# Biodiversity Observations

http://bo.adu.org.za

**An electronic journal published by the Animal Demography Unit at the University of Cape Town**



## PENTAD-SCALE DISTRIBUTION MAPS FOR BIRD ATLAS DATA

**Les G Underhill & Michael Brooks**

Recommended citation format:
Underhill LG, Brooks M 2016. Pentad-scale distribution maps for bird atlas data. Biodiversity Observations 7.52: 1–8.

URL: http://bo.adu.org.za/content.php?id=245

Published online: 22 August 2016

**– ISSN 2219-0341 –**

PROJECT REPORT

# PENTAD-SCALE DISTRIBUTION MAPS FOR BIRD ATLAS DATA

*Les G Underhill & Michael Brooks*

Animal Demography Unit, Department of Biological Sciences, University of Cape Town, Rondebosch 7701, South Africa

Email: les.underhill@uct.ac.za

The primary purpose of a bird atlas project is to generate accurate maps of the distributions of bird species, based on actual observations (Harrison et al. 1997a, b, Dunn & Watson 2008, Underhill 2016). This was inevitably also one of the main objectives of the Second Southern African Bird Atlas Project (SABAP2) when it commenced in July 2007 (Harrison et al. 2008). At the start of SABAP2, the format of the distribution maps which would be produced by the project was totally unclear. Everything hinged on the levels of participation which the project would achieve and also on the intensity of geographical coverage.

Initially, the area for SABAP2 was South Africa, Lesotho and Swaziland, and the grid was a five minute grid, generating 17,000 pentads, each with dimensions 9.2 km north to south, and c. 8.3 km east to west, depending on how far south the pentad is (Underhill 2016). The SABAP2 protocol has also been formally adopted in Namibia, Zimbabwe, Kenya and Nigeria. Two types of checklists are submitted: (1) full-protocol checklists, with at least two hours of diligent fieldwork in the pentad, trying to confirm the presence of as many species as feasible, with species recorded in the order in which they were recorded; (2) ad hoc lists, consisting of lists of species recorded in the pentad, usually made over less than two hours, and without the imperative of trying to be comprehensive. For example, the ad hoc lists might be made at a single spot within the pentad, whereas the essence of a full-protocol list involves moving to as many habitats as feasible, trying to see as many species as possible within the pentad (Underhill 2016).

## Three considerations for mapping distributions using the SABAP2 protocol

The thinking about maps to display distributions based on SABAP2 data has evolved steadily. This paper presents the current (August 2016) proposal. It contains suggestions for the new series of distribution maps which will display the actual observations in the SABAP2 database. It is illustrated by three samples, two from the original SABAP2 region of South Africa, Lesotho and Swaziland, and one from Namibia, where full-protocol checklist coverage was only 10%.

1. The first consideration, which seemed impossible at the start of SABAP2, is to contemplate whether it is possible to create distribution maps on the pentad scale. The alternative would be to merge the data for the individual pentads into the familiar quarter degree grid cells, and plot the maps at this scale. The large volume of SABAP2 data and the excellent coverage, with full-protocol checklists, of almost every pentad over large regions, especially in the east and south of the SABAP2 region, means that it is feasible to generate distribution maps on the pentad scale. The gains in mapping precision are large; the alternative resolutions are 17,000 pentads for the three countries, in contrast with only 2000 quarter degree grid cells.

2. Many pentads have multiple full-protocol checklists, so it is feasible to generate maps representing reporting rates. The reporting rate ($r$) for a species in a pentad is defined as the proportion of full-protocol checklists submitted for a pentad ($n$) which have the species recorded on them (m), so $r=m/n$. Conceptually, as the number of checklists ($n$) increases, the reporting rate gets closer to its “true value”. This is self-evident with the tossing a coin and counting the number of heads; as

the number of times the coin is tossed increases, we intuitively feel that the proportion of heads should get closer and closer to 0.5, assuming heads and tails are equally likely. Likewise, as the number of checklists for a pentad increases, we anticipate that (other things being equal) the reporting rates of each species will slowly get more and more stable. Ultimately, they will converge on the true value of the reporting rate for the species in a pentad. ("Other things being equal" means that factors such as the conspicuousness of birds through the year does not change, the abundance of the species is neither increasing nor decreasing, etc). Reporting rates can only meaningfully be based on full-protocol checklists. Ad hoc lists which contain small numbers of species would tend to bias the reporting rates to be smaller than they ought to be.

