

Environmental correlates of seasonal range dynamics of migratory and nomadic birds

Luvuyo Kani^{a,b,*}, David Maphisa^{a,c}, Res Altwegg^a

^a*Centre for Statistics in Ecology, Environment and Conservation, Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa,*

^b*Endangered Wildlife Trust, Department Name, Street Address, City, Postal Code*

^c*South African National Biodiversity Institute, Department Name, Street Address, City, Postal Code*

Abstract

Birds can respond to environmental variability through movement. Seasonally predictable fluctuations lead to migration while unpredictable variability leads to nomadism. Intermediate strategies also exist. We examined possible environmental drivers associated with arrival and departure of five migratory and six nomadic bird species in western South Africa, using dynamic occupancy models and data from the Southern African Bird Atlas Project. Dynamic occupancy models estimate site occupancy at the beginning of the study. Occupancy probabilities at later time steps are derived from directly estimated colonisation and extinction probabilities. Here, we use these models to estimate monthly changes in occupancy between 2014 and 2018, redefining colonisation as arrival and extinction as departure. We modelled arrival and departure as a function of changes and anomalies in average monthly rainfall, temperature and vegetation. Arrival and departure of obligate migrants were associated with seasonal changes in average rainfall, temperature and vegetation. In nomadic species, arrival was associated with seasonal changes in average rainfall, temperature and vegetation, while departure was associated with anomalies in rainfall, temperature and vegetation. The models successfully captured the generally known seasonal pattern in occupancy of all five obligate migrants, while no regular seasonal fluctuation in occupancy was evident among the nomadic species. Over the four-year period, no species showed a strong shift in seasonal range dynamics in both groups.

Keywords: keyword1, keyword2

*Corresponding author

Email addresses: luvuyo.o.kani@gmail.com (Luvuyo Kani), d.maphisa@sanbi.org.za (David Maphisa), res.altwegg@uct.ac.za (Res Altwegg)

1. Chapter 1: Seasonal range dynamics of migratory birds

1.1. Introduction

(Teitelbaum and Mueller, 2019): review of nomadism

Species range shifts, which have been evident in a wide range of marine and terrestrial taxa (Chen et al., 2011a; Przeslawski et al., 2012; Tingley et al., 2012; Howes et al., 2019), have resulted in renewed interest in factors that limit species' distributions (Gaston, 2009; Sexton et al., 2009; Pigot et al., 2010; Angert et al., 2011). Often, such shifts have been linked to recent climate change (Parmesan and Yohe, 2003; Chen et al., 2011a; Gottfried et al., 2012; Poloczanska et al., 2013) although it is recognized that there is a host of other factors, such as land-use change, biological invasions, atmospheric CO₂ increase and many more, that could trigger species range shifts (Sala, 2000; Jetz et al., 2007; Parmesan et al., 2011; Péron and Altwegg, 2015). In the Southern Hemisphere, relatively fewer studies on species range shifts have been conducted than in the Northern Hemisphere, where a general pattern of poleward and upward range shifts has been reported (Lenoir and Svenning, 2013, 2015). The movement of animals, whether through dispersal, migration or foraging, exists as a result of the potential benefits associated with the movement (Alexander, 1998; Shaw, 2020). Seasonality in abiotic factors influenced by the earth's climate system, for example, the cyclic fluctuation of temperature, rainfall and photoperiod (Pezulli et al., 2005) give rise to seasonal environmental conditions and events in many species of plant and animals (Sun, 2016). Fluxes in resource availability resulting from seasonality present opportunities for the exploitation of seasonal environments. In the case of deteriorating conditions, if movement facilitates an escape, then movement away from unfavourable conditions will be favoured (Shaw, 2020). The escape from unfavourable conditions in pursuit of relatively favourable ones results in occupied sites being vacated and previously vacant sites being occupied. Such movements have been hypothesized to have led to the evolution of seasonal migration as a strategy in response to seasonally varying conditions in migratory birds (Berthold, 2001; Rappole, 1995). Seasonal migration in birds is described as the regular return movement between breeding and non-breeding grounds (Newton, 2008; Rappole, 2013). This definition encompasses different patterns of seasonal movements mainly differing in regularity, distance and population level patterns resulting in different types of bird migration as described by Rappole (2013). Long distance migrants migrate between breeding and non-breeding grounds that are separated by thousands of kilometres while short distance migrants travel relatively shorter distances. In some species a certain portion of the population may migrate to non-breeding grounds while another may choose not to migrate from their breeding grounds in a type of migration called partial migration. Altitudinal migrants migrate to different elevations for their breeding and non-breeding seasons. Wanderers are migrants that have a specific breeding ground to which they return to but do not have a specific non-breeding ground, while stepwise migrants migrate from their breeding grounds to two or more specific non-breeding grounds consecutively.

Nomadic migrants move between breeding and non-breeding grounds with neither being a fixed location (Rappole, 2013). Several of these types of migration are not mutually exclusive, the groupings are somewhat arbitrary and local populations may differ among each other in terms of migratory behaviour (Rappole, 2013). An organism's behavioural strategy is defined as a rule that determines how the organism responds to any set of possible circumstances/environmental conditions (Houston et al., 1999). In the context of migration in birds, strategy determines how environmental cues are perceived and converted into neuroendocrine and endocrine signals that regulate specific morphological, behavioural and physiological systems (Ramenofsky et al., 2012). This results in a migration phenology which influences decisions such as timing of migratory movement, accumulating fat reserves prior to migration, growing of flight muscles, leaving or staying longer in a stopover site, settling for breeding or continuing to travel and so forth. Cues are those aspects of the environment, both physical and social, that are perceivable and/or predictive enough to prompt such decisions (Winkler et al., 2014). They can be divided into two categories namely; initial predictive cues and local supplementary cues (Ramenofsky and Wingfield, 2007; Ramenofsky et al., 2012). Initial predictive cues initiate the transition from a life history stage (e.g. breeding) to migration. These cues need to be reliable and offer predictive information about conditions elsewhere (Ramenofsky and Wingfield, 2007; Ramenofsky et al., 2012). Daylength is an example of a predictive cue because it reflects planetary timing which possesses predictive information about conditions elsewhere and is not influenced by local conditions (Winkler et al., 2014). Local supplementary cues serve to accelerate or slow the migratory process by providing information about current local conditions. Examples include resource availability, social interactions, habitat alterations and predator density (Newton, 2008, 2012; Wingfield and Ramenofsky, 2011). Perception of cues may be similar or different at species, population and individual level, hence the many different types of migration observed (Helm et al., 2009; Ramenofsky et al., 2012). The many different migratory behaviours seen in birds have been historically categorized into two broad groupings namely, obligate and facultative migration (Newton, 2012). This categorization focuses more on what is presumed to be the driving force behind patterns of migratory behaviour. Obligate migration is considered to be under strong endogenous control, prompting an individual to leave its breeding area at a certain time of the year, and return at another time of the year. As such, it is characterised by its regularity, consistency and predictability (Newton, 2008, 2012). In contrast, facultative migration is considered as a direct response to prevailing conditions, for example, food supply or weather conditions. Individuals may migrate in some years and not in others and distance travelled may vary greatly from year to year and between individuals. It is thus mainly characterised by irregularity, inconsistency and unpredictability (Newton, 2008, 2012). Both strategies are adaptive with obligate migration being adapted to resource levels varying regularly and predictably and facultative migration being adapted to resource levels varying unpredictably (Newton, 2008, 2012). However, there are species that seem to employ both types of migration at different stages of their journey. One

