Chatter About the Clouds of White: Using Social Media to Track the Annual Migration of *Belenois aurota*

by

Subhashinidevi Pillay (0504582M)

Primary Supervisor: Associate Professor Jennifer Fitchett

Co-supervisor: Dr Helen Robertson

A research report submitted in partial fulfilment of the requirements for the degree of

Master of Arts in the field of e-Science

in the

School of Social Sciences

University of the Witwatersrand, Johannesburg

February 2021

Abstract

Butterflies are among the most sensitive biota to climate change and variability in the animal kingdom. Shifts in phenological events are spatially and temporally specific. *Belenois aurota* is one of the most commonly occurring butterfly species in South Africa. There is limited phenological data in South Africa about butterfly migrations including that of the *Belenois aurota* migrations. This lack of data poses challenges to the monitoring of local ecosystems, especially in understanding changes under the climate change umbrella. To determine whether data from social media can address this paucity of data, a mixed method approach was used obtain 1063 temporally constrained accounts of the *Belenois aurota* butterfly migration through social media and historical newspapers. For the period 1914-2020, dates of butterfly migrations from newspaper articles were obtained from a secondary source for the initial part of the study period. For the previous decade, social media observations from Twitter, Instagram and Facebook were collected. An advance in the migration date of 2.4 days per decade was calculated for the study period, with a median migration date shift from 269 Julian Days (25 March) to 214 Julian Days (30 January). However, when compared to the citizen science project LepiMAP, a delay of 2.3 days per decade is calculated. Spatially, observations from the database were concentrated in the northern and eastern parts of the country. Peak observations for the entire dataset were between December and February for all years. These changes are consistent with shifts in global phenology whereby butterfly species are experiencing an advance in phenology due to increases in temperatures.

Key words: *Belenois aurota*, butterfly migration, social media, Facebook, South Africa

Acknowledgements

The support of the DSI-NICIS National e-Science Postgraduate Teaching and Training Platform (NEPTTP) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and should not necessarily be attributed to the NEPTTP.

I would like to thank my two supervisors, Associate Professor Jennifer Fitchett and Dr Helen Robertson, for their guidance, time and patience and being such incredible female role models in Academia. The great amount of time spent by Prof Fitchett over email and WhatsApp, explaining and clarifying concepts, is most appreciated. I am thankful to the FitzPatrick Institute of African Ornithology for access to their records on *Belenois aurota*, as well as Antonia Pandazis for access to the historical database that she has compiled.

I would like to thank my two furry four-legged companions Liquorice and Lilypad for the many hours they spent sitting with me while working on my studies, especially during 2020 which turned out to be a very lonely academic year. Most of all I am very grateful to my incredible husband, Fabian Francis, for supporting my return to university for a full-time programme. I am also very appreciative for the time and effort he has invested in proofreading my drafts.

Declaration

I, Subhashinidevi Pillay, declare that this research report is my own, unaided work. It is being submitted for the degree of Master of Arts in the field of e-Science at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.

Pillay

Subhashinidevi Pillay
2nd February 2021

Table of Contents

List of Figures

List of Tables

List of Acronyms and Scientific Terms

AI: Artificial Intelligence

DNA: Deoxyribonucleic acid

DAYS/DECADE: Rate of change in days per decade

ENSO: El Niño Southern Oscillation

FNAI: Florida Natural Areas Inventory

LSA: Lepidopterists' Society of Africa

MLMP: Monarch Larva Monitoring Project

NDVI: Normalised Difference Vegetation Index

r: Pearson's correlation coefficient

R^2: Coefficient of Variation

SABCA: Southern African Butterfly Conservation Assessment

T_{max}: Maximum Temperature (°C)

T_{mean}: Mean temperature (°C)

T_{min}: Minimum Temperature (°C)

UKBMS: UK Butterfly Monitoring Scheme

Chapter 1: Introduction

1.1 Background

In January of 2020 South African residents experienced a noticeably large number of butterflies as they traversed through several parts of the country during their annual migration (Terblanche, 2020). The size of the migration prompted many individuals to take to social media to report their butterfly sightings through posts, pictures, and video footage on their social media accounts (Burger, 2020; Mbhele, 2020). The species partaking in the migration include the African Migrant (*Catopsilia florella*) and the Brown-veined White butterfly (*Belenois aurota*) - the latter being the main participant in the Kgalagadi Butterfly migration (Terblanche, 2014; 2020). Google Trends showed an increased number in Google searches within the country for the terms "*Belenois aurota*" (Fig 1.1a), "Brown-veined White Butterfly" (Fig 1.1b), "Butterfly Migration" (Fig 1.1c) and "White Butterfly" (Fig 1.1d) in January 2020 in South Africa, which illustrated the public interest in the species and the migration thereof. Each of the terms shows a distinct peak in January and February of 2020 (Fig. 1.1a, b, c, and d). Other peaks are observed in January 2013 and 2014. A wealth of social media data is being generated daily (Kemp, 2020). Even with this abundance of information, the social media data around *Belenois aurota* has not been analysed before.