Several studies have linked reporting rate of a species in a pentad to the relative abundance of the species in the pentad (e.g. Robertson et al. 1995, Kemp et al. 2001). The best of these studies was by Griffioen (2001), who rediscovered the mathematical relationship between reporting rate and relative abundance; the original research had been done by Nachman (1981). The inference is that reporting rate does contain information related to relative abundance, and the atlas distribution maps ought to display this information.

3. With hindsight, we made a tactical error the distribution maps produced for SABAP1 (Harrison et al. 1997a, b); we mapped reporting rates even for grid cells with small numbers of checklists. When $n$ is small, only small numbers of different values for the reporting rate are possible. If $n$=1, then the reporting rate $r$ can only be 0 or 1, the species is either recorded or not recorded. If $n$=2, three are possible for $r$ : 0, 0.5 and 1. For $n$=3, the values are 0, 0.33, 0.67 and 1. For $n$=4, the five values for $r$ are 0, 0.25, 0.5, 0.75 and 1. With $n$ checklists, the number of possible $r$-values is $n$+1. Clearly, large numbers of check-lists per pentad provide a good estimate of the reporting rate in the pentad. Our opinion, based on practical experience, is that the minimum number of checklists at which reporting rates become meaningful is $n$=4. SABAP2 is using the expression that four checklists represents foundational coverage for a pentad. However, in broad brush terms, it is not until the number of checklists reaches double figures, that the species list for the pentad includes almost all the species that regularly occur in the pentad (Harrison & Martinez 1995).

The consequence for the SABAP1 maps of plotting reporting rates even when there were only a few checklists per gridcell, is that the maps for many species look like a random chessboard. This was particularly noticeable over parts of Namibia, Zimbabwe and the Northern Cape (Harrison et al. 1997a, b). At the extreme, if there were only single checklists in an area, the reporting rates could only be zero or one. The resulting pattern on the distribution maps consisted of grid cells with highest possible reporting rate shade (when $r$=1) interspersed with unshaded grid cells, where the species was not recorded on the only atlas checklist submitted for the grid cell.

The inference from these three considerations is that maps on a pentad scale are desirable, and that reporting rates should be shown to represent approximate relative abundance, but reporting rates should only be shown for pentads which meet some minimum threshold for coverage, currently taken as $n$=4 full-protocol checklists. This information motivated the development of the new generation of distribution maps, unveiled in this paper (Figures 1 to 3). The new maps are hybrids, showing presence (grey shade) or absence (white dot) in areas with fewer than four checklists, and reporting rates, in seven shades, white if the species was not recorded and six colours from yellow to dark blue dependant on reporting rate (Box 1).

If distribution maps are going to be produced for reproduction at small scales, it would probably be preferable to use a 10-minute grid (four pentads per gridcell) or a 15-minute grid (nine pentads per gridcell). Consideration also needs to be given to reducing the number of colours, from the six proposed here to two or three. The smaller the size of the map, the simpler the representation needs to be.

**Box 1. Interpretation guidelines for new generation SABAP2 distribution maps.**

1. The cells on the maps are pentads, 5 minutes of latitude by 5 minutes of longitude, 9.2 km north-south × c. 8.3 km east-west.

2. There are two shading systems, one for pentads with less than four full protocol checklists, and another for pentads with four or more full protocol checklists. The first system shows presence-absence, the second shows reporting rate.

3. If there are less than four checklists, there are three alternatives: turquoise = no data at all for the pentad; white circle = species not recorded, although there is some data for the pentad, consisting of between one record (incidental or ad hoc) and three full protocol checklists; grey = species demonstrated to be present in pentad. The white circles can be interpreted as possibly absent.

4. If there are four or more checklists, the reporting rates are represented in colour.

If the species has not been recorded, the entire pentad is shaded white, and the species is probably absent.

The reporting rates from the remaining pentads are sorted from smallest to largest, and split into six groups, which are as even in size as possible. The “cut points” for the groups vary with the species.

The pentads with the largest one-sixth of reporting rates are shaded dark blue, indicating the core of the range of the species. The next sixths are shaded light blue, then dark green, the light green, then orange and finally yellow, for the smallest one-sixth of reporting rates, where the species is most rarely recorded.

If a pentad has four or more checklists, and the species has only been recorded as an incidental or on an ad hoc list, then the pentad is shaded yellow.