example is the White Stork (*Ciconia ciconia*) where radio tracking has shown the same individuals departing from the same breeding sites in eastern Europe from year to year varying their wintering grounds, in some years wintering in east Africa and in some years moving further south to southern Africa (Berthold et al., 2002, 2004). Newton (2012) presents an argument that it may perhaps be more sensible not to distinguish between the two types but rather view them as opposite ends of a continuum with one end being fixed and rigid and the other being flexible and variable. All bird migrations can therefore be viewed as having both obligate and facultative phases but with the relative strengths of the phases varying between species and populations such that even the most obligate migrant can sometimes behave as a facultative migrant when experiencing certain circumstances (Newton, 2012). This highlights the importance of the availability and quality of environmental cues for migratory behaviour. As mentioned earlier, migratory birds collect information and integrate environmental and internal physiological conditions before making decisions about timing, direction, distance and cessation of migratory movements (Kumar et al., 2010; Winkler et al., 2014). One of the manifestations of these decisions, interpretable to us, is where and when migratory birds are seen, also understood as their geographic range. Gaston (1991, 1994, 2003) describes two broad measures of geographic range, namely the extent of occurrence and the area of occupancy. Extent of occurrence quantifies the area within the inner most limits of a species' occurrence and is inclusive of those parts that are not necessarily occupied. In simple terms, it can be described as a polygon around known occurrence records. Area of occupancy describes the area within the polygon in which the species is actually found. Assuming equal division of space (for example, grid cells), the area of occupancy of a species can be defined as the set of all grid cells in which the probability of occurrence exceeds a given threshold (Peterson et al., 2011). The probability of occurrence of a species at a given location is not random because a species' range is considered a geographical manifestation of its demographic response to spatial and temporal variation in environmental conditions (Pulliam, 2000; Holt et al., 2005). Therefore, the spatial patterns in rates of local extinction and local colonization determine how the distribution of a species changes through time (Gaston and Blackburn, 2002). In essence, seasonality changes environmental conditions at different times of the year. Birds then perceive environmental conditions or changes thereof partly based on their position on the obligate-facultative migration continuum and make decisions regarding movement and where to settle. This manifests itself in the geographic range of a species in both breeding or non-breeding season. Changes in the general pattern of such decisions over seasons, for example, deciding to settle in a new area, may lead to a shift, contraction or expansion of what is known as the geographic range of a species. Such shifts, contractions or expansions can be achieved through local colonization and local extinction dynamics (Tingley et al., 2009). Range fluctuations can therefore be better understood through local extinction and colonization dynamics which help explain the persistence of populations in dynamic systems over long periods of time (Hanski, 1998; Altwegg et al., 2008). Bird distributions are often climate-limited (Wormworth and Šeker-

cioğlu, 2011). However, climate works in conjunction with many biotic factors resulting in complex relationships that shape species' distributions (Holt and Keitt, 2000; Parmesan et al., 2005; Case et al., 2005). Even so, climate is often assumed to be the ultimate, albeit sometimes indirect, underlying determinant of species' distributions (Wormworth and Şekercioğlu, 2011). Synergy between climate change (IPCC, 2021) and other anthropogenic pressures such as habitat loss through land-use (Jetz et al., 2007; Goudie, 2013) is expected to result in localized and multidirectional patterns of change in communities (Root et al., 2003; Brook et al., 2008; Clavero et al., 2011). In response, birds are expected to either adapt to the new environmental conditions without shifting their general range or they may adjust their geographical distribution to escape by niche tracking (Böhning-Gaese and Lemoine, 2004; Huntley et al., 2006; Wormworth and Şekercioğlu, 2011). Their mobility, through flight, allows them to move to and between different habitats and ecosystems in response to irruptive or pulsed resource availability (Mahendiran and Azeez, 2018) much faster than other organisms that walk or swim (Alerstam, 1991). Relative to other taxa, birds would therefore seem to have an advantage to escape extrinsic threats and perhaps exploit certain changes (Kalle et al., 2018). However, complex annual cycles, long migration routes and dependence on different sites at different times make migratory bird species more susceptible to environmental change than their resident counterparts (Newton, 2004). There have been reports of declining population numbers in migratory birds in recent decades, which are declining at a faster rate in long distance migrants than in short-distance migrants or residents (Sanderson et al., 2006; Kirby et al., 2008; Heldbjerg and Fox, 2008; Thaxter et al., 2010). Subsequently, there is considerable interest in the distribution, ecology and movements of migrants (Newton, 2008). Currently, there is substantial evidence for the impact of global climate change on phenology, breeding biology and population dynamics of migratory birds (Trautmann, 2018). Changes in migratory distance, arrival and departure dates, and diminishing migratory behaviour in migratory species have been reported (Trautmann, 2018). Furthermore, changes in breeding timing and length of breeding seasons have also been reported in some species (Both and Visser, 2001; Parmesan and Yohe, 2003; Both et al., 2004; Doswald et al., 2009; Smallegange et al., 2010; Knudsen et al., 2011; Bussière et al., 2015). Such changes eventually lead to biogeographic changes through spatial abundance shifts, local extinctions and colonization events (Trautmann, 2018). For migratory species, biogeographic changes may be viewed at a seasonal or long-term temporal scale. Over long-term periods species may shift their distributions as individuals adapt and / or move in response to changes in mean conditions in the environment, typically thought as being a response to long-term climate or environmental change (Trautmann, 2018). At a seasonal scale, species may shift their distribution in response to seasonal conditions, for example the white stork, which winters in east Africa in some years and southern Africa in others (Berthold et al., 2002, 2004). In the context of conservation biology, an understanding of these biogeographic responses and the identification of the main determinants of a species' sensitivity to environmental change with regards to its range is important (Pereira et

al., 2010; McMahon et al., 2011), especially because range contractions, local extinctions and declines in species richness often form the basis of conservation action and effort (Böhning-Gaese and Lemoine, 2004). Southern Africa has had several examples of bird range expansions and contractions in the 20th century. Examples of expansions include the Yellow-Billed Oxpecker (*Buphagus africanus*, Stutterheim and Brooke, 1981), South African Cliff-Swallow (*Hirundo spilodera*, Rowan, 1963) and Southern Grey-headed Sparrow (*Passer diffusus*, Ward et al., 2004; Craig et al., 1987). Examples of contractions include the Blue Swallow (*Hirundo atrocaerulea*, Allan, 1988), Red-billed Oxpecker (*Buphagus erythrorhynchus*, Bezuidenhout and Stutterheim, 1980), Burchell’s Courser (*Cursorius rufus*, Hockey, 2005) and Eurasian Bittern (*Botaurus stellaris*, Dean, 2005). Expansions are mostly attributed to the emergence of new habitats resulting from anthropogenic activity, while contractions are attributed to habitat loss and degradation (Loftie-Eaton, 2014; Péron and Altwegg, 2015). In a study involving 408 species of South African birds, Hockey et al., (2011) found that 56 of the species that had undergone range shifts did not all conform to southward movement (towards cooler latitudes) as would be expected under climate change pressures. While most did move south (consistent with climate change predictions), some moved west (towards drier and warmer habitats, inconsistent with climate change drivers but consistent with land use drivers) and few moved to the north and east. Interestingly, while both southward and westward movers were habitat generalists, southward movers were mobile taxa (migrants/nomads), whereas westward movers were species that were generally associated with human-modified elements in the landscapes (Hockey et al., 2011). This may be suggestive of range shifts by mobile species being driven by their relatively higher sensitivity to direct impacts of climate change because of their close affinity to seasonal climatic conditions, while shifts in habitat generalists being driven more by exploitation of emerging new habitats. Climate is defined as the average temperature, humidity, atmospheric pressure, wind, rainfall, and other meteorological elements in an area of the earth’s surface over a long period of time. Climate change is therefore defined as a change in the state of the climate that can be identified by changes in the mean and/or variability of its properties over long periods of time (IPCC, 2014). Climate is a largely influential factor in global patterns in natural systems (Gitay et al., 2002; Thomas, 2010). It is therefore not surprising that there is evidence of ecological responses to climate change in the form of changing phenologies, biotic interactions and range shifts (Walther et al., 2002; Root et al., 2003; Parmesan and Yohe, 2003; Badeck et al., 2004; Chen et al., 2011b; Anderson et al., 2012; Chambers et al., 2013). Since the late 19th century, the earth has experienced a combined land and ocean surface temperature average increase of 0.99 °C (IPCC, 2021). Changes in precipitation patterns and an increase in the occurrence of extreme weather events in different regions have been observed globally (IPCC, 2021). The global impacts of climate change on natural systems are manifold but vary regionally and even locally in their occurrence, severity and implications (IPCC, 2021). Africa has experienced a 1.5 °C increase in near surface temperature over the past 50 – 100 years (IPCC, 2014). For southern Africa, minimum tem-

