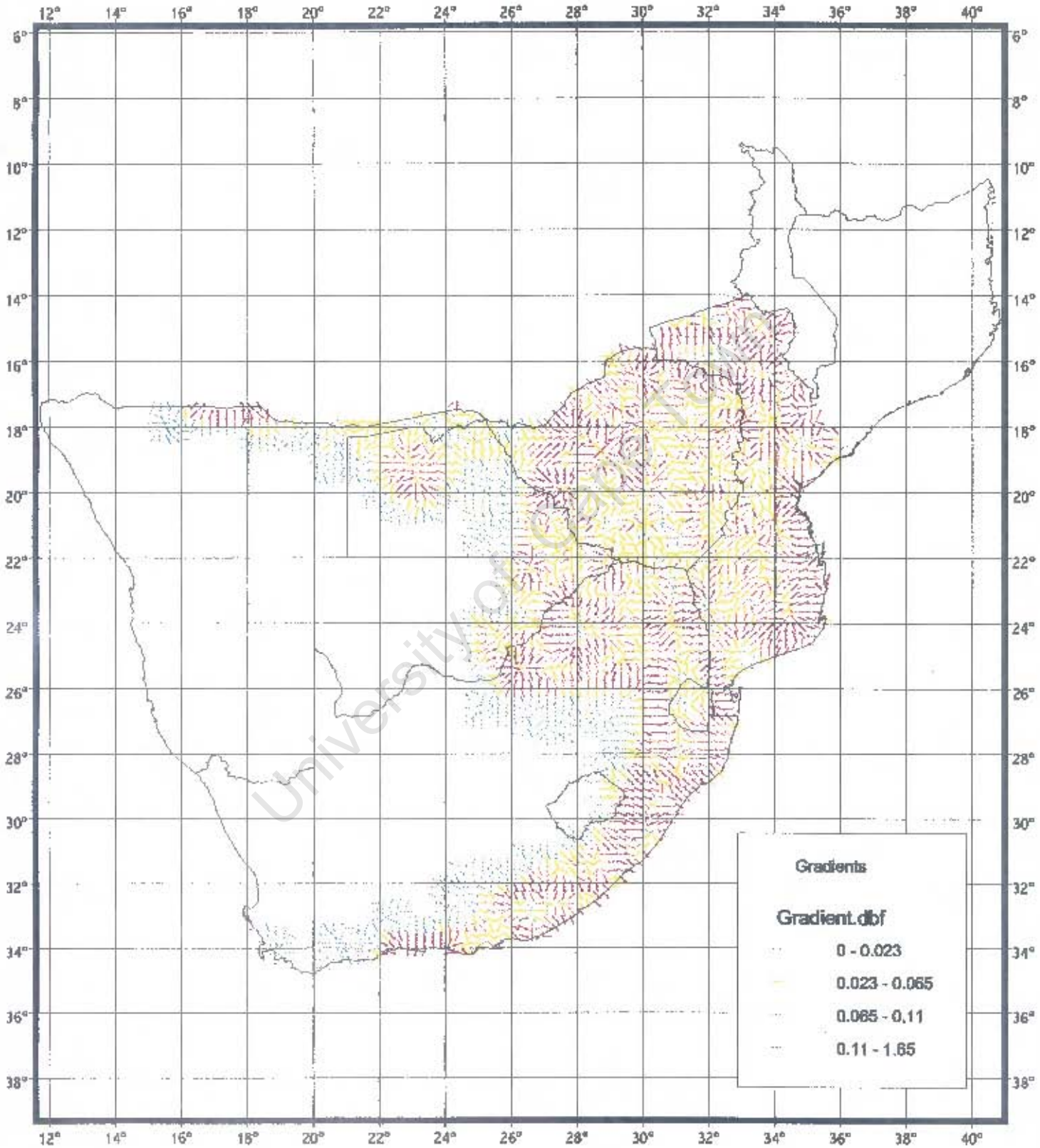


Figure 24: South African Cliff Swallow - Smoothed distribution based on four-parameter model for blocks of 25 grid cell

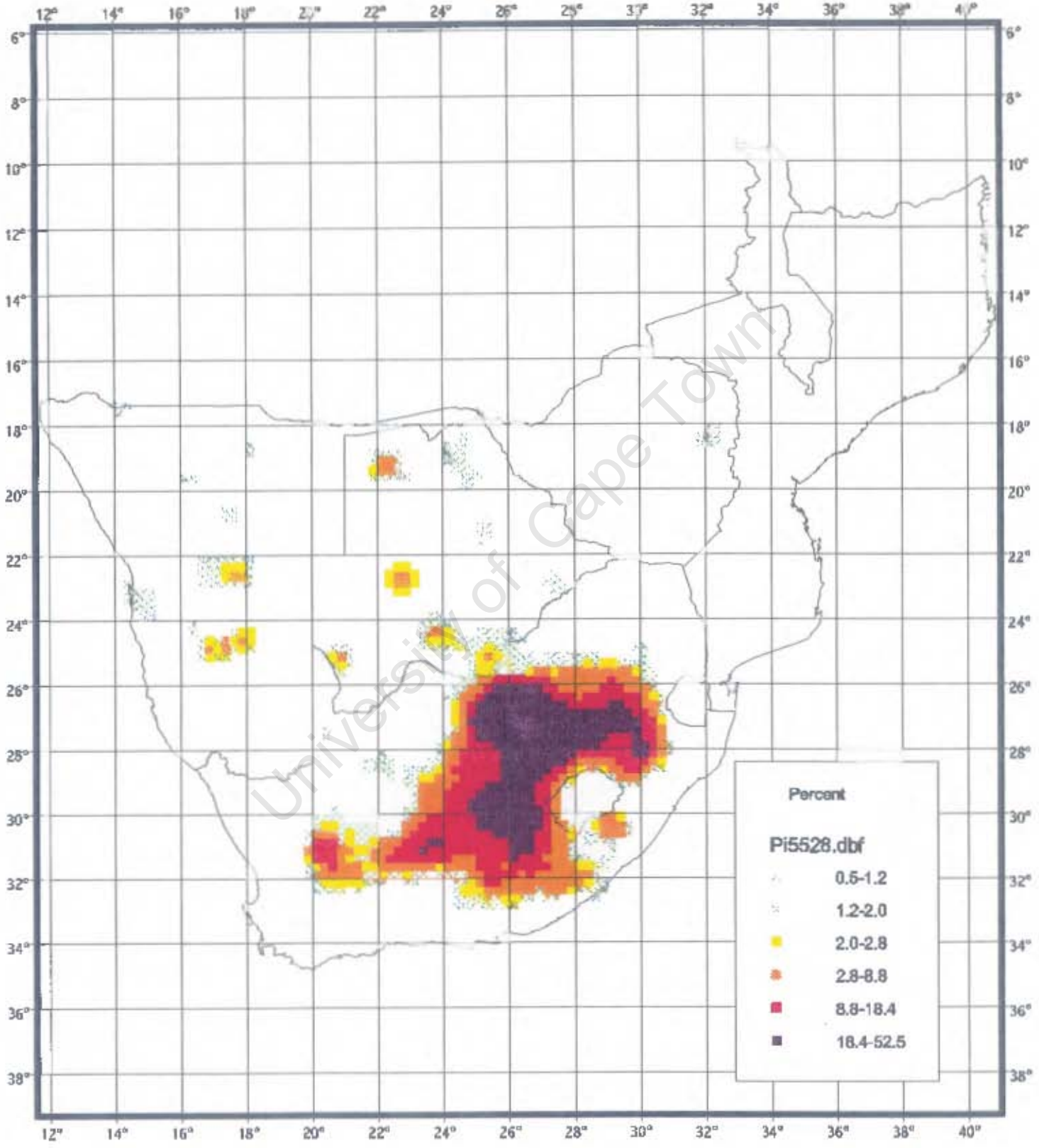


Figure 25: SA Cliff Swallow - Gradient distribution based on four-parameter model for blocks of 25 grid cells

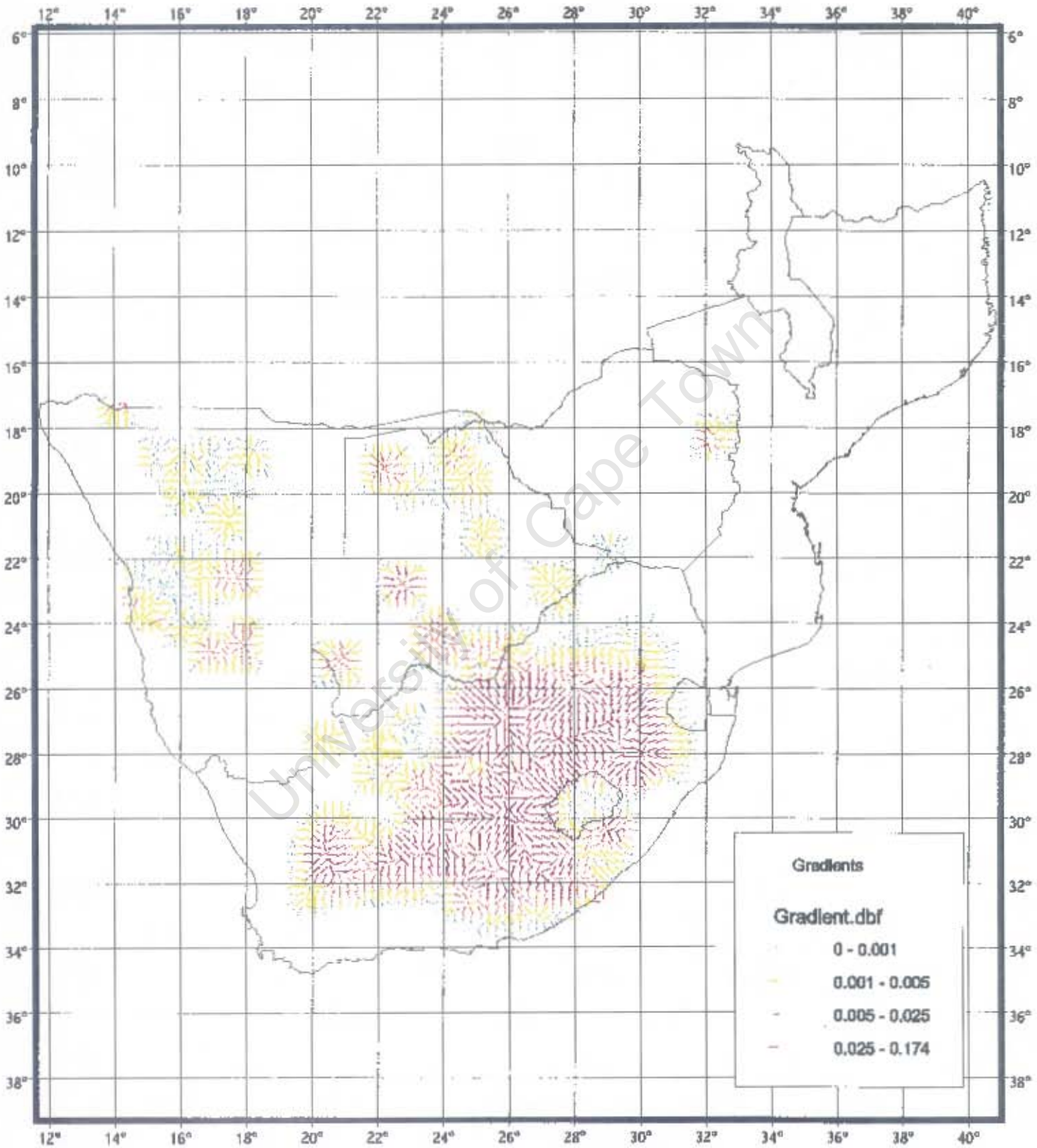


Figure 26: Lesser Striped Swallow - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