The pentads shaded blue, either light or dark, represent the third of the range where reporting rates are largest, the pentads shaded green, either light or dark, show the middle third of reporting rates, and yellow-orange pentads represents the third of the range with the smallest reporting rates. Pentads shaded dark blue, light blue and dark green all have reporting rates above the median reporting rate for the species. Half of the pentads are shaded these three colours.

The reporting rate values for the five cutpoints are only of academic interest. They are deliberately not presented, because they are not comparable between species.

## Three examples

### *Blue Crane*

Most of the pentads in which Blue Cranes *Anthropoides paradiseus* had been recorded by August 2016 already had four or more checklists, and are represented in colour on the distribution map (Figure 1). The core of the range was clearly in the Overberg, the wheat growing area on the coastal plain to the east of Cape Town; this is demonstrated on the distribution map by the predominance of pentads shaded dark blue and light blue over this region. There was a second core to the range in the Swartand, the wheat-growing region north of Cape Town. There was a striking gap of c. 50 km between these two sections of the core, an area which is mostly mountainous terrain. The large expanses of white across the Kruger National Park and coastal KwaZulu-Natal point to the absence of Blue Cranes in these regions. There was a region of mainly yellow, orange and light green pentads in the KwaZulu-Natal Midlands, going northwest into the grasslands of the Free State and Mpumalanga. In the southeast of the Four Degrees region of Greater Gauteng, a cluster of coloured pentads is conspicuous against a background of white pentads. This region is known as the Devon Grasslands and was established as an IBA in 2014 (Marnewick et al. 2015). In the eastern Karoo, most of the range was shaded grey, indicating that the species was present in this region, but that there were insufficient checklists,