peratures have been reported to have increased more rapidly than maximum temperatures (New et al., 2006) resulting in decreased intra-annual seasonality (Böhning-Gaese and Lemoine, 2004). Warming in this region is predicted to continue in all seasons with a mean land surface warming that exceeds that of the global mean (James & Washington 2013). IPCC (2014, 2021) also reported downward trends in rainfall in parts of the region including western South Africa. Besides the overall reduction in mean annual precipitation, the changes in intra-seasonal characteristics of seasonal rainfall such as the onset, duration, dry spell frequencies and rainfall intensity (Dunning et al., 2018, 2016; Kniveton et al., 2009; Tadross et al., 2005) are alarming because they affect the availability and quality of cues that are used for phenology. Precipitation projections are less certain than that of temperature (Rowell, 2012; IPCC, 2014). For southern Africa, however uncertain, projections show a drying trend in the western parts relative to eastern parts (IPCC, 2014, 2021). Western South Africa, defined as all regions west of the Free State province, borders between being arid and semi-arid (CSIR, 2015) and thus is characterized by resource peaks that are variable in space and time owing to the region experiencing extremes in weather such as periods of prolonged drought to exceptionally high rainfall events (Dean, 2004; Dean et al., 2009). It has five distinct biomes namely; Desert, Succulent Karoo, Nama Karoo, Fynbos and Savanna (see figure 1.1; Mucina and Rutherford, 2006). The desert biome stretches from near the mouth of the Orange River running almost parallel to the river and the Namibia-South Africa border until the Onseepkans and Pofadder area. It is characterized by high aridity with most of the little rain received falling in summer months (Jurgens, 2006). Extending southwards from the western part of the Desert biome along the coast is the semidesert Succulent Karoo characterized by a relatively predictable low winter rainfall and high summer aridity (Mucina et al., 2006a). Between Vredendal and Clanwilliam Succulent Karoo shifts inland giving way to the Fynbos biome which occupies the southwestern tip of South Africa. It is characterized by a mediterranean type of climate with hot and dry summers and wet winters, although east of Cape Agulhas rainfall generally falls all-year-round (Rebello et al., 2006). Extending from the eastern part of the Desert biome, in a southeasterly direction to encompass north-eastern Western Cape and fringing of the western border of the Free State is the Nama Karoo biome. It is an arid biome characterized by low unreliable rainfall (Mucina et al., 2006b). The reliability and quantity of rainfall increases towards the east of the biome, while in the north eastern parts, the proportion of rain falling in summer increases (Mucina et al., 2006b). North of the Nama Karoo is the Savanna biome characterized by highly seasonal rainfall mostly falling in summer and subtropical temperatures (Rutherford et al., 2006). Figure 1.1: Vegetation biomes of South Africa (figure taken from Mucina and Rutherford, 2006). Western South Africa (herein WSA), has three rainfall seasonality zones that closely mirror the delineation of the biomes (see figure 1.2; Schulze and Kunz, 1997). The distinction between the rainfall seasonality zones is particularly important in the context of the drought that was experienced in the region between 2015-2017. Although the whole region was affected, the drought was severe in different zones at dif-

ferent times (Conradie, 2018; Archer et al., 2019). In the winter rainfall zone, the drought started towards the end of 2014 and continued to intensify until September 2017. The all-year rainfall south coast zone started experiencing severe drought towards the end of 2016 until late 2017. The Great Karoo interior experienced the peak of the drought from early to late 2017 (Conradie, 2018; Archer et al., 2019). 13 Figure 1.2: Rainfall seasonality in South Africa (figure taken from Schulze and Kunz, 1997) With changes to the environment emanating from climate change, land use and the 2015-2017 drought event, I suspect that some birds, particularly migrant species, would have responded to the changes in some way through their seasonal range dynamics. This is based on idea that, within varying environments, animals should theoretically settle in sites that maximize their fitness (Fretwell and Lucas, 1970). I aim to investigate the seasonal range dynamics of a selection of migratory bird species that use different migratory strategies (obligate and facultative) in western South Africa. In the process, I aim to showcase the value of using dynamic occupancy models over naïve site occupancy estimates, described as the number of times the species of interest is observed divided by the number of surveys conducted in the site, to model site occupancy. Furthermore, I seek to identify the key aspects of seasonality that influence the range dynamics of respective species. 14 Migrant and nomadic species present a complexity in their range dynamics that is not found in their resident counterparts. They occupy partly or wholly different ranges at different times of the year dividing their overall geographic range into two or more seasonal ranges (Newton, 2008). Essentially, they regularly vacate and return to their seasonal ranges, with some species known to vary significantly from year to year depending on where sporadic rainfall has created suitable conditions (Herremans, 1998). Such behaviour could obscure signals of range shifts, contractions or expansions occurring at seasonal and/or overall range scale. Therefore, data used to study changes in range dynamics must be appropriate. The Southern African Bird Atlas Project 2 (2007-currently) is a citizen science project which is aimed at mapping the distribution of bird species in southern Africa over long periods thus creating a “movie” of their distribution rather than a “snap-shot” which was achieved by the preceding SABAP (1987-1991) (Underhill et al., 2017). This makes SABAP 2 unique because it is one of the first bird atlas projects that aims to track bird distributions in both space and time, thereby tracking changes in distributions in real time (Underhill et al., 2017). The fact that data are collected continuously throughout the year opens an avenue for the study of migratory phenology and the changes thereof (Altwegg et al., 2012). Information on the movement patterns of different species of birds varies. Obligate migrants generally have more established and documented movement patterns while the irregularity of movement in facultative migrants makes it difficult to efficiently categorise some species into different types of facultative movement. I therefore relied on literature and expert opinion for species selection. I selected obligate migratory species based on their described migration phenology in Chittenden et al., (2016) and data availability on SABAP 2 (Table 1.1). Similarly, I selected facultative migratory species on the basis of reports of facultative movements such as nomadism from

Chittendem et al., (2016) as well as data availability on SABAP 2 (table 1.2).

Table 1.1: Obligate migratory species and their movement summary with respect to western South Africa according to Chittendem et al., (2016) Species Movement summary African paradise flycatcher (*Terpsiphone viridis*) Intra-african migrant arriving between Sept – Oct and leaves western South Africa during 15 European bee-eater (*Merops apiaster*) Greater striped swallow (*Hirundo cucullate*) White throated swallow (*Hirundo albigularis*) Yellow billed kite (*Milvus aegyptius*) April. It is resident in other parts of southern Africa. Two different populations occur in southern Africa: non breeding Palearctic migrants (mostly arriving in Oct and leaving in Apr) and breeding intra-African migrants (mostly arriving in Aug and leaving Feb). Breeding intra-African migrant arriving between Jul/Aug and departing between Apr/May. Breeding intra-African migrant arriving between Jul/Aug and depart between Apr/May. Intra-African breeding migrant, arriving in southern Africa from August-October and departing in March. Table 1.2: Facultative migratory species and their movement summary within western South Africa according to Chittendem et al., (2016). Species Movement summary African sacred ibis Resident or nomadic in response to water availability. (*Threskiornis aethiopicus*) Black headed canary Nomadic, sometimes irrupting out of usual range. (*Serinus alario*) Grey-backed sparrow lark Nomadic in response to rainfall; partially migratory. (*Eremopterix verticalis*) Lark-like bunting Highly nomadic. (*Emberiza impetuani*) Namaqua sandgrouse Nomadic and partially migratory; present in Namaqualand and Karoo in summer and (*Pterocles namaqua*) southern Kalahari in winter. 16 Southern Pale chanting goshawk (*Melierax canorus*) Largely resident but is known to move in response to conditions and food availability. 17

2. Chapter 2: Methods and dynamic occupancy models

Species distribution modelling (SDM) approaches that can accurately and efficiently quantify complex range dynamics over large spatial and long temporal scales can be used to understand the response of species in changing environments (Rushing et al., 2019). Species distributions often cover large geographic regions, thus no in-situ experiments can be reasonably carried out to investigate range dynamics and possible range shifts (Woodward, 1987). For this reason, modelling has been recognised as the most feasible and efficient way of studying species distributions (Morin and Thuiller, 2009). The different kinds of models can be categorized into two major groups: correlative and mechanistic models (Fletcher and Fortin, 2018). Correlative models exploit the statistical association between spatial environmental data and occurrence records to implicitly capture processes limiting the distribution of species, whereas mechanistic models are based on functional and physiological traits and explicitly incorporate processes that limit distribution (Kearney et al., 2010). In the context of impacts of climate change, there has been a level of confidence in the predictive power of mechanistic models because of their physiological basis (Kearney et al., 2010), however, Kearney et al. (2010) stress that their strength