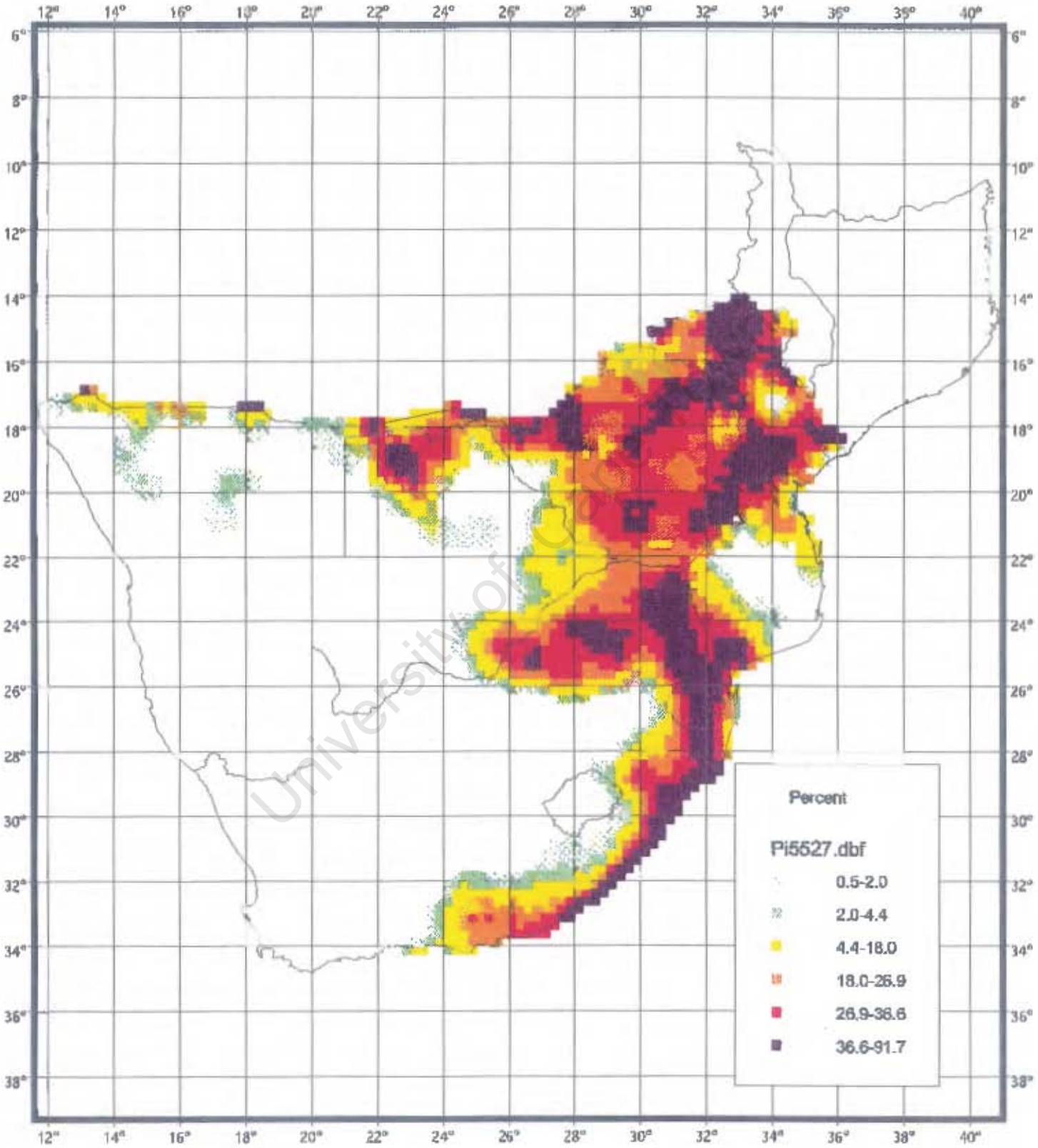


Figure 27: Lesser Striped Swallow - Gradient distribution based on four-parameter model for blocks of 25 grid cells

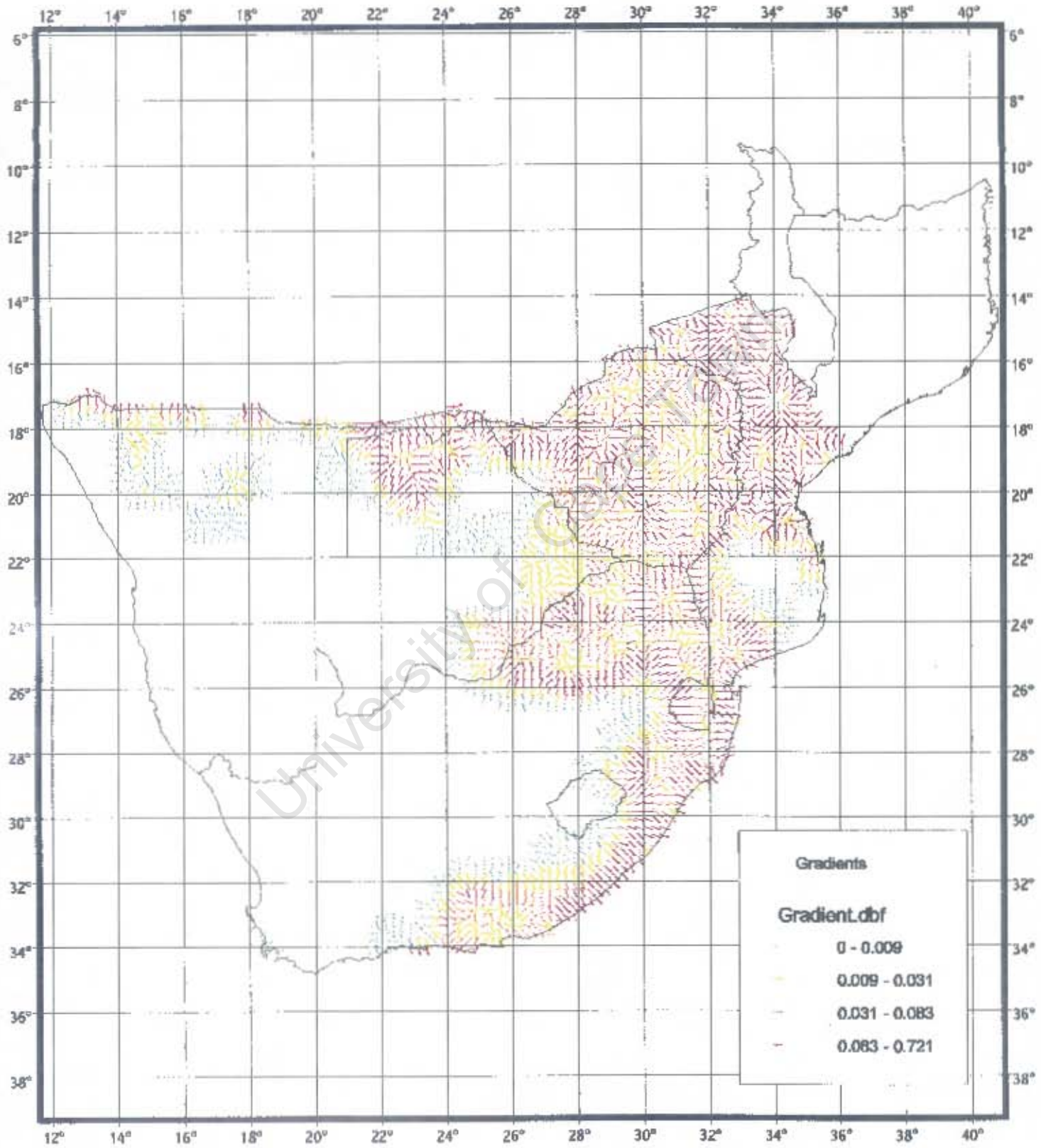


Figure 28: Spikeheeled Lark - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

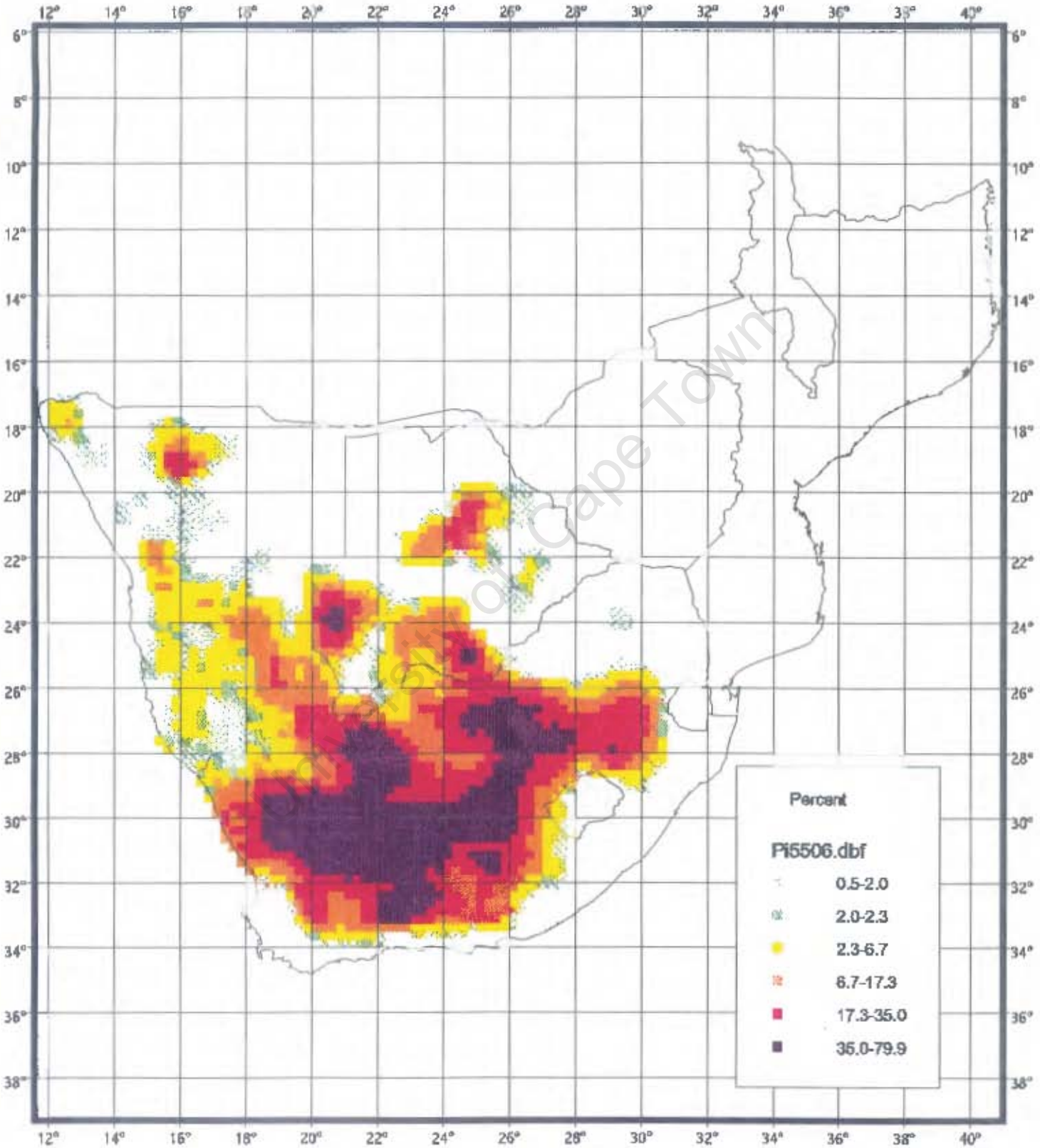


Figure 29: Spikeheeled Lark - Gradient distribution based on four-parameter model for blocks of 25 grid cells