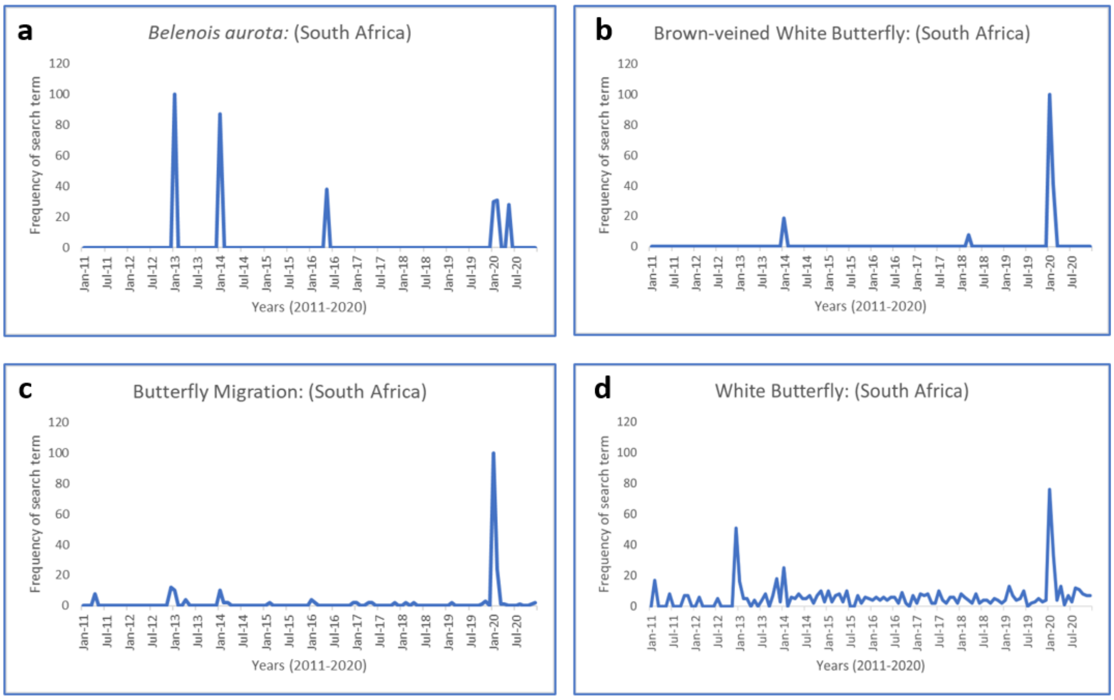


Figure 1.1a-d: Google Trends highlighting the increase in searches for specific terms relating to *Belenois aurota.*

Phenology refers to "the study of the timing of recurrent biological events such as flowering times, migration and breeding, the causes of their timing regarding biotic and abiotic forces, and the interrelationship among phases of the same or different species" (Badeck *et al.,* 2004: 295). Phenology has broad applications, and this study focuses specifically on the migration phenology of the butterfly species *Belenois aurota*. Butterflies are a sensitive biota in that they are affected by ambient temperatures (Illán *et al*., 2012; Roy *et al.,* 2001; Stefanescu *et al.,* 2003). Butterflies are dependent on plants at both larval and adult stages of their life cycle and, as a result, are also sensitive to climatic variables similarly to plants (Colom *et al*., 2020; Stefanescu *et al.,* 2011; Wallis De Vries *et al.,* 2011). As a result, research on the effects of various climatic variables on butterfly phenology is providing many insights into the effects global warming has on butterfly species (Colom *et al.*, 2020). There are several studies on butterfly phenology in the Northern Hemisphere (Altermatt, 2010; Forister and Shapiro, 2003; Roy and Sparks, 2000; Stefanescu *et al.,* 2003). The abundance of studies is primarily attributed to copious databases generated through decades of monitoring in some of these countries (Schmucki *et al*., 2016; van Swaay *et al.,* 2008). These studies detail the advances in butterfly phenology for several species, based on local climatic changes experienced in those regions (Altermatt, 2010; Forister and Shapiro, 2003; Roy and Sparks, 2000; Stefanescu *et al.,* 2003).