depends on the identification of key limiting processes. Correlative models assist in identifying key limiting factors for the distributions and range dynamics of species and thus are useful in the development of hypotheses rather than testing (Jarnevich et al., 2015). However helpful correlative models are, the quality and nature of the data as well as statistical approaches used in modelling are critical (Guillera-Arroita, 2017). The common use of presence-only data in conjunction with environmental data of that specific location is often sufficient for identifying the general range of environmental conditions that a species experiences but largely confounds habitat preference of the species and habitat availability in the landscape (Guillera-Arroita, 2017). An alternative approach is the use of presence-only data in conjunction with characterizations of the environmental conditions on a broader scale. A comparison of the environmental conditions where the species was detected to the frequency of those conditions in the broader landscape is indicative of habitat preference (Guillera-Arroita, 2017). However, this approach provides no information on habitat avoidance, possibly leading to overestimation of occurrence probability. In addition, presence-only data does not permit inference about detection probability. Presence-absence data provide relatively more information because absences assist in identifying low suitability habitats that would have otherwise been considered as preferred habitat if presence-only data was used (Brotons et al., 18 2004). In addition, presence-absence data with some form of replication permit inference about detection probability. Classical SDMs do not account for imperfect detection. The probability of detecting a species, given that it occurs at a site, is almost always < 1 (Tyre et al., 2003; Kéry and Schmidt, 2008), therefore, when not accounting for imperfect detection, one estimates the area where the species is detected rather than the actual distribution of the species. In addition, the strength of covariate relationships with the distribution may be underestimated or overestimated when imperfect detection is not accounted for (Tyre et al., 2003). Occupancy modelling (MacKenzie et al., 2002; Tyre et al., 2003) offers a modelling framework that formally accounts for imperfect detection. In this framework the occupancy probability of a species can be jointly modelled as a function of spatially indexed covariates and the detection process, resulting in two nested logistic regression models to account for detectability (Kéry et al., 2013). An extension to this framework, dynamic occupancy models (MacKenzie et al., 2003), allows for the study of changes in species distribution and range dynamics over time. Here, the occupancy probability changes over discrete time steps and is estimated by explicitly modelling the process underlying occupancy dynamics in a Markovian manner (Kéry et al., 2013). In addition to modelling occupancy as a function of spatially indexed covariates and the detection process, colonization and local extinction probabilities of sites can be modelled as a function of relevant environmental predictors. Detection/non detection data results from two nested processes. (i) The state (or ecological) process generates the true distribution of a species with two possible states for each site (present or absent). (ii) Given the species of interest is present, the observation process results in the species being detected or not detected at a given site (Kéry et al., 2010). The occupancy state of a target species at a given site, z_i , resulting from

the state process is binary and can be modelled as a Bernoulli random variable such that $z_i \sim \text{Bernoulli}(\psi_i)$, where ψ_i is the probability of occurrence of the species at site i and z_i denotes the true state of occurrence with $z_i = 1$ indicating presence and $z_i = 0$ indicating absence. Adding the observation process at site i results in another Bernoulli random variable, with the success probability being the product of z_i and detection probability (p). The detection/non detection data (y_{ij}) observed at site i during survey j can be modelled as $y_{ij} | z_i \sim \text{Bernoulli}(z_i \times p_{ij})$. With a single survey, the true distribution (z_i) and the detection probability (p) are confounded leading to the modelling of the combination of z_i and p_{ij} as the basic parameter rather than occupancy probability (ψ_i) (Kéry et al., 2010). To separate these two components, repeat surveys at sites are required. The repeated surveys provide information that can be used to model the observation process and subsequently derive detection probability (p) separately from ecological process (Kéry et al., 2010). Repeated surveys at each site can be converted to detection histories (h_i) in the form 1-0-0-1, for example, indicating that the species was detected during the first and last survey, assuming no false positive errors, and not detected during the second and third survey, or 0-0-0-0 indicating the species was never detected throughout all four surveys. Assuming “closure” within a surveying season, meaning no extinction or colonization during the repeated visits to a site, the probability of observing a detection history can be modelled as a function of parameters of occurrence/occupancy (ψ) and detection probability (p) and maximum likelihood estimates can be obtained. For example, the target species would be considered present throughout a $h_i = 1-0-0-1$ surveying season being detected with a probability of p in the first and fourth survey, and not detected with a probability of $(1 - p)$ during the second and third survey. This can be expressed as $\Pr(h_i = 1001) = \psi \cdot p^1 (1 - p^2) (1 - p^3) p^4$. In a site with a detection history of $h_i = 0-0-0-0$, the target species may either be present with probability ψ but never detected with probability $(1 - p^1) (1 - p^2) (1 - p^3) (1 - p^4)$ or absent with probability $(1 - \psi)$. The probability of observing such a detection history can be expressed as $\Pr(h_i = 0000) = \psi (1 - p)^4 + (1 - \psi)$ if p is constant. The closure assumption, assuming no extinction or colonisation in a site over repeat surveys, results in the above explained model framework being static. This means for each site with a detection history derived from multiple visits, one occupancy probability estimate will be estimated. Temporal changes in site occupancy probability over a prolonged period can therefore not be estimated. Dynamic occupancy modelling (MacKenzie et al., 2003) extends the static form of occupancy modelling (Mackenzie et al., 2002; Tyre et al., 2003), by dividing detection histories into multiple surveying seasons. For example, a detection history of $h_i = 1001-0000-1110$, indicates the target species was detected twice in the first season, never detected in the second season and detected three times in the third season. The model assumes extinction and colonization (change in occupancy state) in a site may occur between seasons but not within a season. Occupancy state changes between seasons (time step) depend on the 20 occupancy state of the previous time step and the probability of local colonization (ϕ) and local extinction (ϵ). Therefore, the probability that a site is occupied at

a particular time ($t + 1$) is equal to the sum of (i) the probability of the target species being present at time t and not becoming locally extinct in the interim between t and $t + 1$ and (ii) the probability of the target species being absent at time t and colonizing the site in the interim between t and $t + 1$ (Kéry et al., 2013). Occupancy probability at $t + 1$ can be expressed as $\psi_{t+1} = \psi_t \cdot (1 - \epsilon) + (1 - \psi_t) \cdot \rho$. With the addition of the observation process, the model has four parameters: probability of initial occupancy ψ_1 , local extinction ϵ , colonization ρ , and detection p . For site i the probability of observing a detection history of $h_i = 1001-0000-1110$ can be expressed as $\Pr(h_i = 1001-0000-1110) = \psi_1 \cdot p^2 (1 - p)^2 \times \{(1 - \epsilon)(1 - p)^4 (1 - \rho) + \rho \cdot \epsilon\} \times \{p^3 (1 - p)\}$. Each of the parameters can be modelled as a function of covariates using a logit link function. The purpose of the logit link function is to ensure that predictions fall between 0 and 1. The logit link function is defined as: $\text{logit}(\psi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_U x_{iU}$, where ψ_i is the probability (or parameter) of interest for the i th sampling unit and $x_{i1}, x_{i2}, \dots, x_{iU}$ are the values for the U covariates of interest measured at the i th sampling unit. The regression coefficients $\beta_1, \beta_2, \dots, \beta_U$ determine the size of the effect of the respective covariates, and β_0 is the intercept term. The probability of initial occupancy (ψ_1) can only vary spatially and thus can only be modelled as a function of site-specific covariates, which are covariates that do not change between different surveys of a site, for example, height above sea level. Local extinction (ϵ), and local colonization (ρ) probabilities can vary both spatially and temporally between seasons and thus can be modelled as a function of site-specific covariates and seasonal site-covariates, which are covariates that may change over different seasons, for example, the average rainfall in a season. Detection probability (p) varies within a season, thus is modelled as a function of survey-specific covariates, which are covariates that can change each time a site is visited, for example, the number of hours spent while surveying the site. 21

2.1. Data

Citizen science projects present a way of collecting data at spatial and temporal scales that would be difficult to achieve through surveys carried out by professionals (Dickinson et al., 2010). However, the strength of any inference drawn from citizen science data depends on the appropriateness of the survey design and analytic approach (Altwegg & Nichols 2019). Collecting data without putting thought into how the data will be analysed often produces datasets that are not ideal for addressing focal questions (Altwegg and Nichols, 2019). In addition, the nature of atlas data inherently brings with it issues of heterogeneous and non-random sampling, false absences, false detections and spatial correlations within the data. Most of these problems can either be dealt with in the survey design or the analytic approach (Altwegg and Nichols, 2019). SABAP 2 has a protocol to ensure that data are collected in a manner that allows robust statistical inference and motivates volunteers to contribute. Contributors select a geographical pentad scale grid cell (5' x 5') where they bird for a minimum of two hours and record the species seen in the order in which they were seen on a survey card. The minimum of two hours as well as reporting of the hours spent