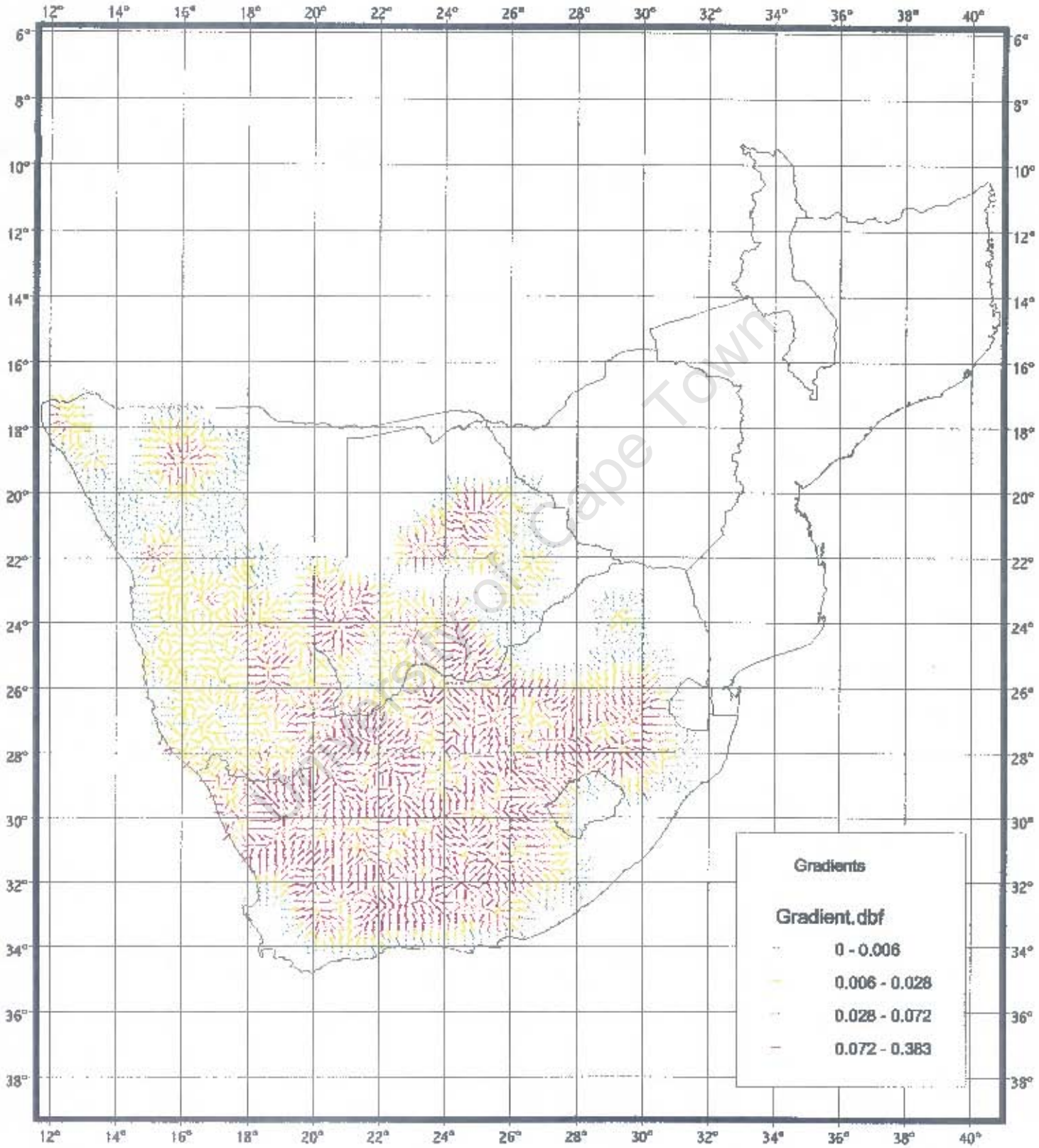


Figure 30: Pied Starling - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

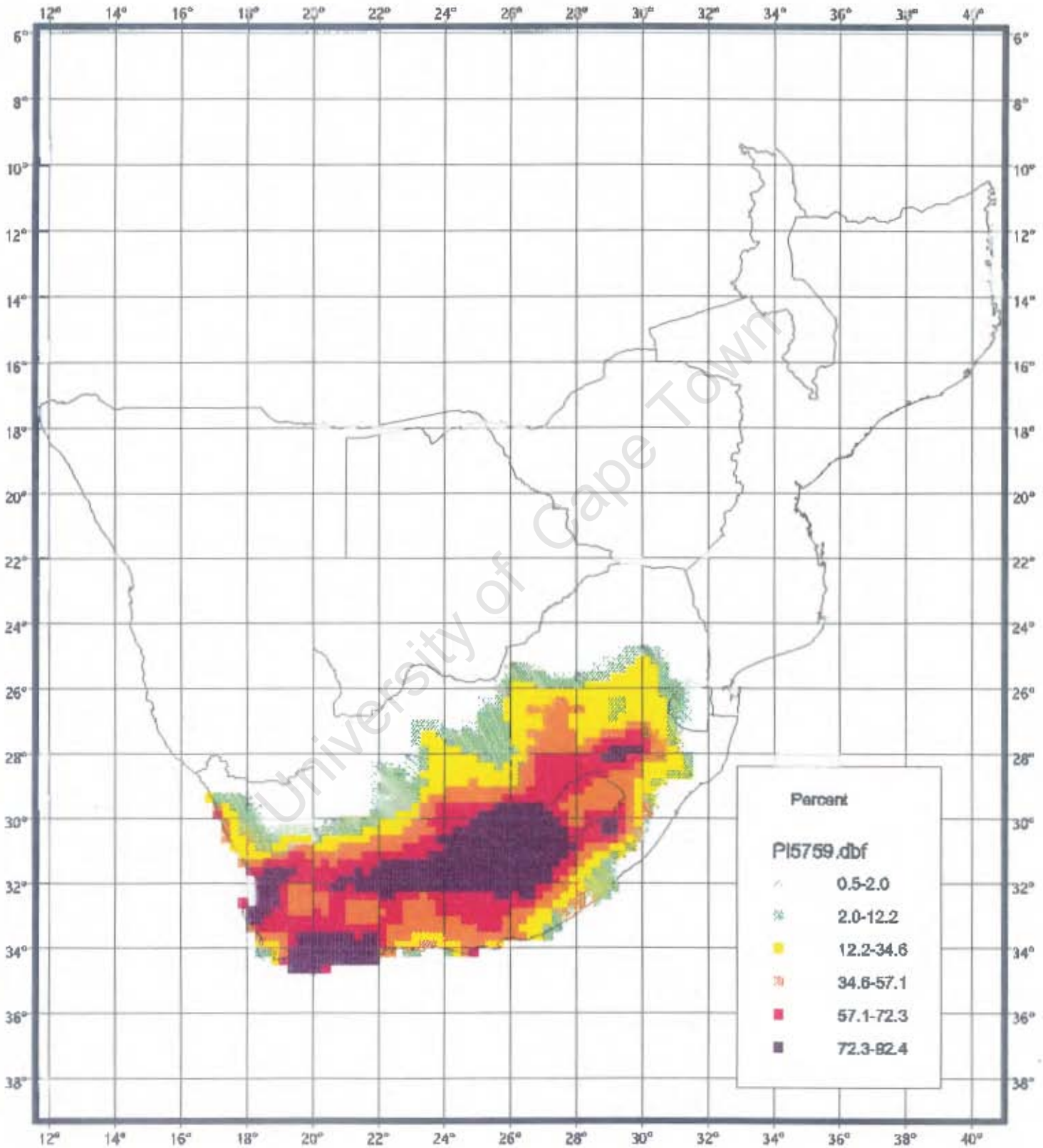


Figure 31: Pied Starling - Gradient distribution based on four-parameter model for blocks of 25 grid cells

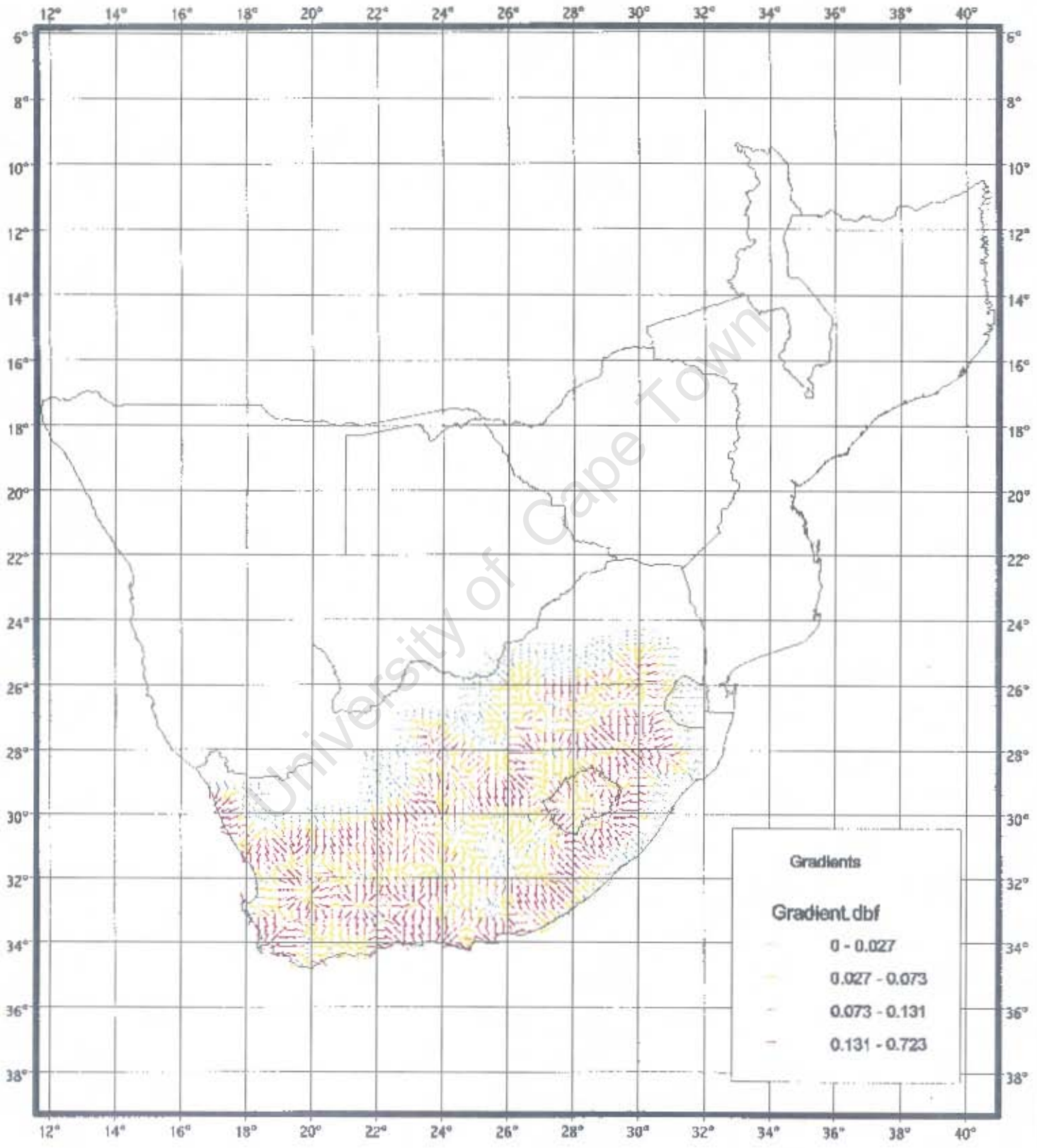


Figure 32: Bokmakierie - Smoothed distribution based on four-parameter model for blocks of 25 grid cells