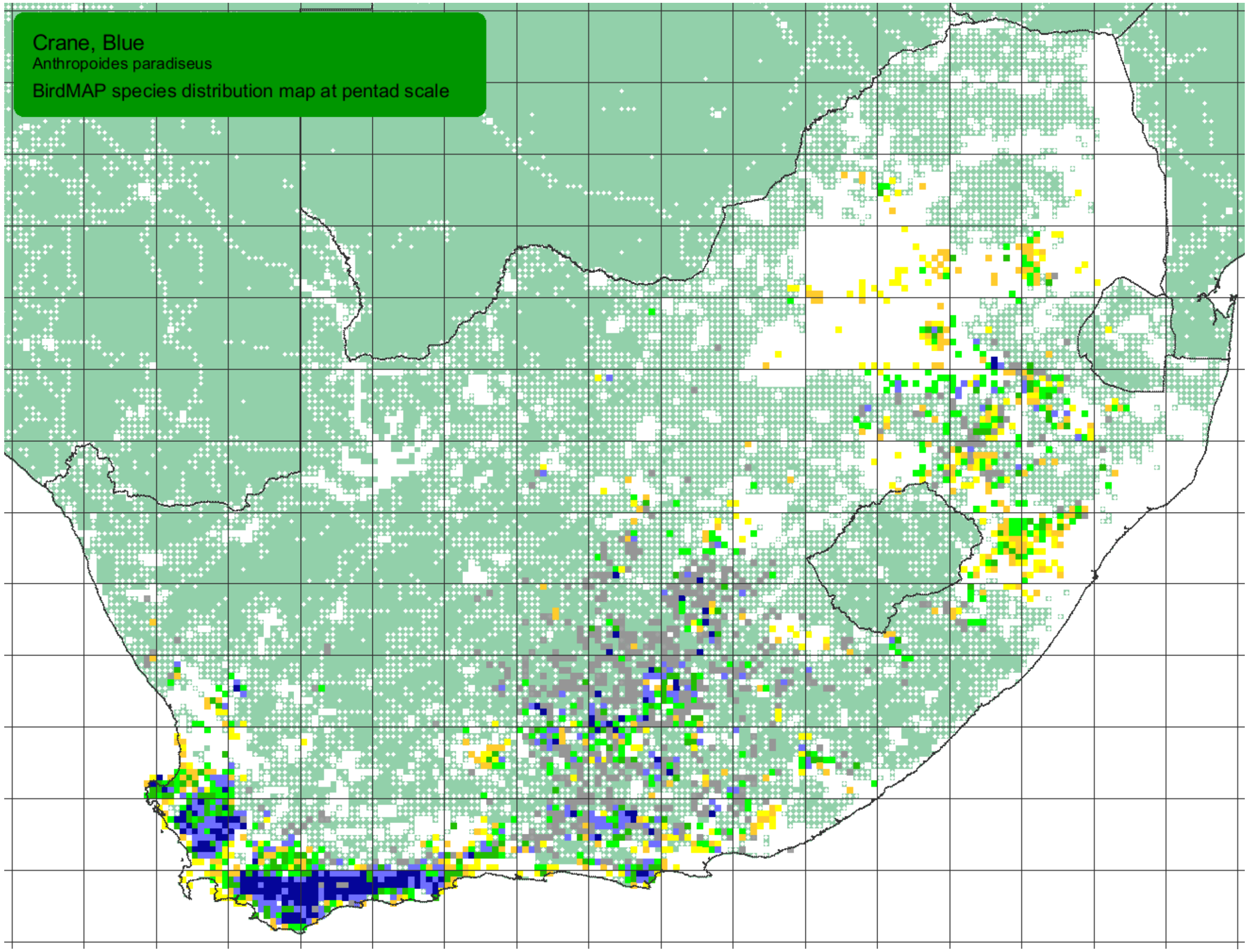


*Figure 1. Distribution map for the Blue Crane, on a pentad scale, August 2016. Box 1 explains the process for the interpretation of the map*

by August 2016, for the grey pentads to be able to compute a reliable reporting rate for them (Figure 1). This simply indicates that the Blue Crane was present in these pentads.

With the exception of the eastern Karoo, this example falls not far short of the target being aimed at with this mapping system. The objective is at least four checklists for each grid cell within the range of the species. Every pentad will then have a reporting rate which can be displayed in colour.

### *Red-billed Quelea*

The pentad distribution map for the Red-billed Quelea *Quelea quelea* is largely grey, especially across the Free State and Limpopo (Figure 2). This large area has too few checklists to be able to compute the reporting rate, but does probably contain much of the core of the distribution of this species, and will receive either dark or light blue shades when coverage gets to four pentads. Based on the subset of the range with four or more checklists, it appears that the core of the distribution of the Red-billed Quelea is mainly in the Free State. This map has several other interesting features: (1) the "hole" in the distribution in the conurbation of Pretoria-Johannesburg in Gauteng, shown as yellow with small reporting rates was not apparent in early distribution maps; (2) there is a new centre of distribution along the Orange River east and west of Upington in the Northern Cape; this was not evident during the late 1980s on the SABAP1 map (Mundy & Herremans