Where monitoring programmes do not exist, citizen science projects have played a big role in generating data for phenological studies (Kosmala *et al.,* 2016; MacKenzie *et al.,* 2017; Taylor, 2019; Willis *et al.*, 2017). The projects vary for plant and animal studies; for plant phenology, this includes remote sensing, digital repeat photography, and phenological models and experiments (Fitchett *et al.,* 2015). The Normalized Difference Vegetation Index (NDVI) is a satellite remote sensing technique that has been used to detect phenological shifts at a continental, regional and local scale by quantifying the 'green wave' (Fitchett *et al.,* 2015; Fitchett and Fani, 2018). These remote sensing methods are advantageous in that they can detect phenological changes across broader spatial ranges which would be challenging to achieve with more ground-based recording (Zhao *et al.*, 2013; 2015). These methods are, however, not always efficient at detecting species-level phenological changes (Zhao *et al.*, 2015). For animal phenological studies, camera traps are increasingly being used for migration phenology of large mammals (Burton *et al.*, 2015; Hofmeester *et al*., 2019; Willi *et al.,* 2018). In the case of butterflies specifically, the field has adapted to data shortage through citizen science projects due to challenges with large scale and sufficiently sensitive monitoring (Dennis *et al*., 2017; Golumbic, 2015; Hart *et al*., 2018; Howard and Davis, 2009; Jue and Daniels, 2014; Prudic *et al.*, 2017; Ryan *et al*., 2019; Schultz *et al.*, 2017; Wilson *et al*., 2015). Butterfly tagging projects have also been used for a better understanding of migration phenology in butterflies (Mouritsen *et al.*, 2013; Taylor *et al.*, 2020).

1.2 Rationale

Overall, there is a scarcity of ground-based observations of phenological events and associated phenological research in the African and Middle Eastern region (Fitchett *et al.*, 2015; Fitchett and Fani, 2018). In South Africa, the same holds true (Fitchett and Fani, 2018; Fitchett *et al.,* 2015; Whitecross *et al.,* 2017). South Africa does not have well established butterfly species monitoring programmes such as those in the Northern Hemisphere. The limited data for *Belenois aurota* exists predominantly in the form of citizen science data from LepiMAP. LepiMAP is an initiative by the University of Cape Town's Animal Demography Unit and the Lepidopterists' Society of Africa (Loftie-Eaton, 2016) which serves as an Atlas of African and South African Lepidoptera (Loftie-Eaton, 2016). LepiMAP, however, also has a limited amount of data as this is dependent on the contributions made by citizen scientists who need to input data onto the LepiMAP virtual museum platform (Loftie-Eaton, 2019). There is also currently butterfly specific content being generated on Facebook via the "The Great Little White Butterfly Migration" Facebook page. However, it appears that the purpose of the social media component is to entice citizen scientists to post their observations on LepiMAP. It is not known to what extent the social media data being generated is analysed. Valuable data generated on social media may not be utilised effectively in tracking environmental phenomena (Kirilenko *et al.*, 2015; Lopez *et al.*, 2019).

Butterflies in general are an indicator of an ecosystem's health (Woodhall, 2012). They mostly play an important role in pollination and form a key component of the food chain (Woodhall, 2012; 2020). Butterflies serve as important flagships for invertebrate conservation as they are regarded as the most popular of invertebrate species (Willis and Woodhall, 2010). As with many other butterflies, *Belenois aurota* is most likely a species sensitive to changing climates and environments, however, without data to analyse, inferences cannot be made about the local impacts of climate on this species. Unfortunately for the Brown-veined White butterfly in the South African context, details of these interactions are poorly understood as limited research is being conducted and far less is being published about this species. The limited available research on *Belenois aurota* and their migrations in South Africa was published in the 1950s and 1960s (Broekhuysen and Broekhuysen, 1960; Dickson and Schoefield, 1958; Milne, 1964; Taylor, 1959). This data shortage has implications for the understanding of regional and local phenological shifts, especially under climate change (Fitchett *et al.,* 2015).

This project will contribute to generating data in the field of phenology which is a scientific field that has a paucity of data in the South African context (Fitchett *et al.,* 2015; Fitchett and Fani, 2018). There is an increasing number of South Africans who have consistent internet access across class and economic divides (Stats SA, 2019). Of those who have internet access, it is estimated that 23 million South Africans are active on social media platforms (Barr, 2020; Kemp, 2020). This project, therefore, determines whether posts from news and social media sources such as Facebook, Twitter and Instagram can provide meaningful insights into the annual *Belenois aurota* butterfly migration, given the quantity of data that is being generated on social media each day. Social media is also said to be a cost-effective and abundant resource for research (Breckheimer *et al.,* 2019; ElQadi *et al*., 2017) and this is ideal for South Africa, a country that has other financial focuses and may not have the resources for elaborate monitoring programmes.