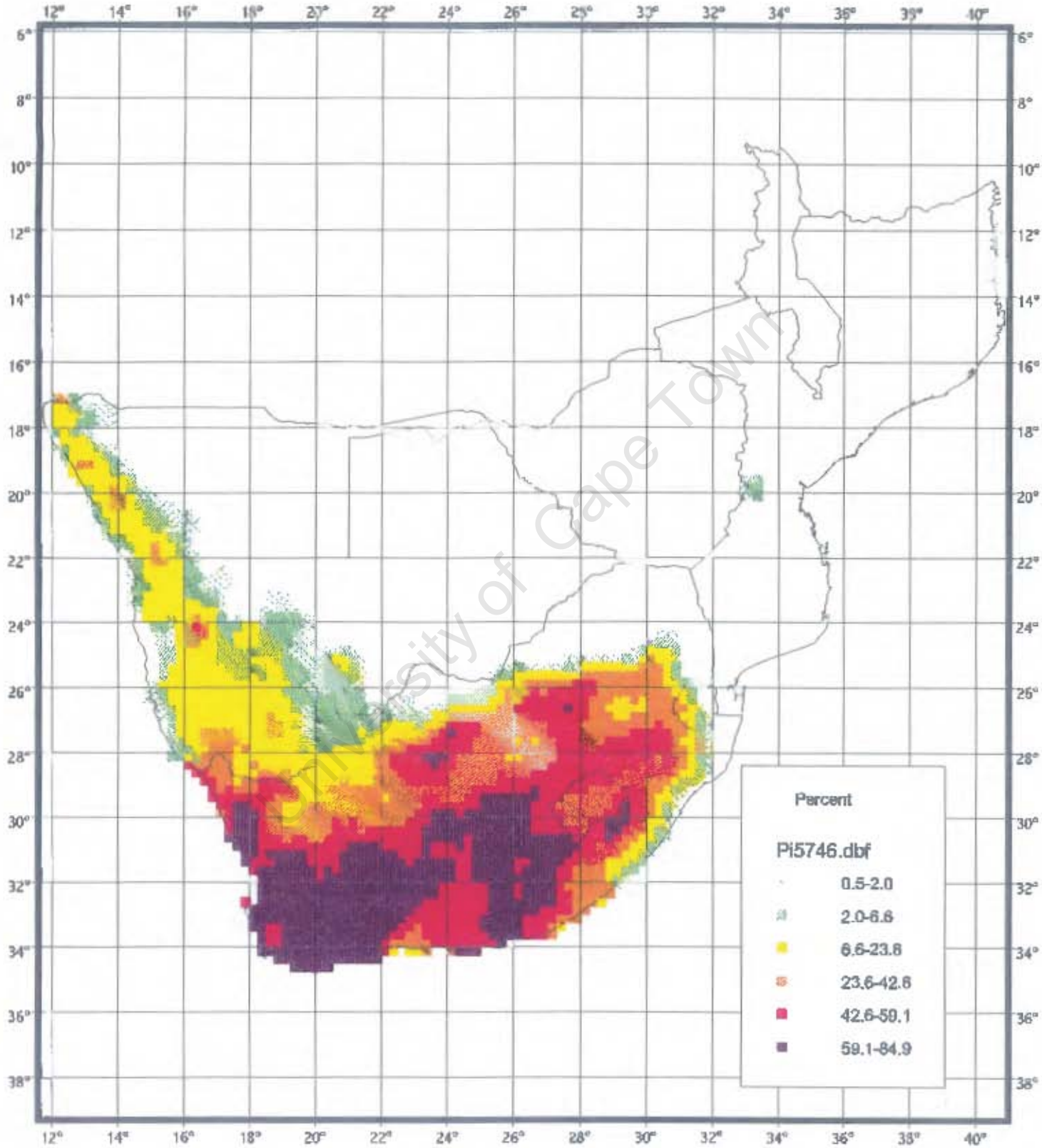


Figure 33: Bokmakierie - Gradient distribution based on four-parameter model for blocks of 25 grid cells

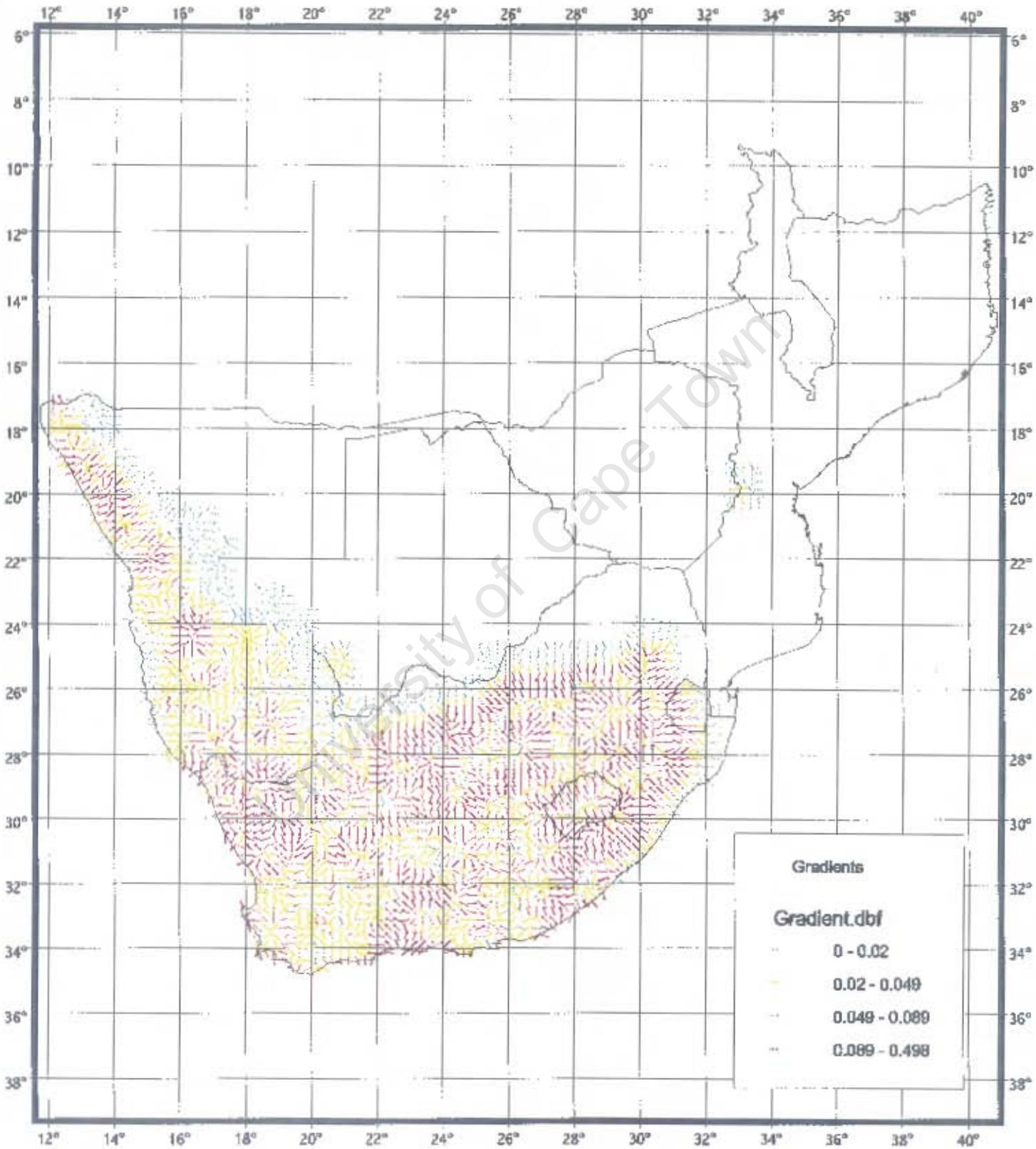


Figure 33a: Bateleur - Smoothed Distribution based on four-parameter model for blocks of 25 grid cells

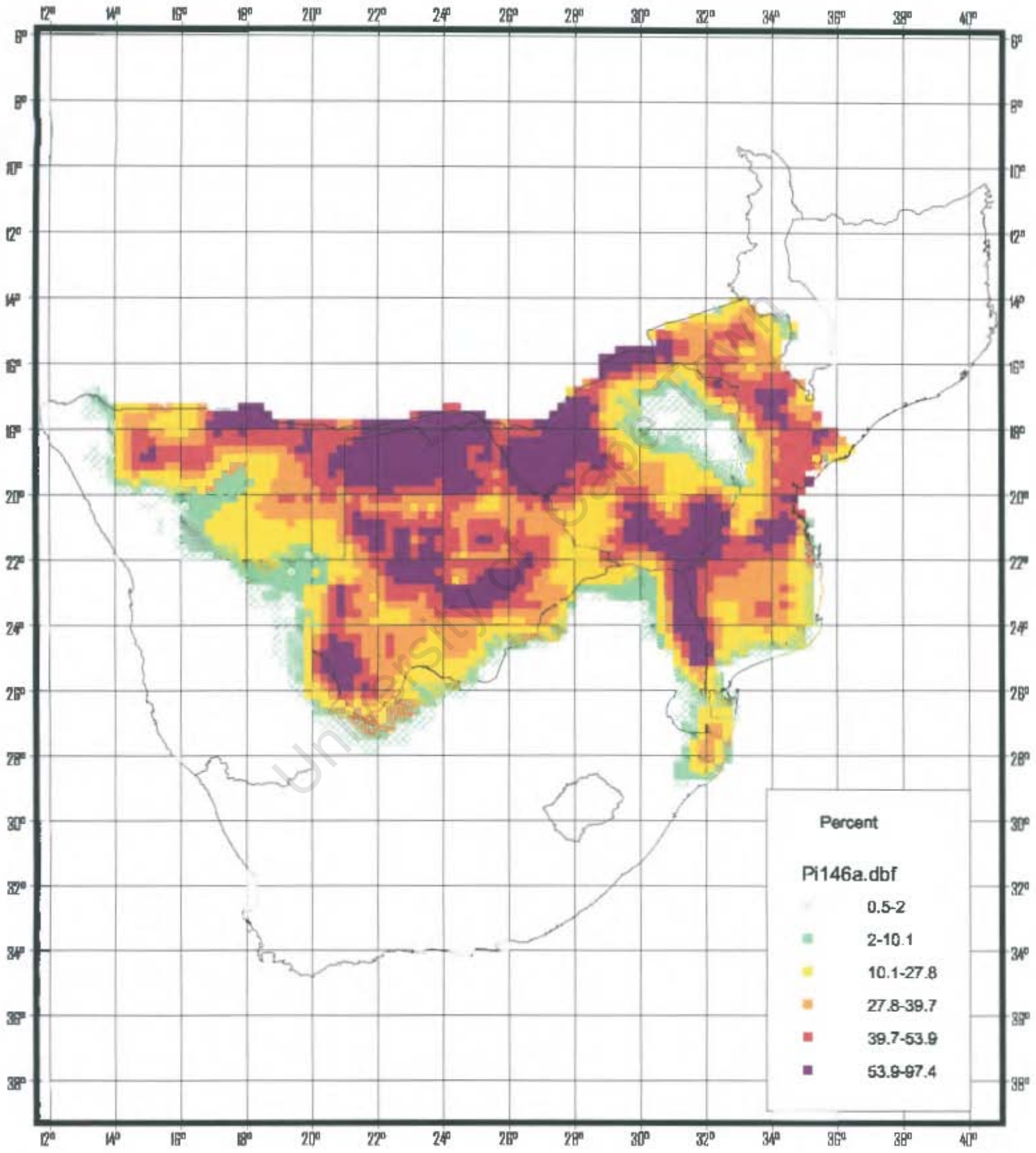


Figure 33b : Bateleur- Gradient distribution based on four-parameter model for blocks of 25 grid cells