is aimed at reducing heterogeneity in detection and quantifying survey effort. Contributors are encouraged to sample as many habitats as possible within the bounds of the grid cell. They can continue adding to the checklist/survey card for up to five days. The requirement of starting a new list after five days is so that each successive list submitted by the same observer/s for a particular grid cell is not just a duplicate of the previous one (Underhill et al., 2017). Upon submission of each list, a rigorous vetting process is used to identify and further investigate dubious records. I obtained detection/non detection data from survey cards occurring throughout the area of interest - western South Africa (fig 2.1). Only survey cards with a minimum of two hours of effort between January 2014 to December 2018 from SABAP 2 resulting in a total of 33 505 unique survey cards over 4 748 pentads. I chose the period between January 2014 to December 2018 because it was the most data rich period. To reduce the number of pentads for computational purposes, I aggregated the data for all the survey cards from pentad scale ($5' \times 5'$) to quarter degree scale ($15' \times 15'$) resulting in a total of 975 quarter degree grid cells, herein called sites. I grouped the data by month with each month representing a surveying season resulting in a total of 60 surveying seasons between January 2014 and December 2018. I will herein refer to these surveying seasons as months to avoid any confusion with the 22 conventional meaning of the term 'season' as times of the year. The total number of surveys throughout the period for respective sites ranged between 13 and 88 per month. To obtain the monthly reporting rate for each site, also known as the naïve occupancy estimates, I divided the number of surveys where the species of interest was detected by the total number of surveys conducted in that month. Surveys were spatially sparse on a monthly scale. I therefore used the overall reporting rate to get an idea of the spatial pattern in occupancy of each species of interest (figure 2.2). Overall reporting rate is defined as the total number of surveys where the species of interest was detected throughout the period of interest divided by the total number of surveys conducted throughout the period of interest (2014-2018). In order to explore differences in spatial and temporal patterns in occupancy within the region and quantify to what extent a species moves its range within the study area on a monthly basis, I arbitrarily divided the region into four quadrants (quadrant 1, quadrant 2, quadrant 3, quadrant 4) using a horizontal line at -30° latitude and a vertical line at 22° longitude (figure 2.2). I classified sites within each grid based on probability of occupancy estimate. I defined site i to have high occupancy if $i \geq 0.8$, medium occupancy if $0.4 < i < 0.8$ and low occupancy if $i \leq 0.4$. This categorisation was only created to allow for the spatial visualisation of the spread of occupancy probabilities in the respective quadrants of the different months. Figure 2.1: A map indicating the region of interest. 23

Figure 2.2: A map of the overall reporting rate of the European bee-eater in Western South Africa (WSA) divided into four quadrants. The white grid cells did not have any data during the period of interest.

2.2. Covariates

Local temperatures have been shown to be an environmental cue that can influence individual-level departure decisions without a genetic change in reaction thresholds in migratory birds (Burnside et al., 2021). Similarly, variation in rainfall and food resources have been identified as cues that can influence departure times in migratory bird species (Studds and Marra, 2011). I therefore considered the average temperature, rainfall, and plant productivity as three main environmental variables that birds in the study region might be responding to. Rainfall monthly averages data at pentad scale were obtained from the Climate Hazards Centre InfraRed Precipitation with Station (CHIRPS) dataset dating back to 1989 to near-present. Temperature data for each grid cell were obtained from the MODIS land product MOD11C3 (Land Surface Temperature & Emissivity Monthly 0.05°). Similarly, vegetation index (NDVI) data were obtained from the MOD13C2 (Vegetation Indices Monthly L3 Global 0.05° CMG) land data 24 product. Both products provide monthly averages dating back to 2000 to near-present. NDVI provides a measure of land surface vegetation activity at spatiotemporal scales (Didan et al., 2015). Changes in vegetation activity reflect changes in net primary productivity, thus NDVI can be used a proxy for food availability (Thorup et al., 2017). All data products were aggregated to quarter degree scale ($15' \times 15'$). I used three types of covariates in the modelling process, namely site-specific covariates, survey-specific covariates and seasonal site covariates. To create site-specific covariates, which are variables that do not change between surveys of a site, I averaged all the monthly averages throughout the period of January 2014 to December 2018. Site-specific covariates characterise the different sites such that spatial patterns in temperature, rainfall and vegetation can be recognised (figure 2.3). For both groups of species, I used site-specific covariates to model the initial occupancy probability at each site. Survey-specific covariates are those covariates that may change over different surveys, for example, the hours spent surveying. I used survey-specific covariates to model the variation in detection probability. I defined seasonal site covariates as variables that vary between seasons but are constant between repeat surveys of the same site in the same season, where a season is a month. I then created variables that would reflect the change and anomaly in all three environmental variables between different months. I defined “change” variables as the current month average minus the previous month average in a site. “Anomaly” variables were defined as the current month average minus the average for the that month across the years, for example anomaly for May 2018 would be May 2018 minus average of May 2014, May 2015, May 2016, May 2017 and May 2018. Figure 2.3: Site-specific covariates. (A) Average rainfall (mm) for each quarter degree cell throughout the period of interest. (B) Average temperature ($^{\circ}\text{C}$) for each quarter degree cell throughout the period of interest. (C) Average NDVI for each quarter degree cell throughout the period of interest. The relevance of change in all three variables lies in the assumption that drastic change in conditions prompts migratory and nomadic birds to move in efforts to track their niche. The relevance of anomaly lies in the assumption that birds may move to other sites in efforts to exploit periodic abundance of

resources. In addition, change and anomaly variables capture the local temporal changes while eliminating static spatial differences which are described by the site covariates mentioned earlier. I explored the correlation between covariates using Kendall Rank Correlation Tests. None of the covariates were very highly correlated with the exception of rainfall and NDVI which had a fairly high correlation (0.72) while rainfall and temperature (-0.42) and temperature and NDVI (-0.46) were not strongly correlated.

Table 2.1: A table listing the names, abbreviations and justifications for the each of the covariates used in the models. Covariate name Average Rainfall (site-specific) Abbreviation Average Temperature (site-specific) Average NDVI (site-specific) Rainfall Anomaly (seasonal site-covariate) Temperature Anomaly (seasonal site-covariate) R T V RA TA Definition and justification This is the average of all the monthly rainfall averages throughout the period of interest (2014- 2018) within each pentad. When compared across the area of interest, it provides a spatial pattern showing differentiation between relatively dry and wet pentads. This is the average of all the monthly temperature averages throughout the period of interest (2014- 2018) within each pentad. When compared across the area of interest, it provides a spatial pattern showing differentiation between relatively warm and cold pentads. This is the average of all the monthly NDVI averages throughout the period of interest (2014- 2018) within each pentad. When compared across the area of interest, it provides a spatial pattern showing differentiation between pentads with relatively high and low primary productivity. Rainfall anomaly was calculated as the monthly rainfall average at any one month minus the average of all the monthly rainfall averages of that particular month throughout the period of interest (2014- 2018). This covariate provides an indication of how much any one monthly rainfall average deviates from the overall average rainfall in a particular pentad. Temperature anomaly was calculated as the monthly temperature average at any one month minus the average of all the monthly temperature averages of 26 NDVI Anomaly (seasonal site-covariate) Change in Rainfall (seasonal site-covariate) Change in Temperature (seasonal site-covariate) Change in NDVI (seasonal site-covariate) Log hours spent surveying (survey-specific) VA CR CT CV hours that particular month throughout the period of interest (2014-2018). This covariate provides an indication of how much any one monthly temperature average deviates from the overall average temperature in a particular pentad. NDVI anomaly was calculated as the monthly NDVI average at any one month minus the average of all the monthly NDVI averages of that particular month throughout the period of interest (2014- 2018). This covariate provides an indication of how much any one monthly NDVI average deviates from the overall average NDVI in a particular pentad. Change in rainfall was calculated as the monthly rainfall average at any one month minus the average rainfall of the previous month. This covariate provides an indication of how drastic the change is for rainfall from one month to the next. Change in temperature was calculated as the monthly temperature average at any one month minus the average temperature of the previous month. This covariate provides an indication of how drastic the change

is for temperature from one month to the next. Change in NDVI was calculated as the monthly NDVI average at any one month minus the average NDVI of the previous month. This covariate provides an indication of how drastic the change is for NDVI from one month to the next. Natural logarithm total number of hours spent per survey in each pentad. This covariate gives an indication of the surveying effort which influences detection probability. Change variables represent the change from month to month and reflect the seasonal changes in environmental conditions in a calendar year (summer, autumn, winter, spring). I, therefore, hypothesised that obligate migrants will show sensitivity to change variables due to the consistency and predictability they offer. In addition, I hypothesised that any given change variable will have opposite effects of colonization and extinction probabilities. For facultative migrants, I hypothesised a higher sensitivity to anomaly variables as anomaly variables reflect sudden changes in environmental conditions that prompt individuals to vacate or colonize a site. 27