1.3 Aims and Objectives

This aim of the study is to address the data paucity in butterfly phenology by generating a database from social media regarding the timing of adult *Belenois aurota* sightings during the annual butterfly migration. To determine the changes in the phenology of these migrations, historical newspapers sourced from a concurrent academic study was used. This aim will be achieved through the following objectives:

1. To compile a database of dates and locations of tweets relating to the butterflies.
2. To assess the interannual variability in the timing of posts on the migration.
3. To compile a spatial inventory of butterfly sightings which are differentiated by year.
4. To determine how the interannual variability of the social media and newspaper posts compare to that of the LepiMAP data set for the same period.

1.4 Overview of the Report

Chapter 2 provides a review of the literature relating to this research project. An overview of phenological research is given both in a broad context and specifically to butterflies. An overview of South African phenological studies is provided. Citizen science research is also covered in this chapter. A discussion of butterfly-related citizen science is given. The chapter concludes with a discussion of social media research and more specifically details biometeorological butterfly research.

Chapter 3 describes the study area for this research which is at the scale of South Africa. A brief description of the vegetation and climate of South Africa is given. Social media usage in the country is also covered briefly.

Chapter 4 describes the research methodology that was utilised to conduct this research project. Manual and automated web scraping methods used to obtain the data is explained. A combination of automatic text analysis and content analysis were used to analyse the social media posts. A linear regression model was utilised to analyse both social media and LepiMAP data. Spatial records are visually analysed.

Chapter 5 presents the findings of the research. These findings are divided into the four sections corresponding to each of the research objectives described in section 1.3. These are a description of the database; the variability in the timing of records relating to *Belenois aurota* (reflected in bar graphs, box plots and a regression model); the spatial inventory of sightings (shown in summary bar graphs and maps relating to the spatial distribution of records); and finally the social media data and LepiMAP citizen science project is compared using bar graphs, box plots, and regression analysis.

Chapter 6 provides a detailed discussion of the results and relates the results to broader literature. The content of the social media posts is discussed. A comparison is drawn between the timing of the social media posts and the LepiMAP data.

Chapter 7 concludes the report by discussing how objectives were achieved. Future work that can be conducted in this field is outlined.

Chapter 2: Literature Review

The existing literature on phenology, citizen science and the use of social media as a data source in biometeorological and ecological studies are discussed in this chapter. The phenology subsection will broadly discuss research relating to the field. Thereafter, the phenological studies conducted in a South African context are mentioned. Citizen science is one of the key components of this research. This subsection will examine the type of research that is being conducted and then focus specifically on butterfly-related citizen science projects. The final section of this chapter will deal with the use of social media as a data source. It will provide details of some of the popular applications of social media data and applications to ecological research.

2.1 Phenological Research

There is evidence to suggest that shifts in phenological events, such as migrations, may not occur at the same time each year due to changing climates (Parmesan and Yohe, 2003; Schwartz *et al.,* 1999), thus resulting in advances and delays in phenological events due to global warming (Chambers et al., 2013; Parmesan and Yohe, 2003; Root *et al.,* 2003). As a result of the change in the timing of these events, there is often a mismatch in the interaction of species. One such example is between flowers and pollinator species (Hegland *et al.,* 2009; Rafferty and Ives, 2011) whereby an advance in flowering was experienced when species were not constrained by pollinators (Rafferty and Ives, 2011). However, Hegland *et al.* (2009) argue that pollination networks are more robust and offer methods for studying these mismatches.

In terms of phenological data, Europe has detailed records of phenological events (Chmielewski and Rötzer, 2002; Fitchett *et al*., 2014; Menzel *et al.,* 2006). In areas where high-resolution datasets do not exist, there have been attempts to find new records of phenological data that have been captured in sources such as diaries, logbooks, and newspapers (Chmielewski and Rötzer, 2002; Fitchett *et al.*, 2014; Fitchett and Fani, 2018; Menzel *et al.*, 2006). There have also been initiatives to include phenological observations from local citizens (Leocadio *et al*., 2018; Mackenzie *et al.,* 2017). In the Northern Hemisphere, mobile phone applications are used to facilitate citizen science initiatives which include data collection and analysis of scientific phenomena by the public (Graham *et al.,* 2011). In the

Southern Hemisphere, efforts have been predominantly to understand indigenous knowledge systems and local understanding of phenological events (Chambers *et al.,* 2017).