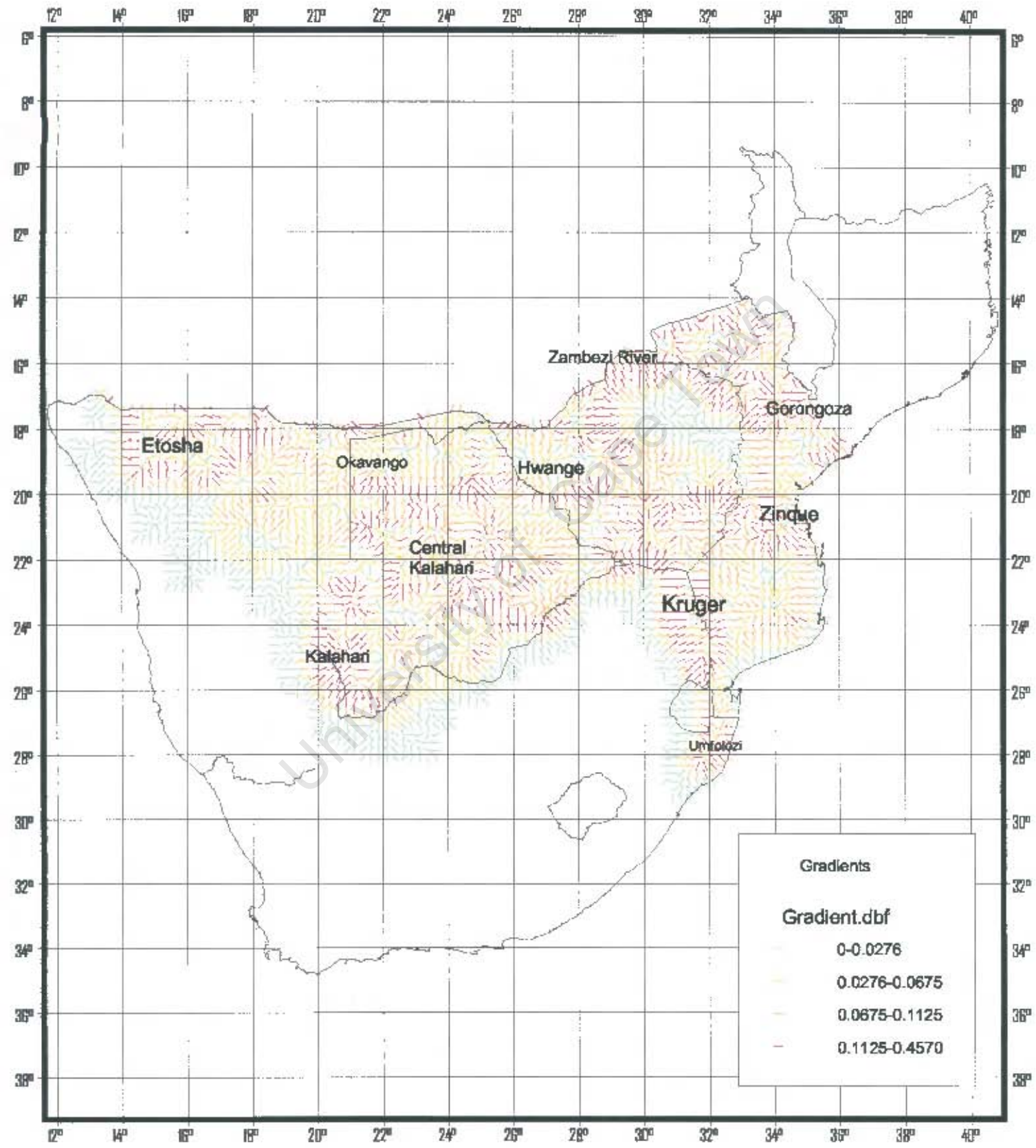


Figure 34: Distribution of sum of gradients for 12 selected species (Table 1)

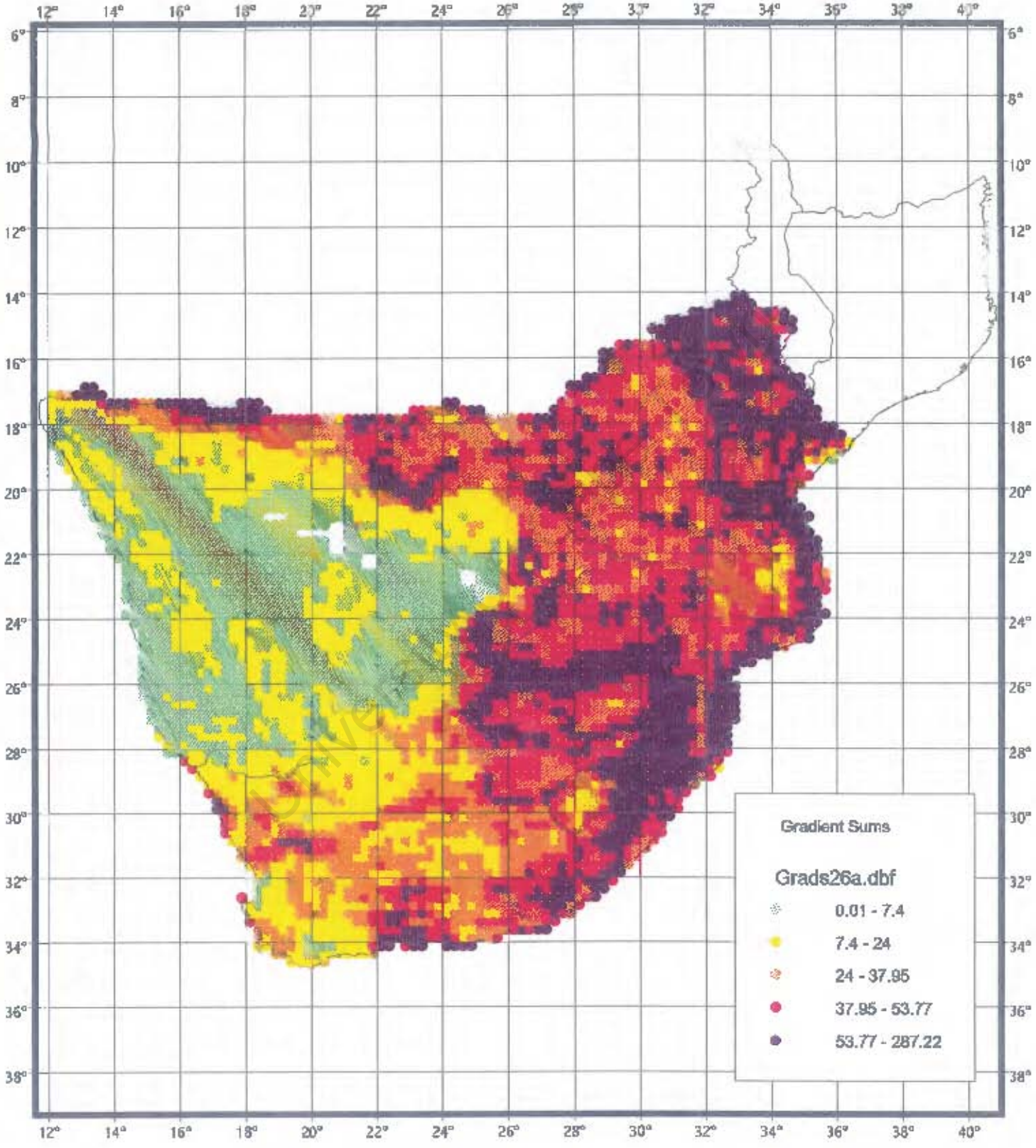


Figure 35: Distribution of top 10% of directional gradient sums for 12 selected species (Table 1)

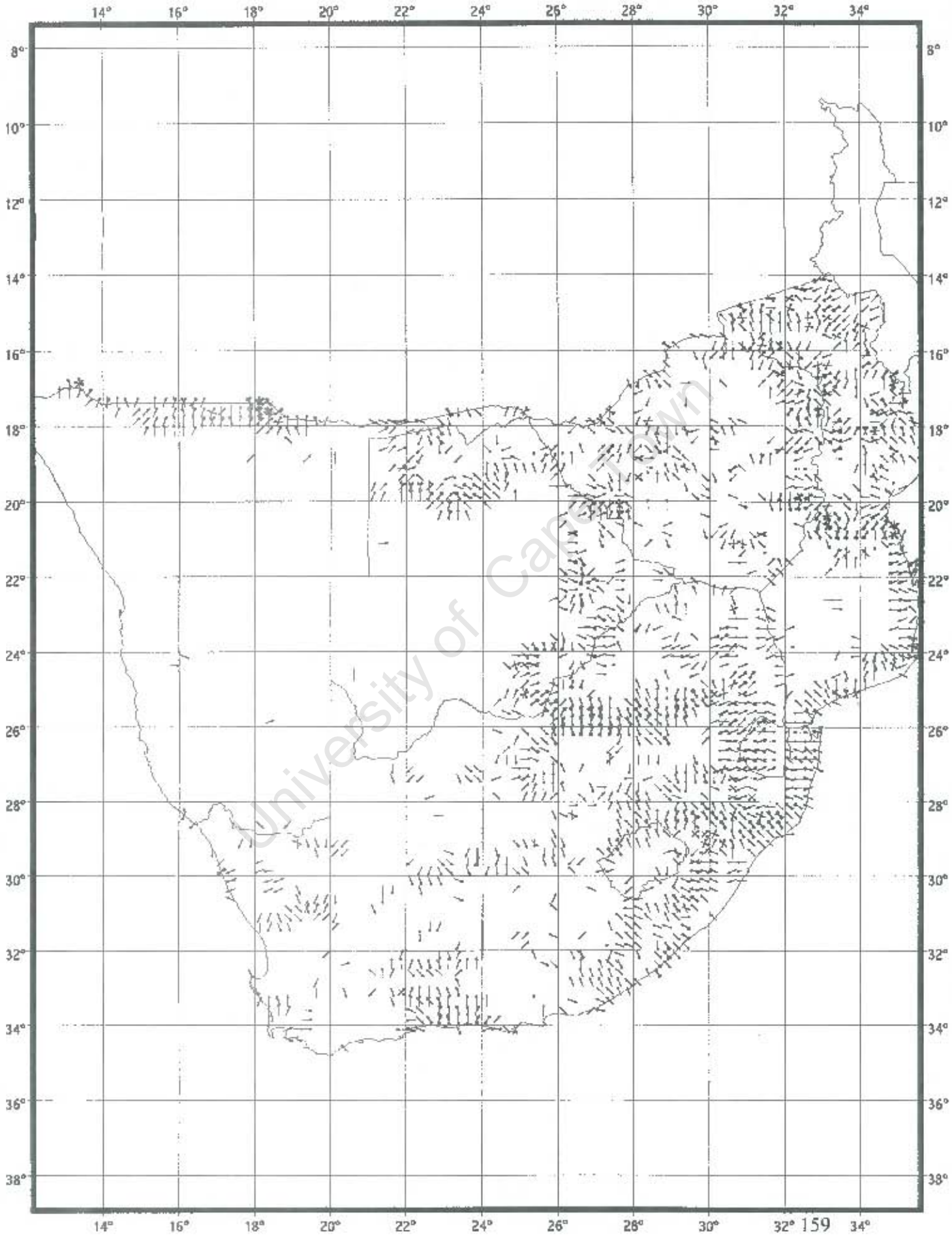


Figure 36: Distribution of median gradients for 12 selected species (Table 1)