2.3. Models

There was a total of 60 months between 2014 and 2018. I fitted dynamic occupancy models using the “occMod” function in RPresence 2.13.2 package (MacKenzie and Hines, 2021) in R version 4.0.0 (R Core Team, 2021). I limited the combination of variables for each of the four parameters to combinations that align with the aim of my study. Subsequently, I drew up a set of models that were general enough to be fitted on species that have different life-history traits. The subscript text indicates whether the covariate is site-specific (ss), seasonal site-covariate (ssp) or survey-specific (surv) A. (Rss + Tss + Vss) (Rss+Tss+Vss + change in rainfallssp + change in temperaturessp + change in NDVIssp) (Rss+Tss+Vss + change in rainfallssp + change in temperaturessp + change in NDVIssp) p (hours surv) B. (Rss + Tss + Vss) (Rss+Tss+Vss + rainfall anomalyssp + temperature anomalyssp + NDVI anomalyssp) (Rss+Tss+Vss + rainfall anomalyssp + temperature anomalyssp + NDVI anomalyssp) p (hours surv) C. (Rss + Tss + Vss) (Rss+Tss+Vss + rainfall anomalyssp + temperature anomalyssp + NDVI anomalyssp) (Rss+Tss+Vss + change in rainfallssp + change in temperaturessp + change in NDVIssp) p (hours surv) D. (Rss + Tss + Vss) (Rss+Tss+Vss + change rainfallssp + change in temperaturessp + change in NDVIssp) (Rss+Tss+Vss + rainfall anomalyssp + temperature anomalyssp + NDVI anomalyssp) p (hours surv) Note: Abbreviations were used for variables that were common in all four models (see Table 2.1). I assumed that average rainfall, temperature, and vegetation are variables that broadly capture the spatial patterns in environmental conditions throughout the study region and thus the spatial pattern in range for different species. Subsequently, I used all three variables to model the initial occupancy probability (1) of each species. To deduce derived temporal occupancy 28 probability estimates for each site at $t + 1$, I used the colonization and extinction parameters. I used different combinations of the seasonal site-covariates to model colonization and extinction probabilities. In addition to seasonal site-covariates, I added site-specific covariates to account

for the spatial pattern in average rainfall, temperature, and vegetation in each site. This was to separate between spatial and temporal effects on colonization and extinction probabilities. To avoid convergence failure, I used a min-max normalization to transform all variables that had different scales to a scale of between zero and one while preserving the relationships between variables. The probability of detecting a species, given that it occurs at a site, is almost always < 1 (Tyre et al., 2003; Kéry and Schmidt, 2008). The more time spent surveying would theoretically improve the probability of detection. I therefore used the natural logarithm total number of hours spent per survey in each site to model detection probability. To introduce more flexibility in the modelling of 1 in all four models, I used regression splines with two knots on all three variables leading to four more models (A1, B1, C1, D1). My implementation of regression splines follows that of Crainiceanu et al. (2005), where a basic function is calculated for each knot; these functions are then included as additive terms in a linear model. Unfortunately, introducing regression splines to the colonization and extinction parameters led to convergence issues. Splines were thus only used to model initial occupancy for model A1, B1, C1 and D1. Model comparison was done through Akaike weights (w_i) which were interpreted as the probability that a particular model is the best fitting model given the data and the set of candidate models (Wagenmakers and Farrell, 2004). I drew up conclusions about the strength of evidence of relationships between covariates and initial occupancy, local colonization, local extinction and detection probability parameters based on the direction of relationships and 95% confidence intervals on the obtained model coefficients. I used the “predict” function in the RPresence package to obtain predictions on all model parameters (t , t , t , pt). I used the predicted values to produce maps and plots of covariate relationships. Plots showing the relationship between any given parameter and covariate were plotted assuming mean values of all other non-focal covariates. I calculated the proportion of each occupancy category within the grids in different seasons (high, medium & low occupancy) as a measure of spatiotemporal patterns in occupancy within the region. 29

3. Chapter 3: Results and discussion

3.1. Obligate migrants

Occupancy probability estimates for all five obligate migrants tended to increase in spring and summer months (September – February) and decrease in the autumn and winter months (March – August) (fig: 3.1.1 – 3.1.5). This confirmed that the models were at least able to successfully capture the generally known seasonal pattern in occupancy (arrival and departure) of all five obligate migrants. For each species in the group, the temporal pattern of which months were estimated to have high occupancy and which months were estimated to have low occupancy was mostly consistent throughout the five year period. Similarly, the spatial pattern indicating the geographical position of high and low occupancy probability estimates was fairly consistent throughout

the five year period (fig: 3.1.1 – 3.1.5). This suggests that there were no obvious changes in seasonal range dynamics of all five obligate species throughout the period of interest. This may be interpreted as that the general environmental changes and changes in seasonality (mentioned in chapter 1) have not yet led to changes in seasonal range dynamics for the species included in this study. Alternatively, it is possible that change in seasonal range dynamics were not detected due to the seasonal and spatial sparsity in surveys carried out in the region or the five year period being too short to detect any real changes in seasonal range dynamics. The aggregation from pentad to quarter degree cells might have had influence on the ability of the models to capture any changes in seasonal range dynamics. Given more efficient computing power, I recommend the use of the same models on pentad scale data spanning over a longer period. Regardless of the possible alternative explanations, the consistency in spatial and temporal patterns of occupancy among the obligate migrant group of species concurs with the earlier introduced idea (chapter 1) that obligate migration is characterised by regularity, consistency and predictability (Newton, 2008, 2012). The comparison between the average of predicted occupancy estimates and the average of naïve occupancy estimates in each season revealed that, even though both averages followed the same pattern, being relatively high in spring and summer months and relatively low in autumn and winter months, the average of predicted occupancy estimates in the region of interest was generally always higher than the average naïve occupancy estimates in all seasons for all obligate species (fig 3.2.1 - 3.2.5). This result supports the idea that there is value in using dynamic occupancy models to estimate occupancy rather than naïve occupancy when using citizen science data. However, the fact that both averages followed the same patterns in 30 each species indicates that looking at the average of naïve occupancy estimates is still useful if the intention is to get an idea of when and where more sightings of a particular species are occurring. In southern Africa the african paradise-flycatcher is known to occur mostly along east southern Africa from Mozambique and Zimbabwe, extending south into eastern South Africa avoiding the more arid central and western parts of South Africa. In WSA (as defined in this study) it is known to have a relatively smaller range along the coast of Eastern Cape extending into Western Cape as far as Cape Town. As such, the proportion of sites estimated to have high probabilities of occupancy was generally low in the region in all seasons with the few relatively higher estimates only occurring in quadrant 3 and 4 (Appendix B: fig B1). During peak occupancy of the european bee-eater, the northern parts of the region (quadrant 1 & 2) have a higher proportion of sites that have occupancy probabilities above 0.8 than the southern parts (quadrant 3 & 4) (Appendix B: fig B2). This may be due to a gradual descent from the northern parts of southern Africa and eastern parts of South Africa, or it may be indicative of preference of conditions in Savana, Nama and Succulent Karoo biomes over the Mediterranean type of conditions found in the Fynbos biome beyond the Cape Fold belt. Interestingly, even though there was a generally consistent pattern, it is worth noting that 2017 showed slight deviation where the proportion of sites having occupancy probability estimate over 0.8 during

the peak month of November in all four quadrants dropped drastically in comparison to preceding and succeeding year. The greater striped swallow has a relatively big range within WSA, therefore during months of peak occupancy (Nov-Feb), there was generally a higher proportion of sites with high probability estimates within all quadrants besides grid 1 (Appendix B: fig B2). During the first month of arrival (September), sites with high probability estimates were almost always within quadrant 3 and 4. This indicates of preference of the coastal areas along the cape within the WSA, and then a shift towards inner regions of WSA (quadrant 2) as numbers increase and factors such as competition and resource availability intensify (Appendix B: fig B2). In a fashion similar to that of the greater striped swallow, during the first month of arrival (September) of the white-throated swallow, sites with high probability estimates were almost always within quadrant 3 and 4. During the peak estimated occupancy months, more high estimates start appearing within quadrant 2. Quadrant 1 has few to no sites with high estimates, which could be indicative of avoidance of the sub-region (Appendix B: fig B4). During the months of peak occupancy, the range of the yellow billed kite is limited to the coastline along 31 Eastern and Western Cape. This is a small portion of the region of interest (WSA) and therefore the figure showing the proportions of sites with high, medium and low estimated occupancy is not of much use because it is dominated by low proportions (Appendix B: B5).