Citizen science has also proven to be an efficient way to contribute to phenological research (Bison *et al.,* 2019). There is a large range of projects including plant-related citizen science phenology projects (Bison *et al*., 2019; MacKenzie *et al.,* 2017; Taylor, 2019) and animal-related citizen science phenology projects (Eisenhauer *et al*., 2018; Newson, 2016).

Global climate change has impacted on the phenology of species and therefore climate is a key factor in understanding species phenology (Araújo and Rahbek, 2006; Bellard *et al.,* 2012; Parmesan and Yohe, 2003; Thackeray *et al*., 2016; Thomas *et al*., 2004). Research being conducted on species responding to climate change is plentiful (Devictor *et al.,* 2012; Radchuk *et al*., 2019; Socolar *et al.,* 2017). Butterflies are said to be one of the primary indicators of climatic changes (Illán *et al.*, 2012; Stefanescu *et al.,* 2003). These taxa are poikilothermic - affected by the ambient temperature - and these species, therefore, show rapid responses to changes in temperature (Illán *et al.,* 2012; Roy *et al.,* 2001; Stefanescu *et al.,* 2003). Butterflies also develop dependencies on plants in both the larval and adult phases of their life cycle; this makes butterflies sensitive to climate variables, particularly to precipitation and humidity (Colom *et al*., 2020; Stefanescu *et al.*, 2011; Wallis De Vries *et al.*, 2011).

An abundance of precise phenological data that is being generated through monitoring programmes (Schmucki *et al*., 2016; van Swaay *et al.*, 2008) facilitates climate change research; therefore, the impacts of climate change on these species are well known. Overall, the rise in temperatures has resulted in the seasonal advance of butterfly activity over time (Altermatt, 2012; Sparks and Yates, 1997; Stefanescu *et al.,* 2003). There have been reports of the advance of flight periods in butterfly species that were connected to increasing temperatures (Roy and Sparks, 2000). Several accounts of this relationship occur in the Northern Hemisphere that document similar advances, including studies in Central Europe (Altermatt, 2010) and the Mediterranean (Forister and Shapiro, 2003; Stefanescu *et al*., 2003). However, differences in the species' response to warming climates may occur due to factors such as narrow diets (Diamond *et al.,* 2011), habitat (Altermatt, 2012), elevation gradient (Illán *et al.,* 2012), microclimate (Wallis De Vries and Van Swaay, 2006), and impacts of climate on seasonality (Walther *et al.,* 2002). Species phenology is dependent on seasonality and this is distinct for each species (Snyman, 2020; Walther *et al.,* 2002). Some studies have highlighted the mismatch between interdependent species (Balfour *et al.,* 2018; Schleuning *et al.,* 2020) and have shown how changes in phenology due to climate warming is causing extinction (Balfour *et al.,* 2018).

In comparison to research done globally in this field, a limited number of studies in South Africa exist (Fitchett *et al.,* 2015; Fitchett and Fani, 2018; Whitecross *et al.,* 2017). In the South African context, much of the phenology-related research focuses primarily on plants. These studies include research on flowering phenology of *Oxalis* (Dreyer *et al.*, 2006), the changes in the blossoming of jacaranda (Fitchett and Fani, 2018), advance in pear and apple tree blossoming (Grab and Craparo, 2011), the flowering of baobabs (Venter and Witkowski, 2019), and the advance of first flowering and full bloom of Namaqualand daisies (Snyman, 2020). Other studies include remotely sensed vegetation phenology (Wessels, 2011), savanna tree phenology (Moyo *et al.*, 2015; Whitecross *et al.,* 2016; Whitecross *et al.,* 2017), grassland phenology (Moyo, 2018), and woody community phenology (Shackleton, 1999). Limited phenology studies in South Africa include the migrations of barn swallows (Altwegg *et al.,* 2012; Burman, 2016; Møller *et al.,* 2011), delays in the timing of the sardine migration (Fitchett *et al.,* 2019), photoperiods as a predictive cue for migration of long-fingered bats (Pretorius *et al.,* 2020), and the resistance of maize plants to Lepidoptera (van Wyk *et al*., 2008). While these studies show broad trends, the species specificity means that the assumption cannot be made that changes in plants and sardines are shared by butterflies. This is demonstrated in the rising issue of species mismatch (Fitchett *et al*., 2015; Snyman, 2020).