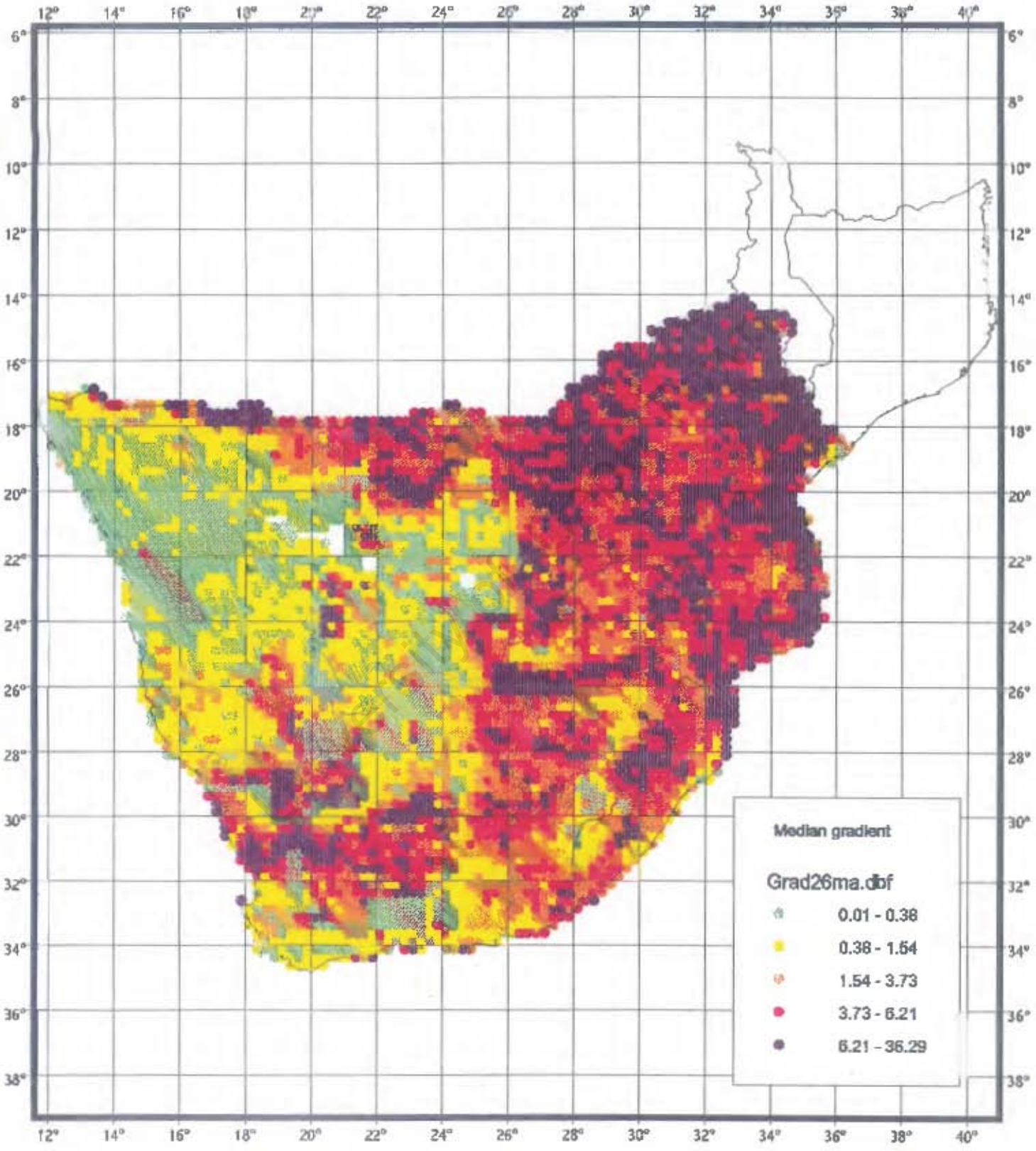


Figure 37: Distribution of sum of gradients for all terrestrial species

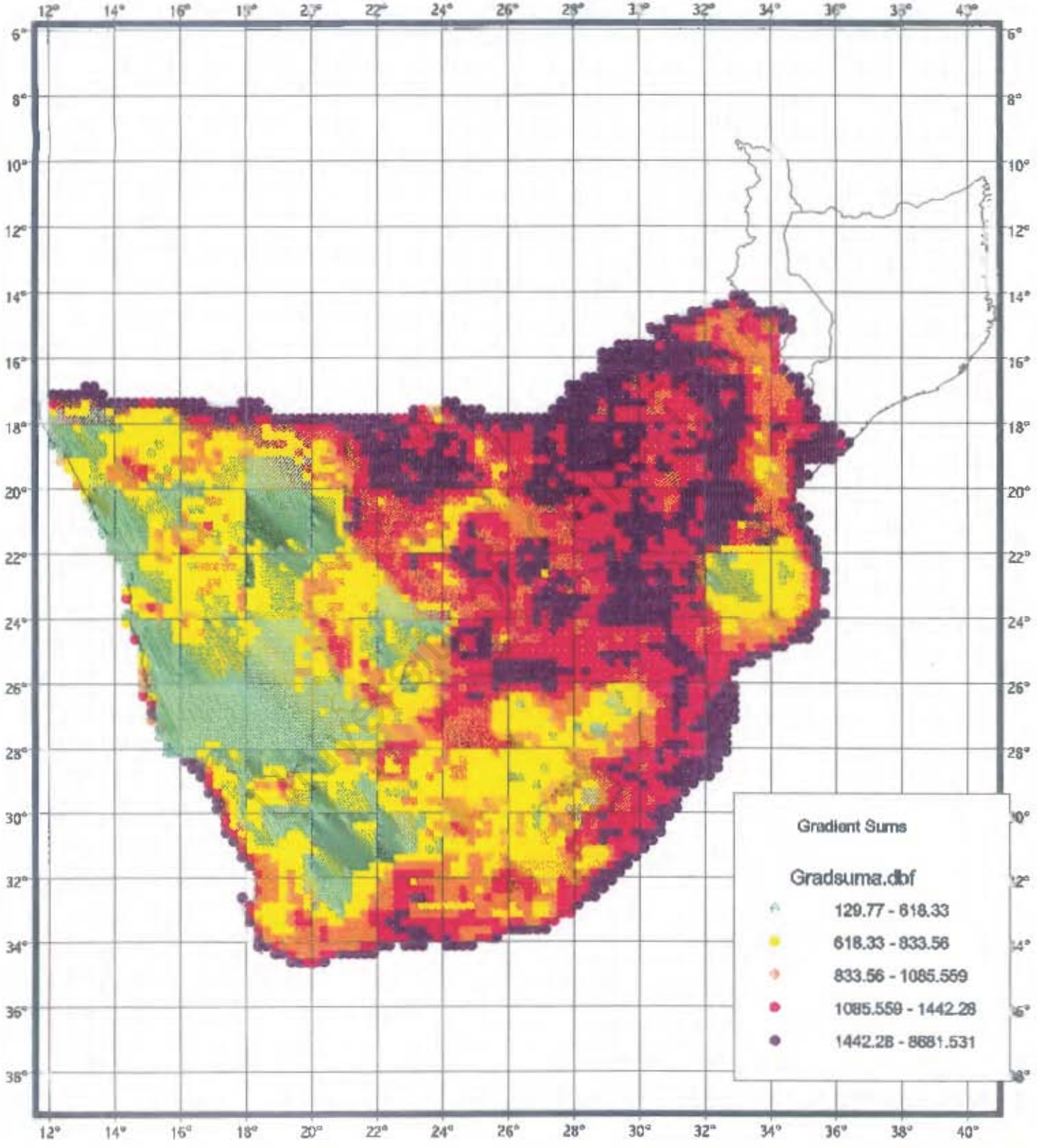


Figure 38: Distribution of top 10% of directional gradient sums for all terrestrial species