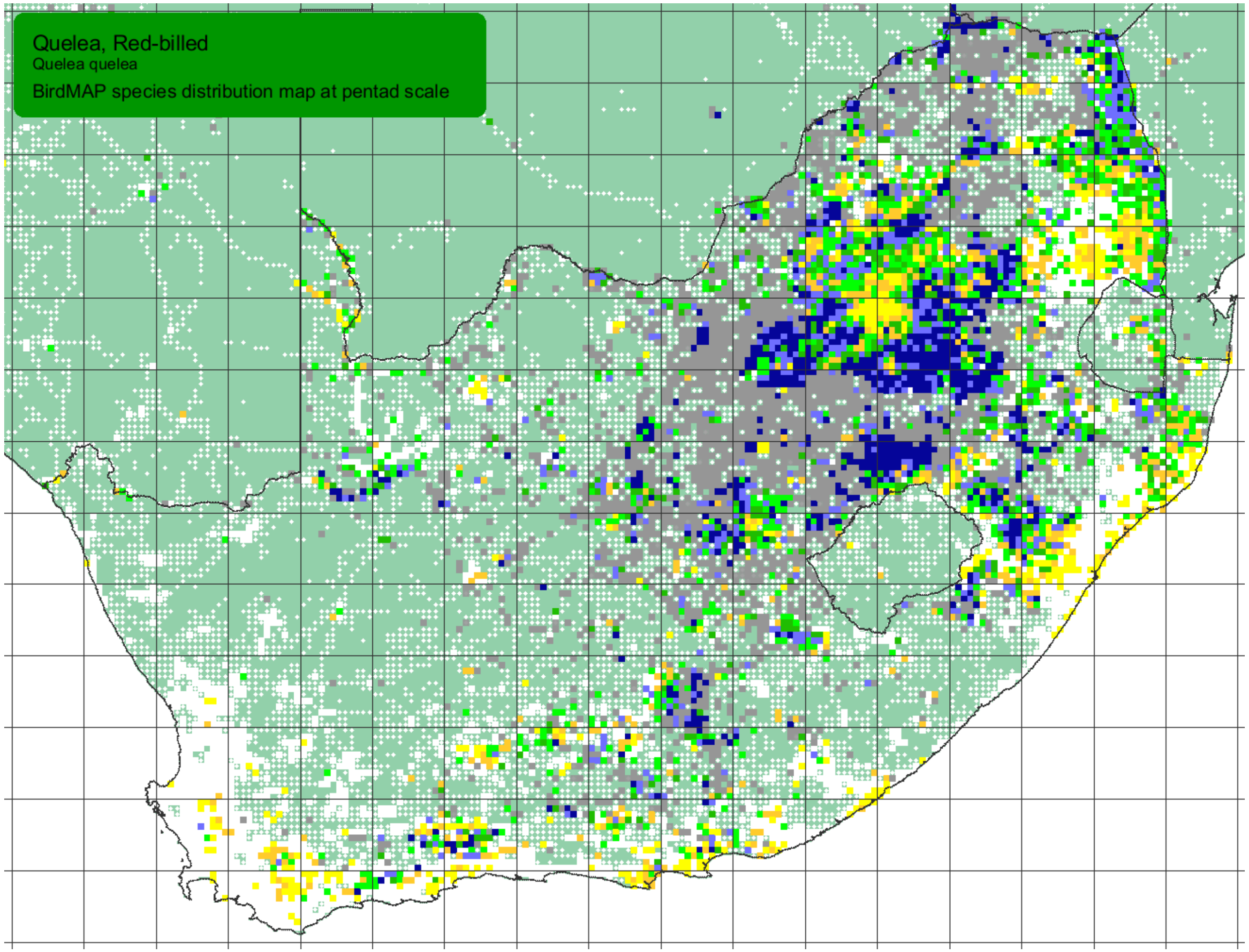


*Figure 2. Distribution map for the Red-billed Quelea, on a pentad scale, August 2016. Box 1 explains the process for the interpretation of the map*

1997); (3) a large-scale range extension into the Western Cape has taken place, such that in the Little Karoo there a several pentads with reporting rates in the upper one-sixth of reporting rates, and therefore shaded dark blue on the coverage map (Figure 2).

The distribution map for the Red-billed Quelea falls short of the target. Large numbers of pentads within the range of the species have fewer than four checklists. In these regions, the map downgrades itself to a display of presence (pentads shaded grey) and possible absence (pentads have small white dots) (Box 1).

***Cape Turtle Dove in Namibia***

The third example illustrates the start of the development of a distribution map at an early stage of an atlas project. There are 10617 pentads Namibia; in July 2017, 1102 (10.4%) of them had at least one full protocol checklist. Only 227 Namibian pentads (2.1%) had four or more checklists and therefore qualified for reporting rates in colour on species distribution maps. The total number of pentads in Namibia with data from all sources (full protocol checklists, ad hoc checklists and incidental) was 2744, 25.9%. Thus at this stage in the development of the bird atlas in Namibia, there is coverage for 15.5% more of the country than covered by full-protocol checklists. The species recorded in the most pentads in Namibia was Cape Turtle Dove *Streptopelia semitorquata*, in 1273 pentads (Figure 3). 768 of these records were in pentads with full-protocol lists, and an additional 505 were in pentads for which there were only ad hoc lists and incidental records.

This example illustrates the value of ad hoc checklists in helping to get the distribution maps started in a country. In the early phases of a project, there is great value in obtaining additional