The rate of change in phenological studies is an important factor to be considered. The rate of changes for South African phenological studies are summarised in Table 2.1. The plant species that have been investigated have all experienced an advance of between 1.1 and 2.38 days/decade in the flowering of these plants (Grab and Craparo, 2011; Fitchett and Fani, 2018; Snyman, 2020). In contrast with the trend of advancing seen in flowering plants, the one animal species that showed a delay in an associated rate of change is the sardine run, with a delay of 1.3 days/decade (Fitchett *et al.*, 2019).

In terms of the data used for the analysis of these studies, only one of the three has a more traditional phenological dataset (Grab and Craparo, 2011) which was obtained from a network of orchards. The other two were reconstructed using newspaper records (Fitchett and Fani, 2018; Fitchett *et al.*, 2019) and social media data (Fitchett and Fani, 2018). Linear regression models were the main methods used to analyse the data in these studies (Grab and Craparo, 2011; Fitchett and Fani, 2018; Fitchett *et al.*, 2019). However, in contrast to the other studies, Fitchett and Fani (2018) did not use climate variables as part of their linear regression models while Grab and Craparo (2011) and Fitchett *et al.* (2019) did use climate variables. Furthermore, the climatic variables in the sardine study by Fitchett *et al.* (2019) were ENSO conditions rather than T_{min}, T_{max}, and T_{mean} used in Grab and Craparo (2011).

Table 2.1. Summary of the phenological changes for South African species (adapted from Mahlangu and Fitchett, 2019).

Species	**Paper**	**Data Used**	**Rate of Advance (day/decade)**	**Province**
Golden Delicious Apple	Grab and Craparo (2011)	Petal drop data from orchards in SW cape	1.9 day/ decade advance	Western Cape
Sayaka Apple	Grab and Craparo (2011)	Petal drop data from orchards in SW cape	1.4 day/ decade advance	Western Cape
Granny Smith Apple	Grab and Craparo (2011)	Petal drop data from orchards in SW cape	1.1 day/ decade advance	Western Cape
Bon Chrétien Pear	Grab and Craparo (2011)	Petal drop data from orchards in SW cape	1.8 day/ decade advance	Western Cape
Jacarandas	Fitchett and Fani (2018)	Social media and newspaper articles	2.38 day/decade advance	Gauteng
Sardines	Fitchett *et al.*, (2019)	Newspaper articles	1.3 day/decade delay	KwaZulu-Natal
Namaqualand daisies	Snyman (2020)	Newspaper articles, Flickr and iNaturalist	2.1 day/decade advance	Northern Cape/ Western Cape

These studies summarised in Table 2.1 are based on only seven species from the multitude of fauna and flora that South Africa is home to. In terms of the locality of these studies, three apple and one pear species were based in the Western Cape province (Grab and Craparo, 2011). The sardine study was specific to KwaZulu-Natal (Fitchett *et al.,* 2019). The study on jacarandas was based on trees in the Gauteng province (Fitchett and Fani, 2018). The study on the Namaqualand daisies was based on wildflowers in the Northern Cape and Western Cape provinces. The required amount of data to make inferences regarding climate change must span at least over 30 years and therefore only a limited number of studies exist as listed inTable 2.1 (Fitchett and Fani, 2018; Snyman, 2020).

2.2 Citizen Science for Species Conservation and Ecological Research

Citizen science is a process whereby volunteers from the public are utilised to collect, categorise, transcribe, or analyse data about species within specific localities over time (Bonney *et al.,* 2009; Bonney *et al.,* 2014). The aims of a number of these initiatives are twofold: firstly, the initiatives assist in advancing scientific knowledge and secondly, it can be used as educational tools for those volunteers who undertake studies in citizen science (Bonney *et al.,* 2009; Cohn, 2008). The process can also provide knowledge for social and environmental change in the communities where the volunteers reside (Bonney *et al.,* 2014).

Under the restraints of traditional data collection methods, citizen science has been a long-established tool to collect data for ecological research (Dickinson *et al.*, 2010; Dickinson *et al.*, 2012; Lopez *et al.*, 2019). Cooper *et al.* (2007) argue that citizen science projects can collect data across spatial and temporal scales where traditional projects may not have the capacity. According to Follett and Strezov (2015), citizen science projects can be categorised into three main types based on the level of the citizen scientist involvement. The first is contributory, which involves the contribution to citizen science projects by collecting data, and in some instances the interpretation of data (Bonney *et al.*, 2009; Follett and Strezov, 2015). The second type is a collaborative citizen science project whereby citizen scientists may help with the design of the project and to interpret data and disseminate results (Follett and Strezov, 2015). The third type of project is co-created where citizen scientists are involved in all phases of the project (Follett and Strezov, 2015).