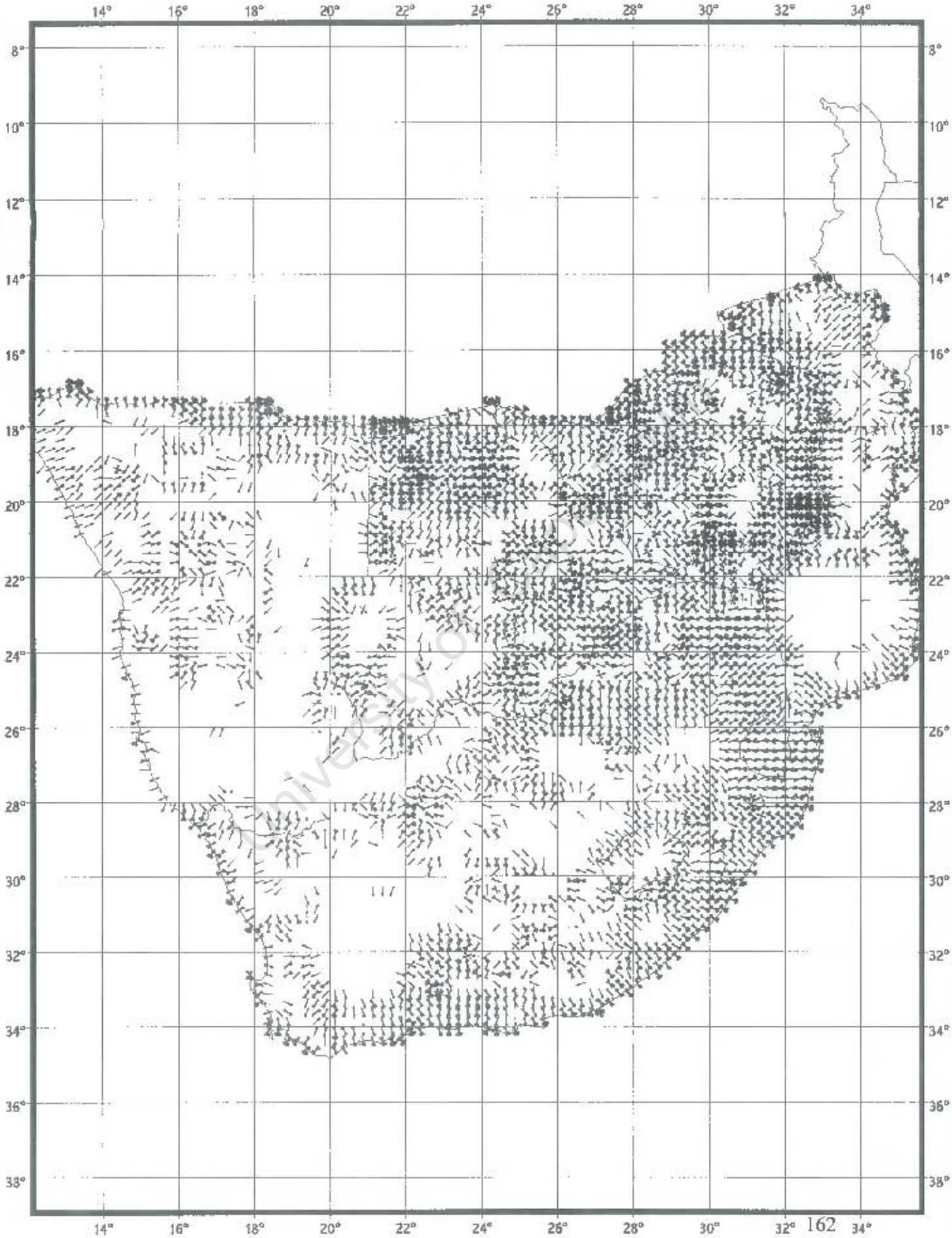


Figure 39: Distribution of median gradients for all terrestrial species

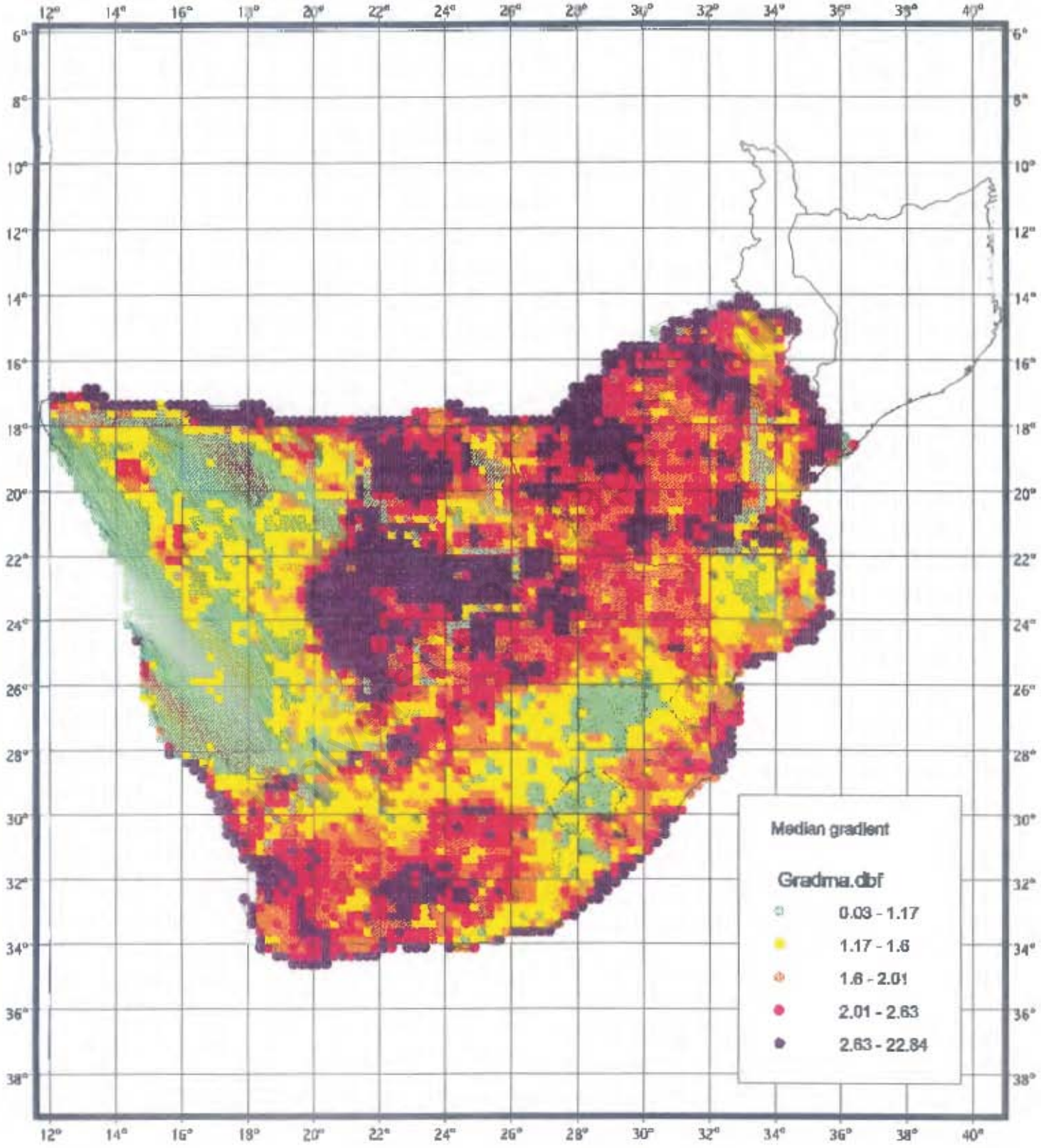


Figure 40: Distribution of top 10% of median directional gradients for all terrestrial species

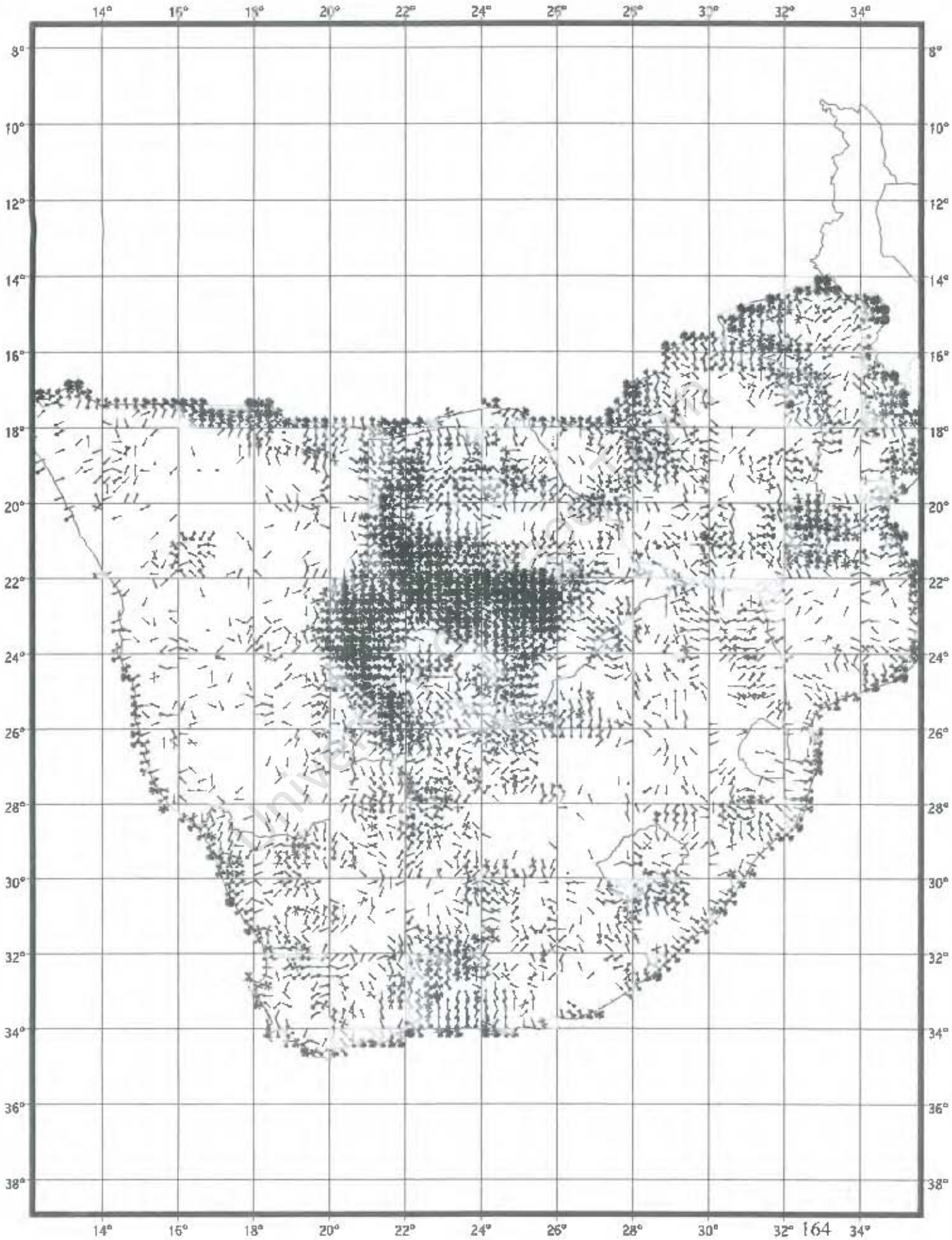
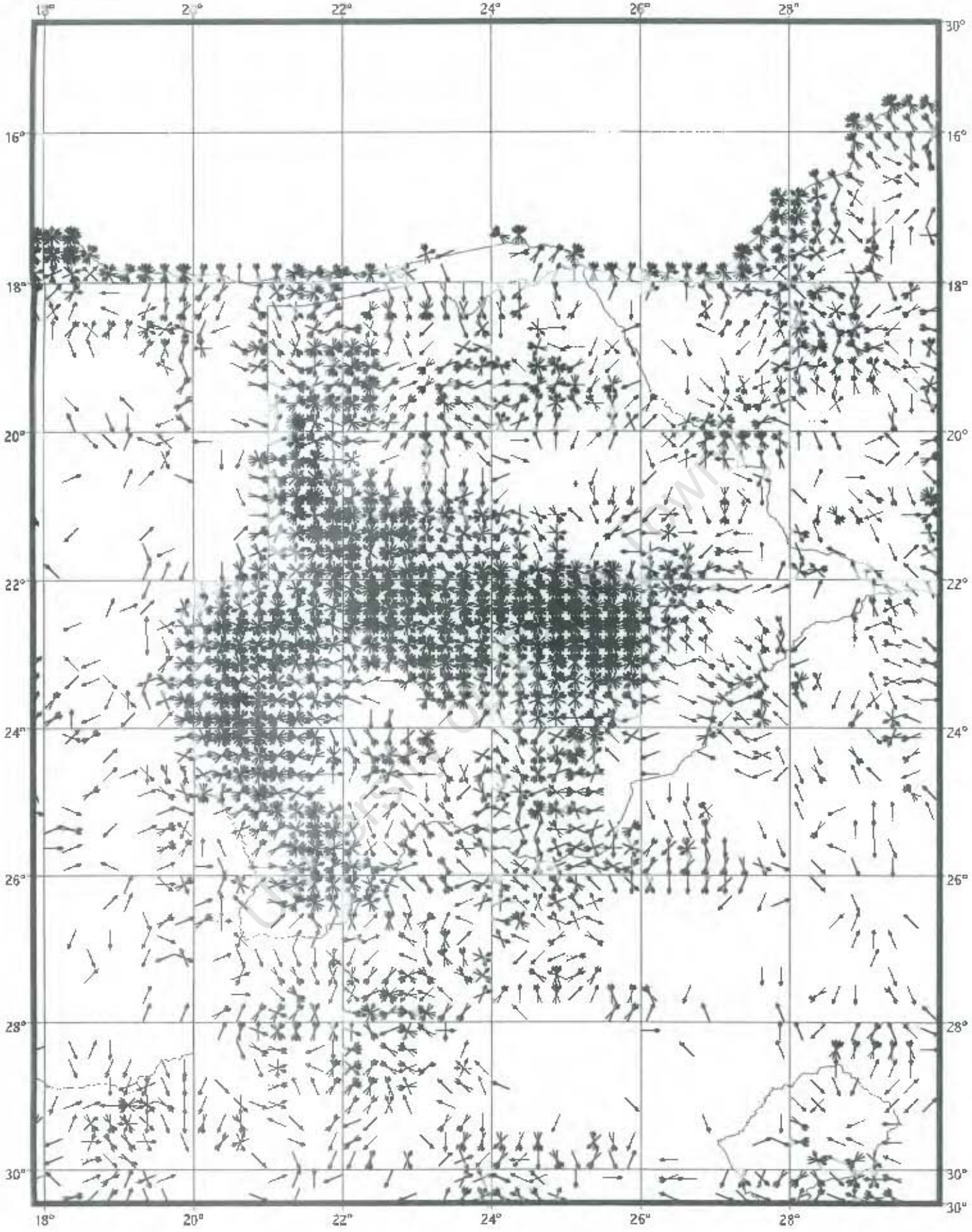


Figure 41: Distribution of top 10% of median directional gradients for all terrestrial species in Botswana



Chapter 6

An Overall Approach to the Analysis of Biogeographical Patterns in the Distributions of Southern African Bird Species.

F. LITTLE

Avian Demography Unit, Department of Statistical Sciences, University of Cape Town, Rondebosch, 7701 South Africa

Data Hierarchy

The impetus for this thesis was the availability of a different kind of data for bird species in southern Africa. Traditionally, biogeographical information for bird (and other ecological) species has been in the form of presence-absence data. The Southern African Bird Atlas Project (Harrison *et al.* 1997a, b) and the Mozambique Bird Atlas Project (Parker 1999) generated data on reporting rates for birds that takes into account the likelihood of species detection in a given area. The observed reporting rate for a species in a given grid cell corresponds to the proportion of checklists on which the species was recorded. However, these data contain sampling errors as discussed in Chapter 2. Through the use of a regression model (see Chapter 2) we replace these observed reporting rates with smoothed detection probabilities that give a more accurate reflection of the likelihood of observing a species in a given area. These smoothed detection probabilities form the basis of the species-specific distribution maps in this thesis.

When we wish to combine information for several species, we need to take into account the differential abundance of these species. Some species have large reporting rates, and others have small reporting rates. This artefact needs to be eliminated. To this end we replace the actual detection probabilities with their decile values as described in Chapter 4. These detection probability deciles are derived by dividing the species-specific ranges of detection probabilities into 10 intervals with equal numbers of probabilities in each interval. The actual probabilities are then replaced by one of the 10 integers between one and 10, divided by 10 to transform them back to the probability range. A value of zero will indicate species absence. These detection probability deciles reflect the relative probability of detecting a species in a given grid cell in such a way that the cores of the distribution of each species are treated equitably. They are dependent on the study are and ideally the study area should be large enough to include complete distributions for the chosen species, from core areas with the largest reporting rates to pheripheral regions with small reporting rates. The approach is not valid if the study area is so small that species have roughly uniform reporting rates throughout the area.