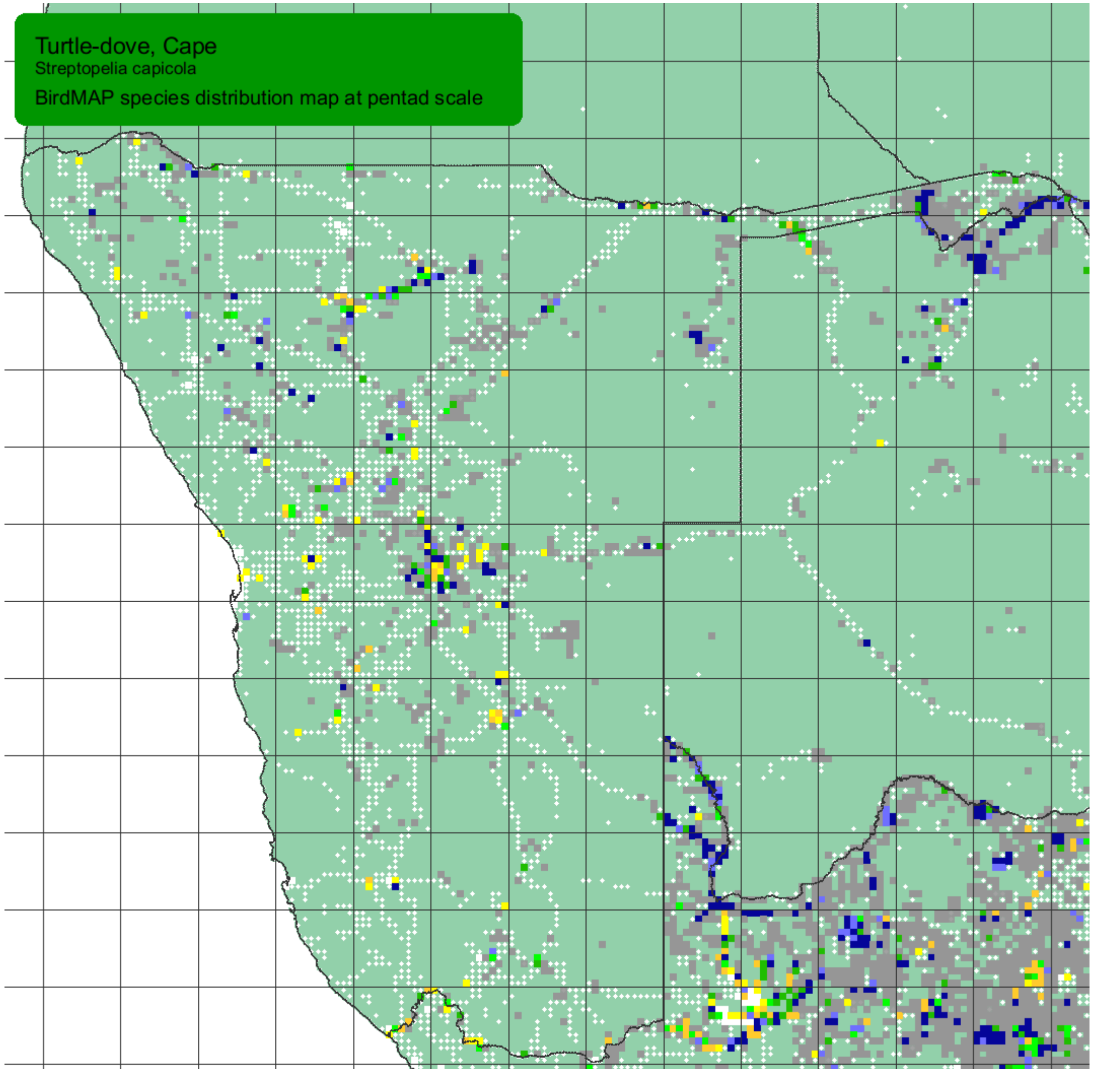


*Figure 3. Distribution map for the Cape Turtle-Dove, in Namibia,, on a pentad scale, August 2016. Box 1 explains the process for the interpretation of the map.*

data over and above full-protocol checklists in these contexts. In fact, every distribution record is valuable, and citizen scientists can make a substantive contribution to the project by making ad hoc lists for every pentad they visit, even though they are unable to spend the two hours required to make a full protocol checklist (Underhill 2016).

In Figure 3 (and in Figures 1 and 2), no distinction is made between records on full-protocol checklists and those from ad hoc checklists. Presence-absence mapping continues in a pentad until its gets its fourth full protocol checklist, and the reporting rate is displayed in colour (Box 1).

## Future developments

Gaps in coverage can potentially be filled by statistical modelling (Franklin 2009). There are families of methods to undertake species distribution modelling (SDM) and developing these models has been an industry keeping statisticians busy for two or three decades (Franklin 2009). There are limitations to the use of species distribution models, which were discussed by Carneiro et al. (2016), and in references in that paper. Carneiro et al. (2016) recommended that an investment should be made in long-term ecological research projects which monitor biodiversity. They also recommended that such projects, which are essentially atlas projects, should be undertaken prior to the short-term biodiversity assessments undertaken for EIAs at individual sites.