Citizen science projects face many challenges. The lack of quality control, quality assurance and the general reliability of the data collected is always of concern and is cited as one of the main challenges (Bison *et al.,* 2019; Burgess *et al.,* 2017; Dickinson *et al.,* 2010; Feldman *et al.,* 2018; Kosmala *et al.,* 2016). The data generated through citizen science projects are often difficult to analyse as the effort across volunteers is variable and their willingness to participate may not be uniform temporally (Greenwood, 2007; Isaac *et al.,* 2014). The individual skill levels among volunteers may also differ drastically (Greenwood, 2007). Some projects may also be poorly designed (Tulloch and Szabo, 2012). A well-designed project is essential for keeping volunteers engaged and motivated throughout the process (Tulloch and Szabo, 2012; Wright *et al.,* 2015). Clear protocols, which are easy to follow, play an important role in a project's success (Dickinson *et al.,* 2012). Despite the challenges that citizen science projects face, many advantages do exist. Studies suggest that results from certain citizen science projects are consistent with the results of studies conducted by research professionals (Bison *et al.,* 2019). Citizen science projects collect data at spatial and temporal scales that would otherwise be more challenging for researchers (Dickinson *et al.,* 2010). This is especially true when volunteers participate in several surveys over a long period (Bison *et al.,* 2019). However, citizen science has not been widely accepted as a valid method for conducting research (Bonney *et al.,* 2014). In the past, these projects have often had challenges in getting reviewed for publishing in journals (Bonney *et al.,* 2014).

Some authors have identified data science methods for conducting citizen science-related research which in turn generates 'big data'. Big data is defined as a volume of structured and unstructured data that is so large that more traditional databases cannot handle its storage (Panchariya *et al.,* 2015). Novel ways of collecting big data have been described by several authors. These include using applications on mobile phones (Graham *et al.,* 2011; Liebenberg, 2016; Panchariya *et al.,* 2015), Facebook applications (Sachs and Finin, 2011), and citizen science websites (Kelling *et al*., 2018; Newman *et al.,* 2010). Willi *et al.* (2018) illustrate how the combination of camera trap datasets, in conjunction with Deep Learning and Citizen Science was used to evaluate these large datasets that were generated. Allan and Redden (2017) highlight the risks associated with big data within the field of citizen science, namely data quality, challenges in algorithm investigation accountability and transparency concerns. Callaghan *et al.* (2019) also highlight how the big data generated from these citizen science projects create information and statistical problems that makes analysis of the data very challenging. The engagement levels of citizen scientists on a project can dwindle over time (Greenwood, 2007; Isaac *et al.,* 2014). Panchariya *et al.* (2015) identify the potential of gamification to keep the users of citizen science websites and mobile phone applications engaged in a project.

In South Africa, citizen science projects typically comprise studies on various plant and animal species. Many of these projects are managed by the University of Cape Town's Animal Demography Unit (FitzPatrick Institute of African Ornithology, 2020). Generally, these projects aim to identify the distribution of various species across South Africa and in some cases Africa as a whole (FitzPatrick

Institute of African Ornithology, 2020) as geographic coordinates are captured on the platform. Another purpose is that the platforms serve as repositories for photographs of species (FitzPatrick Institute of African Ornithology, 2020). There have been several formal citizen science research projects published within the South African context. These include early warning systems (Barnard and De Villiers, 2012), roadkill studies (Périquet *et al.*, 2018) and several bird studies including secretary birds (Hofmeyr *et al.*, 2014), African Woolly-necked storks (Thabethe and Downs, 2018), barn swallows (Burman, 2016), and a study looking more generally at the impact of environmental change on bird species (Hofmeyr, 2012). The most recent studies discuss the BirdLasser project (Lee and Nel, 2020; Rose *et al.*, 2020), with Rose *et al.* (2020) specifically dealing with the changes to citizen science contribution under the COVID-19 lockdown.

As described in the preceding paragraphs, there is a variety of citizen science projects. The remainder of this section will focus on butterfly-specific citizen science projects (Table 2.2) focusing on various butterfly species. Citizen science projects were selected to include projects that occur at varying scales in different regions (Table 2.2).

Table 2.2. Summary of the butterfly-related citizen science projects.