The range of 11 detection probability deciles can be converted to a binary outcome by grouping all non zero values together. We are then essentially reducing the number of intervals into which to subdivide the non zero range to one. This is equivalent to the presence-absence approach. In addition, the use of detection probabilities gives us the means by which to define "presence". Instead of considering all non zero probabilities, we can look at probabilities in excess of a large non zero probability cut-off value.

Methodology

Our choice of analytical methods and summary measures depended on the characteristics of the data and the scale of the problem. We motivate the choice of analytical techniques and point out their advantages.

The Choice of Smoothing Model

The varying number of checklists for each grid cell, that form the denominators for calculating the reporting rates, suggests that our data follow a binomial distribution with error variances dependent on the true probabilities of detection and on the varying sample sizes. The binomial nature of the data was the prime motivation for using a generalised linear model with a logistic link function to generate smoothed detection probabilities. As discussed in Chapter 2, not all the assumptions of the binomial distribution were valid for our data. In particular, the problem of non-constant variance lead to overdispersion but did not affect the predicted probabilities of detection. We were not interested in the statistical inference but in using the logistic model as a tool for generating smoothed surfaces.

Several other approaches to smoothing spatial data are available and discussions and comparisons of these can be found in Cressie (1993), Burrough & McDonnell (1998) and McNeill (1991). Erni (1998) evaluated previous approaches to smoothing and their applicability to binomial type data as motivation for developing the smoothing model used in Chapter 2. The most well-known smoothing methods include the collection of methods known as "kriging". These methods model the data as a stochastic process with a covariance matrix that assumes stationarity. Standard kriging methods assume constant sample sizes. McNeill (1991) tried to adapt kriging techniques to deal with binomial data with varying sample sizes. However, her smoothing approach assumed uniform detection probabilities over the entire species range. Erni (1998) demonstrated that this global approach is inappropriate because it leads to oversmoothing, especially in areas with small sample sizes.

Modern methods of Image Analysis refer to obtaining data on natural resource inventories through remote sensing, using satellites or high-flying aircraft (Cressie 1993). The continuous images obtained consist of the intensity of radiation in different bands of the electromagnetic spectrum. These images are then digitized to a finite set of colours on a rectangular lattice to form a two-dimensional array. Bayesian statistical methods are used to restore these digitized data back to continuous representations of the distribution of natural resources. It is difficult to see how this kind of data collection can be feasible for species so small in size as most of the birds. Furthermore, the methods used for image enhancement are not appropriate for binomial type data.

Our choice of smoothing method takes account of the varying sample sizes through the underlying binomial assumption. We model the spatial auto-correlation explicitly by including the directional coordinates in the model. We have illustrated how this smoothing algorithm reduces observer and sampling error and provides interpolated rates in areas of sparse sampling, thus resulting in more accurate and clearer representations of the true probability of species detection for species with a reasonable degree of continuity in their distributions. We control the degree of smoothing mainly through our choice of block size that defines the neighbourhoods. However, we have introduced a simple adjustment to our smoothing approach for species with more fragmented distributions by taking a weighted average between observed and modelled rates based on the degree of fragmentation and the extent of the sampling. Our model is simple to implement and allows us to cope with the very large estimation problem of generating smoothed detection probabilities for over 900 species at more than 5000 point locations.

The aim of this thesis was not to develop a new method of smoothing. Rather we implemented and extended a method developed by Erni(1998) and used this smoothing model and the resulting detection probabilities for our research into the derivation of measures of species richness and species endemism and the calculation of species gradients. The measures of species richness and species endemism could be applied to the observed reporting rates or to smoothed rates generated by any other smoothing method. The gradient analysis, on

the other hand, directly derives from our choice of smoothing model and provides further motivation for using this approach.

Summary Measures for Species Richness and Species Endemism

To incorporate the smoothed probabilities of detection into measures of species richness and endemism, we transformed them into the detection probability deciles as described above. These detection probability deciles then replaced the more uniform values of one or zero which traditionally indicate species presence or absence. The transformed measures of species richness and species endemism are thus just extensions of the measures based on presence-absence data. However, the resulting measures contain additional information on the underlying patterns of species distributions and are based on the areas of occupancy of species rather than their range of occurrence (Van Jaarsveld *et al.* 1998). The advantage of the measures based on detection probability deciles is that they focus on the core of the distributions of species and down-weight the peripheral edges where relatively small detection probabilities are indicative of abundance levels that may be too low to guarantee future survival of the species. This does assume that the study area is large enough to include complete distributions and a large degree of variability in reporting rates for all species.

The use of detection probability deciles when deriving a measure of species richness means that not only can we determine overall species richness as a simple count of the number of species present in a given area, but we can also make species richness reflect the relative likelihood of species detection in a given area. We can home in on explicitly and objectively defined core-ranges for species and identify areas where a significant number of species occur in the core of their distributions. The differences between the measures of endemism based on detection probability deciles and the measures of endemism based on presence-absence data are

(1) for presence-absence data the numerator stays constant at 1 for those species that contribute towards the sum, while when using a range of detection probability deciles, the numerator varies between 0.1 and 1.0.

(2) using the sum of the detection probability deciles to measure the extent of the species range shrinks the denominator because it downweights the cells with relatively small detection probabilities. These two features combined have the effect of narrowing the species distributions and concentrating on the grid cells with large relative probabilities within the range for that species.

The detection probability deciles thus enable us to locate the hotspots of narrow endemism in areas where species with resticted distributions have the maximum chance of continued survival. The summations based on differential relative detection probabilities lead to a more sensitive ranking of different areas with respect to endemism and as such they highlight areas that may have been missed by the uniform presence-absence approach.

The measures of species richness and narrow endemism are simple weighted summations based on relative probabilities of detection. We map the distributions of these summations and observe the distributional patterns that emerge. Our methods contain little element of subjectivity. We do not use any of the multivariate clustering techniques to classify species into groups that will identify distinct and homogenous avifaunas. These techniques generally require subjective decisions regarding stopping rules, the number of distinct groups, the number of species in each group and the degree of similarity between species in a group (De Klerk 1998).

Measuring Gradients of Species Distributions

The use of a mathematical model to estimate smoothed probability distributions for species, enabled us to use the theory of directional derivatives to calculate gradients for the detection probability surfaces for species. In

particular, it opened up the possibility of defining individual species gradients, a concept that was undefined for distributions based on presence-absence data. The gradients accurately reflect the relative degree of change among detection probabilities for a species and identify both the edges of the species distributions and changes among detection probabilities within the overall range of occurrence for the species. In addition, we can determine the direction in which these large changes occur.

We combined the gradients for all species in several different ways. Large values for the overall sums of the gradients indicate areas where there are large changes in the area of occupancy for many species. These areas will reflect both species richness gradients and changes in species composition. In general, they are areas of large fluctuation in species composition. On the other hand, small values for the overall gradient sums indicate areas of relative stability. The directional gradient sums distinguish between areas where the changes in species detection probability distributions occur in isolated directions thus indicating ecological transition zones, and areas of random fluctuation indicative of species fragmentation. The distribution of median gradients removes the species richness component from measurements of species gradients and focuses on areas where gradients are steep regardless of the underlying species richness. We calculated detection probability gradients and not abundance gradients for which the data were not available. Though no simple mathematical relationships exist between detection probability gradients and abundance gradients, and especially not between the accumulation of these gradients over many species, the distributions of the individual species detection probability gradients and the combined gradients for groups of species revealed known patterns in species abundance distributions.

The gradients and gradient sums thus enable us to measure the "strength" and "breadth" (Williams 1999) of zones of turnover for individual species and groups of species. Both Chapin (1923) and Moreau (1966) expressed their reservations about regional boundary lines generated by classification techniques that represent abrupt changes for a process that is in actual fact much more gradual. Williams (1999) recognised this and derived neighbourhood turnover indices to measure the strength and breadth of areas of transition. The availability of reporting rate data allowed us to overcome many of the potential shortcomings of his analyses. These included the problems associated with data interpolation, the impact of a large degree of fragmentation and the lack of information on the variations in species abundance. Our smoothing model is a totally objective tool for data interpolation which generates detection probabilities for empty grid cells. Our model assumes a reasonable degree of continuity in the underlying species distribution but makes no use of subjective knowledge about habitat preferences. Hence it does not suffer from the potential problem of circularity where habitat information is used to generate distributional information for species that are then used in algorithms aimed at identifying avifaunal groupings. Our smoothing model decreases the degree of fragmentation inherent in observed reporting rates. In addition, the directional gradients are able to distinguish areas of high turnover due to fragmentation from areas where the turnover is due to species enrichments or species replacement. Most importantly, the use of reporting rate data enables us to take into account the variations in species distributions within their larger ranges of occurrence.

The increased sensitivy of methods based on reporting rate data gives additional insight into the patterns of distribution of species that will enable biogeographers and conservationists to make more informed decisions regarding the placement of conservation areas and the mangement of factors that influence the chances of sustainable future survival.

Contribution to Science

In this thesis we have not so much derived new statistical theories, as adapted existing techniques to deal with the changed and improved data on bird species distributions generated by the bird atlas projects. One of the large challenges was to cope with the scale of the problem. We had to find an efficient smoothing method that could generate smoothed detection probabilities for a large number of bird species in more than

5000 grid cells. We derived simple summary measures of species richness and species endemism that are easy to compute and easy to interpret but that also contain more accurate and sensitive accounts of these concepts. We defined and computed individual species gradients directly from the mathematical models that generate the smoothed detection probabilities. Simple summations of these species gradients generated information on overall changes in combined species distributions without requiring the computing-intensive techniques of existing neighbourhood heterogeneity and neighbourhood segregation techniques (Williams *et al.* 1999).

Our contribution has thus not been in the form of new and innovative techniques. We have shown how to incorporate reporting rate data in the analysis of biogeographical distribution patterns and how these improved data add substantial insight into the patterns of distribution of bird species in southern Africa - insight that was simply not available from previous analyses based on presence-absence data. Our research confirms that the results of an analysis are only as good as the data on which they are based. We have provided the analytical tools, as requested by Van Jaarsveld *et al.* (1998) that will enable biogeographers to make full use of the improved data in their efforts to address the biodiversity crisis. Our major achievement has been to show how these techniques open up all manner of opportunities to biogeographers to use the "abundance" component of the bird atlas data better.

References

Burrough PA, McDonnell RA 1998. Principles of Geographical Information Systems. Oxford University Press 1998.

Chapin JP 1923. Ecological aspects of bird distributions in tropical Africa. American Nat. 72: 106-125.

Cressie NAC 1993. Statistics for Spatial Data. Johyn Wiley & Sons. New York.

De Klerk HM 1998. Biogeography and Conservation of Terrestrial Afrotropical Birds. Unpublished Ph.D. thesis, University of Cape Town.

Erni B 1998. Analysis of Distribution Maps from the Bird Atlas Data. Unpublished M.Sc thesis, University of Cape Town.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) 1997a. The Atlas of Southern African Birds. Vol 1: Non-passerines. BirdLife South Africa, Johannesburg.

Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) (1997b) The Atlas of Southern African Birds. Vol 2: Passerines. BirdLife South Africa, Johannesburg.

Lawes MJ, Piper SE 1998. There is less to binary maps than meets the eye: The use of species distribution data in the southern African sub-region. South African Journal of Science 94, 207-210.

McNeill L 1991. Interpolation and smoothing of binomial data for the southern African bird atlas project. South African Statistical Journal 25, 129-136.

Moreau RE 1966. The bird faunas of Africa and its Islands. Academic Press, London.

Parker V 1999. The Atlas of the Birds of Sul do Save, Southern Mozambique. Avian Demography Unit and Endangered Wildlife Trust, Cape Town and Johannesburg.

Van Jaarsveld AS, Gaston KJ, Chown SL, Freitag S 1998. Throwing biodiversity out with the binary data? South African Journal of Science 94, 210-214.

Williams PH, De Klerk HM, Crowe TM 1999. Interpreting biogeographical boundaries among Afrotropical birds: spatial patterns in richness gradients and species replacement. Journal of Biogeography 26: 459-474.