Figure 3.2.5: A graph showing the average of naïve occupancy and the average predicted occupancy estimates between January 2014 and December 2018 for the Yellow-billed kite in the study area. Among the obligate migrants, the best performing models indicated that both colonization and extinction parameters were mostly driven by seasonal changes in average rainfall, temperature and vegetation (table 3.1). This is consistent with the already established understanding that obligate migrants respond to the change in the environment associated with changing seasonality. This was slightly different for the yellow billed kite where the best performing model indicated colonization to be mostly likely driven by anomalies in average rainfall, temperature and vegetation while extinction was mostly likely driven by changes in rainfall, temperature and vegetation. The relationship between initial occupancy probability estimates and the three environmental variables generally had noticeable uncertainty for all five species (fig 3.3.1 – 3.3.5, A – C; see Appendix A for coefficients). This was expected due to the sparse nature of the data. Even though uncertainty is relatively high, the strength and nature of the relationships may lend themselves as suggestions of preferred conditions for the different species. The more important relationships in the context of this study are the relationships between change or anomaly variables and the probability of colonization and extinction because they reflect on how a 40 species is likely to move in response to seasonal changes. In addition these relationships may suggest which environmental cues are most perceivable for each species.

As mentioned in the methods section, in addition to seasonal site-covariates, site-specific covariates were added on both the colonization and extinction pa-

rameters in efforts to separate the spatial and temporal effects on colonization and extinction probabilities. For the colonization parameter, the spatial variables generally had minimal effects for all five species (fig 3.3.1 – 3.3.5, D – F). This suggests that permanent physical characteristics of a site do not have a strong effect on the probability of a species colonizing a particular site. On the other hand, although varying between species, there were stronger relationships between colonization probability and change variables (fig 3.3.1 – 3.3.5, G – I). Relative to change in rainfall and change in NDVI, change in temperature stood out as having strong positive relationship with colonization probability for the african-flycatcher, european bee-eater, greater striped swallow and the white-throated swallow (fig 3.3.1 – 3.3.4, H). This suggests that with increasing temperatures between seasons (months) the probability that these species will occupy a site increases. Spatial variables used to model the extinction parameter had minimal effects on the extinction probabilities of all the species with the exception of the european bee-eater (fig 3.3.1 – 3.3.5, J – L). In contrast, temporal change variables had pronounced negative relationships with extinction probability (fig 3.3.1 – 3.3.5, M – O). This suggests that, with decreasing rainfall, temperature and NDVI between months, obligate species are more likely to vacate a site. It is worth noting that the relationship between change in temperature and extinction probability was predicted to have the least amount of uncertainty across all five species (fig 3.3.1 – 3.3.5, M – O). This, in combination with change in temperature standing out as having a strong positive relationship with colonization probability for most species in this group may suggest that temperature related environmental cues are the most reliable cues for obligate migrants in the region. As expected, the total hours spent surveying had a strong positive relationship with detection probability for all species (fig 3.3.1 – 3.3.5, P). As mentioned above, the yellow billed kite stands out as the only obligate migrant species to have its colonization probability modelled as a function of anomaly variables. Among the anomaly variables, NDVI anomaly (fig 3.3.5, I) had a much stronger effect on colonization probability than rainfall and temperature anomalies (fig 3.3.5, G,H). Dependency on anomaly variables may indicate a more facultative style of migration as per Newton’s (2008, 2012) description of facultative migration (see chapter 1). Considering it is a bird of prey and NDVI a proxy for food availability – NDVI results in more insects which are preyed on by small vertebrates – the yellow-billed kite could possibly present an example of an obligate migrant 42 that is on the more flexible side of the obligate-facultative continuum relative to the other obligate migrants. This would mean that it would behave as an obligate migrant and respond to general seasonal change, however, if given the opportunity, it would behave in a facultative manner colonizing sites to exploit abundance in resources that are not necessarily associated with changing seasonality. 43

Figure 3.3.1: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the african-paradise flycatcher where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in

rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total number hours spent surveying. 44 Figure 3.3.2: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for European bee-eater where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total number hours spent surveying. 45 Figure 3.3.3: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for greater-striped swallow where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total number hours spent surveying. 46 Figure 3.3.4: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the white-throated swallow where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total number hours spent surveying. 47 Figure 3.3.5: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the yellow-billed kite where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total number hours spent surveying. 48

3.2. *Facultative migrants*

Facultative migrants are different from obligates in that they do not have an established known seasonal movement pattern (Rappole, 2013). In addition, unlike obligate migrants, they do not completely vacate the country. This makes their seasonal range dynamics difficult to identify and understand. What is clearly understood is that they are expected to have irregular, inconsistent and unpredictable patterns in occupancy (Newton, 2008, 2012). However, this was not the case as the maps indicated that occupancy estimates for all facultative migrants did not show any irregular and inconsistent spatiotemporal patterns in occupancy between the different seasons, ultimately suggesting stable occupancy throughout the period of interest (figure 4.1.1 – 4.1.6). The analysis aimed at capturing differences in spatial and temporal patterns in occupancy within the region using quadrats revealed average occupancy did not differ between seasons in the respective quadrats for all species (Appendix B). Assuming the models performed as well they did for the obligate migrants, this result suggests that the general environmental changes and changes in seasonality (mentioned in chapter 1) were not drastic enough to prompt a response from facultative migrants. Alternatively, the irregular, inconsistent and unpredictable movements associated with nomadism and facultative migration are perhaps not as pronounced among the selected species as literature suggests. Alternative explanations for this result associated with the models not performing well include the possibility that changes in occupancy were not detected due to the seasonal and spatial sparsity in surveys carried out in the region or the five year period being too short to detect any real changes in seasonal range dynamics. A second alternative is that the species may have been responding to different covariates than those I considered leading to the model not being flexible enough to capture seasonal range dynamics. A third alternative explanation may be that the aggregation from pentad to quarter degree cells might have affected the ability of the models to capture any changes in seasonal range dynamics. Given more efficient computing power, I recommend the use of the same models on pentad scale data spanning over a longer period. A fourth and final alternative explanation would be that facultative movements did occur however they affected abundance rather than occupancy. The absence of regular and repetitive seasonal patterns of occupancy, such those seen among obligate migrants, was expected because facultative migrants are not known to have strong seasonal movement patterns as obligate migrants do. However, upon further investigation using the comparison between the average of predicted occupancy estimates and the average 49 of naïve occupancy estimates in each season, I was able to detect slight patterns of seasonality in occupancy for the lark-like bunting and namaqua sandgrouse (figure 4.2.4 and figure 4.2.5). The lark-like bunting displayed a fluctuating pattern in the average predicted estimated occupancy, where June and July consistently had the lowest averages while November and December had the highest averages throughout the period of interest (figure 4.2.4). For the namaqua sandgrouse highest averages were between November and January, while the lowest averages were between April and June (figure 4.2.5). This slight pattern of seasonality in species that are

considered to be highly nomadic may be interpreted as support to the idea that obligate and facultative migrants are not necessarily two different categories but rather two opposite ends of a continuum (Newton, 2012). Alternatively, this pattern may be a result of changes in abundance rather than changes in occupancy. It is worth noting that, similarly to the obligate migrants, the average of the predicted occupancy estimates in the region of interest was always higher than the average of the naïve occupancy estimates in all seasons for all species in the group (figure 4.2.1 – 4.2.6). This further highlights the value dynamic occupancy models. 50

59 Among the facultative migrants, the best performing models indicated the colonization parameter in all species was best modelled as a function of seasonal changes in average rainfall, temperature and vegetation, while the extinction parameter was best modelled as a function of anomalies in average rainfall, temperature and vegetation in all species besides the namaqua sandgrouse and the southern pale chanting goshawk. (Table 4.1; see Appendix A for model coefficients). With the understanding that facultative migrants move in a way that allows them to exploit irruptive or pulsed resource availability, one may be tempted to expect colonization to be driven by anomaly variables. However, individual birds cannot respond to irruptive or pulsed resource availability occurring elsewhere because the environmental cues flagging the anomaly would not be immediately available to them. If an anomaly event occurring in a site an individual is already occupying produces conditions that are favourable and exploitable to a facultative individual there would be no incentive to vacate the site. This would then not affect colonization probabilities because an individual cannot colonize a site it is already occupying. If the anomaly produces unfavourable conditions, an individual may vacate the site in search of better conditions, hence I see the extinction parameter being best modelled as a function of anomaly variables. Similar to obligate migrants, the relationship between initial occupancy probability estimates and the three site specific variables for facultative migrants varied and had considerable uncertainty for some species (fig 4.3.1 – 4.3.6, A – C; see Appendix A for coefficients). As mentioned in the methods section, in addition to seasonal site-covariates, site-specific covariates were added on both the colonization and extinction parameters in efforts to separate the spatial and temporal effects on colonization and extinction probabilities. The relationships between spatial variables and the colonization and extinction parameters may reflect the affinity of species to the different variables (fig 4.3.1 – 4.3.6, D – F). However, caution must be exercised when interpreting these relationships because the model is operating on aggregated spatial data which are not ideal for capturing fine scale movements that facultative migrants might be embarking on. For example, average NDVI had a relatively strong negative effect on colonization probability (fig 4.3.2, F) and a strong positive effect on extinction probability (fig 4.3.2, L) for the black-headed canary, indicative of its preference of the dryer central karoo region known to have relatively less vegetation (see map: fig 4.1.2). On the other hand, the average rainfall had a positive relationship with extinction probability for the african sacred ibis (fig