In other words, Carneiro et al. (2016) proposed that it is better to produce distribution maps from real data than from modelled data. The operating philosophy which underpins the fieldwork for SABAP2 is that gaps in coverage needs to be minimized.

But, inevitably there are gaps. Many pentads are inaccessible. The quandary is between "telling the truth", as was done in SABAP1 (Harrison et al. 1997a, b) and only presenting the actual observations, and doing some judicious interpolation and smoothing. In fact, this is not a point of discussion at all. It is necessary to do both to present the actual data, and to do appropriate species distribution modelling. The two approaches can be used to provide different insights.

Two of the key properties of the SABAP2 protocol (Underhill 2016) simplify the species distribution modelling process; one is spatial and the other is statistical. (1) The regular five minute grid system, generating pentads, provides a geometric framework which aids spatial analysis (McNeill 1994). (2) The statistical distribution which underpins the data is the binomial distribution, i.e. there are $n$ checklists for a pentad, and the species has been recorded on $r$ of them (Underhill 2016). Two analytical approaches, which exploit these two properties, can be used not only to fill gaps, but also to smooth reporting rates, are contained in the PhD theses of Lindsay McNeill and Francesca Little (McNeill 1994, Little 2003). Both approaches exploit the spatial autocorrelation present in the atlas data, and both can incorporate explanatory variables, such as habitat and climate.

## Acknowledgements

It is an amazing tribute to the citizen scientists who participate in SABAP2 that we can contemplate producing distribution maps at a pentad scale.

## References

**Carneiro LRdA, Lima AP, Machado RB, Magnusson WE** 2016. Limitations to the use of species-distribution models for Environmental-Impact Assessments in the Amazon. PLoS ONE 11(1): e0146543. doi:10.1371/journal.pone.0146543

**Dunn AM, Weston MA** 2008. A review of terrestrial bird atlases of the world and their application. Emu 108: 42–67.

**Franklin J** 2009. Mapping species distributions: spatial inference and prediction. Cambridge: Cambridge University Press.

**Griffioen P** 2001. Temporal changes in the distributions of bird species in eastern Australia. PhD 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. Johannesburg: BirdLife South Africa: i–cii, 1–785.

**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. Johannesburg: BirdLife South Africa: i–xvi, 1–732.

**Harrison JA, Martinez P** 1995. Measurement and mapping of avian diversity in southern Africa: implications for conservation planning. Ibis 137: 410–417.

**Harrison JA, Underhill LG, Barnard P** 2008. The seminal legacy of the Southern African Bird Atlas Project. South African Journal of Science 102: 82–84.

**Kemp AC, Herholdt JJ, White I, Harrison JA** 2001. Birds of the two largest national parks in South Africa: a method to generate estimate of population size for all species and assess their conservation ecology. South African Journal of Science 97: 393–403.

**Little, Francesca** 2003. The smooth is better than the rough: an exploitation of reporting rate information in southern African bird atlas data. PhD, University of Cape Town. Available online at http://open.uct.ac.za/bitstream/handle/11427/8592/thesis_sci_2003_little_f.pdf?sequence=1

**Marnewick MD, Retief EF, Theron NT, Wright DR, Anderson TA** 2015. Important Bird and Biodiversity Areas of South Africa. Johannesburg: BirdLife South Africa.

**McNeill, Lindsay** 1994. Topics in interpolation and smoothing of spatial data. PhD, University of Cape Town. Available online at http://open.uct.ac.za/bitstream/handle/11427/15969/thesis_sci_1994_mcneill_lindsay.pdf?sequence=1

**Mundy PJ, Herrmemans M** 1997. Redbilled Quelea *Quelea quelea*. In: Harrison JA, Allan DG, Underhill LG, Herremans M, Tree AJ, Parker V, Brown CJ (eds) The atlas of southern African birds. Vol. 2: Passerines. BirdLife South Africa, Johannesburg: 573–575.

**Nachman G** 1981. A mathematical model of the functional relationship between density and the spatial distribution of a population. Journal of Animal Ecology 50: 453–460.

**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** 2016. The fundamentals of the SABAP2 protocol. Biodiversity Observations 7.42: 1–12. Available online at http://bo.adu.org.za/content.php?id=235