4.3.1, D), suggesting avoidance of higher rainfall areas. This is unlikely because the African sacred ibis is a water bird. 60

Among the temporal change variables, change in NDVI had strong positive effects on the colonization probabilities of the african sacred ibis, black-headed canary, grey-backed sparrowlark and lark-like bunting (fig 4.3.1 – 4.3.4; G – I). When compared to the obligate migrants, where change in temperature stood out as the variable influencing colonization probabilities the most for most obligate migrants, this supports the idea that obligate migrants have higher sensitivity to climatic cues while facultative migrants are more sensitive to resource availability which is usually associated with high NDVI. Rainfall anomaly, although differing in the nature of the relationships for different species, stood out as having the strongest effects on extinction probabilities of the african sacred ibis, black-headed canary, grey-backed sparrowlark and lark-like bunting (fig 4.3.1 – 4.3.4; M – O). In theory, because these species are associated with rainfall, the relationship between extinction probability and rainfall anomaly for all of them should be negative, such that the higher rainfall anomaly, the lower the extinction probability. The positive relationship between extinction probability and rainfall anomaly seen in the african sacred ibis and the grey-backed sparrowlark implies that they would likely move away from areas with high rainfall anomalies. This seems theoretically unlikely and is possibly an artifact of the shortcomings of aggregating the data. The extinction parameter for the namaqua sandgrouse and the southern pale chanting goshawk was modelled as a function of change variables. Change in NDVI had the strongest effects on extinction probabilities for both species (fig 4.3.5 – 4.3.6; M – O). The sensitivity to rainfall anomalies of four out of the six species is in agreement with what facultative movements are theoretically supposed to be sensitive to. However, both the namaqua sandgrouse and the southern pale chanting goshawk did not show any signs of sensitivity to anomaly variables although they are considered as facultative species. This may be indicative of the existence of a continuum between facultative and obligate movements. Alternatively, due to using aggregated quarter degree spatial scale, the model may not be sufficiently capturing relationships for some species. 62

Figure 4.3.1: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the african sacred ibis where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total 63 number hours spent surveying. Refer to table 3.3 for coefficients and associated standard errors. Figure 4.3.2: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the black-headed canary where A to C is the mean occupancy probability related to average rainfall

(A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total 64 number hours spent surveying. Refer to table 3.3 for coefficients and associated standard errors. Figure 4.3.3: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the grey-backed sparrowlark where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total 65 number hours spent surveying. Refer to table 3.3 for coefficients and associated standard errors. Figure 4.3.4: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the lark-like bunting where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total number hours spent surveying. Refer to table 3.3 for coefficients and associated standard errors. 66 Figure 4.3.5: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the namaqua sandgrouse where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total number hours spent surveying. Refer to table 3.3 for coefficients and associated standard errors. 67 Figure 4.3.6: Estimated relationships between parameters and covariates (95% CI) from the best dynamic occupancy model for the southern pale chanting goshawk where A to C is the mean occupancy probability related to average rainfall (A), average temperature (B) and average NDVI (C). D to I is the estimated colonization probability related to average rainfall (D), average temperature (E), average NDVI (F), change in rainfall (G), change in

temperature (H), and change in NDVI (I). J to O is the estimated extinction probability related to average rainfall (J), average temperature (K), average NDVI (L), change in rainfall (M), change in temperature (N) and change in NDVI (O). (P) is the estimated probability of detection related to the total number hours spent surveying. Refer to table 3.3 for coefficients and associated standard errors. 68

4. Conclusion

Citizen science projects present a way of collecting data at spatial and temporal scales that would be difficult to achieve through surveys carried out by professionals (Dickinson et al., 2010). However, the strength of any inference drawn from citizen science data depends on the appropriateness of the survey design and analytic approach (Altwegg and Nichols, 2019). This study demonstrates how dynamic occupancy models can be used with citizen science data to study seasonal range dynamics of bird species. It also highlights some of the shortcomings of the data as well as the modelling approach that can be addressed in future studies. According to the models, the two groups of species in this study did not show pronounced patterns of change in their seasonal range dynamics over the years of the study. The obligate migrants maintained their general known arrival and departure months while also maintaining a fairly consistent pattern in the spatial distribution when they were present. This result supports the characterisation of the fixed side of the obligate-facultative migration continuum as being regular, consistent and predictable (Newton, 2008, 2012). As hypothesised in chapter 2, all obligate migrants except the yellow-billed kite, showed sensitivity to change variables for both colonization and extinction probabilities. In addition, the effect of any one change covariate on colonization probability had the opposite effect on extinction probability. The yellow-billed kite was the exception because the best performing model indicated colonization to be most likely driven by anomaly variables and extinction by change variables. This was the first example of an obligate migrant species that could possibly be considered to be on the more flexible side of the obligate-facultative continuum relative to the rest of the species in the obligate migrant group. The models indicated that, for the majority of the obligate migrant species, colonization probability had a stronger positive relationship with change in temperature relative to change in rainfall and vegetation. For all five species, the relationship between change in temperature and extinction probability was predicted with the least amount of uncertainty compared to relationships with change in rainfall and vegetation. This suggests change in temperature as a key aspect of seasonality influencing the range dynamics of obligate migrants. 69 The facultative migrants did not show the expected irregular and inconsistent spatiotemporal patterns in occupancy between the different seasons, ultimately suggesting stable occupancy throughout the period of interest. This may be due to a number of reasons based on how well the models performed. Assuming that the models worked as well as they did for the obligate migrants, it is possible that nomadism is perhaps not as pronounced as expected in the facultative migrant species group with

some species being more resident than nomadic. Possible alternative explanations associated with the models not performing well due to the limitations of this study include changes in seasonal range dynamics not detected due to the seasonal and spatial sparsity in surveys carried out in the region or the five year period being too short to detect any real changes in seasonal range dynamics. A second alternative possibility may be that the species may have been responding to different covariates than those I considered leading to the model not being flexible enough to capture seasonal range dynamics. Examples of such covariates may include social interactions, habitat alterations and predator density. I did not consider these kinds of cues because the scope of this research was to investigate key aspects of seasonality that influence the range dynamics of respective species considering the changing climate in the region. A third alternative explanation may be that the aggregation from pentad to quarter degree cells might have affected the ability of the models to capture any changes in seasonal range dynamics. Given more efficient computing power, I recommend the use of the same models on pentad scale data spanning over a longer period. A fourth and final alternative explanation would be that facultative movements did occur, however, they affected abundance rather than occupancy. The best performing models for the facultative migrants indicated that the extinction parameter was best modelled a function of anomaly variables for the majority of the species in the group. This was in agreement with the hypothesis that facultative migrant species would show sensitivity to anomaly variables. The colonization parameter was best modelled as a function of change variables for all the species in the group. This was contrary to the hypothesised sensitivity to anomaly variables, however, it is explained by the fact that individual birds cannot respond to irruptive or pulsed resource availability elsewhere because the environmental cues flagging the anomaly would not be immediately available to them. In addition, if the anomaly event produces favourable and exploitable conditions at a site an individual is already occupying the individual would theoretically continue occupying the site, therefore, the colonization probability would not be affected because an individual cannot colonize a site it is already occupying. If the anomaly produces unfavourable conditions, the individual may vacate the site in search of better conditions, hence the models indicate sensitivity to anomaly variables only for the extinction parameter. Although differing between species, the modelled relationships between environmental covariates and colonization and extinction parameters indicated that the two groups generally respond to different covariates. For obligate migrants, changes in temperature had a relatively strong positive effect on the colonization probability of the majority of the species, while all three change variables had strong negative effects on extinction probability. For facultative migrants, changes in NDVI had a relatively strong positive effect on the colonization probability of most species, while it was anomalies in rainfall that influenced the extinction probability for most species in the group. The general response to different covariates by the two groups supports the idea of obligate migrants being more likely sensitive to direct climatic environmental cues while facultative migrants being more sensitive to environmental cues that yield some form of immediate

benefit introduced in chapter 1. The exceptions within these group trends support the idea of migration strategies being a continuum rather than two separate groups (Newton, 2012).

References

Teitelbaum, C.S., Mueller, T., 2019. Beyond migration: causes and consequences of nomadic animal movements. *Trends in ecology & evolution* 34, 569–581.