Author	Project	Species	Scale
Jue and Daniels, 2014	The Florida Natural Areas Inventory (FNAI)	At-risk butterfly species	Local Florida, United States of America
Wilson *et al*, 2015	Peninsular Malaysia Butterfly Count	Any species	National (Malaysia)
Dennis *et al.*, 2017	The Big Butterfly Count	18 specific species	Regional (United Kingdom)
Prudic *et al.*, 2017	eButterfly	Monarch	Continental (North America)
Schultz *et al.*, 2017	The Western Monarch Thanksgiving Count	Monarch	Continental (North America)
Davis and Howard, 2005; Howard and Davis, 2009; Howard *et al.*, 2010	Journey North	Monarch	Continental (North America)
Kountoupes and Oberhauser, 2008; Prysby and Oberhauser, 2004	The Monarch Larva Monitoring Project (MLMP)	Monarch	Continental (North America)
Loftie-Eaton, 2019	LepiMAP	Any species	Continental (Africa)
Ryan *et al.*, 2019	Project Pieris	Pieris *rapae*	Global

There are several methods in which projects are conducted and data are collected for each of the butterfly specific citizen science projects. For eButterfly and LepiMAP, participants upload observations, which include date and location information and optional photographs, to the respective platforms (eButterfly, 2020; Loftie-Eaton, 2019). The Western Monarch Thanksgiving Count requires participants to choose a specific overwintering site in which to count Monarch butterflies (Schultz *et al.*, 2017; Xerces Society for Invertebrate Conservation, 2017). Another two projects focusing on the Monarch butterfly are the Journey North (Journey North, 2020) and The Monarch Larva Monitoring Project (MLMP) (Davis and Howard, 2005; Howard and Davis, 2009; Howard *et al.*, 2010; Kountoupes and Oberhauser, 2008; Prysby and Oberhauser, 2004). The Journey North requires participants to record the date and location of roosting Monarch butterflies during fall or autumn (Davis and Howard, 2005; Howard and Davis, 2009; Howard *et al.*, 2010; Journey North, 2020), whereas the MLMP project includes other focus areas on Monarch life stages and events in summer, spring and winter. In contrast to The Western Monarch Thanksgiving Count and The Journey North, MLMP focuses specifically on the species during the larval stages of the butterfly's life cycle (Kountoupes and Oberhauser, 2008; Xerces Society for Invertebrate Conservation, 2017). The project also focuses on the Milkweed habitat as this is the larval host plant (Kountoupes and Oberhauser, 2008; Prysby and Oberhauser, 2004). The project was unique in that the researchers were also trying to gain insights into the citizen science research process (Prysby and Oberhauser, 2004). The participants collected spatial and temporal information of monarch egg and larval abundances for the Northeast and Upper Midwest parts of the United States of America (Prysby and Oberhauser, 2004). The Big Butterfly Count is a butterfly project in the United Kingdom (Dennis *et al.,* 2017). Participants are required to count 18 butterfly species during July and August under sunny weather conditions each year (Dennis *et al.,* 2017). The Florida Natural Areas Inventory (FNAI) is another example of a butterfly survey specific to Florida in the United States of America (Jue and Daniels, 2014). FNAI capitalised on Florida based butterfly enthusiasts who already spent time weekly observing butterflies to compile a database of locations of at-risk butterflies in the region (Jue and Daniels, 2014). A unique project is Project Pieris, which is an international citizen science project that focuses specifically on *Pieris rapae* (Pieris Project, 2020). This project does genetic sequencing on mitochondrial deoxyribonucleic acid (DNA) on the butterfly samples that are collected from across the world (Ryan *et al.*, 2019). However, there is poor representation of Africa, India, and South East Asia in this study. Antarctica and South America are also excluded; the authors justify this exclusion claiming that *Pieris rapae* has not been introduced to these regions (Ryan *et al.*, 2019). According to GBIF (2020b), records on the western coast of South America have been documented, albeit not as abundantly as the other areas. The concentrations of data are from North America and western Europe which may have resulted in skewed and incomplete results (Ryan *et al.*, 2019). A study like Ryan *et al.* (2019), using butterfly DNA, was conducted by Wilson *et al.* (2015) on the Peninsular Malaysia Butterfly Count. In contrast to the work of Ryan *et al.* (2019), the process was non-lethal and instead of killing the butterfly, a sample of the butterfly leg was taken to extract DNA (Wilson *et al.*, 2015). The study is one of the only other examples of research published outside of North America.

A majority of the butterfly citizen science projects do not have any specific selection processes or criteria for contributing to a project (Dennis *et al*., 2017; Jue and Daniels, 2014; Kountoupes and Oberhauser, 2008; Loftie-Eaton, 2019; Prudic *et al.*, 2017; Prysby and Oberhauser, 2004; Ryan *et al*., 2019; Schultz *et al.*, 2017; Wilson *et al.*, 2015). There are, however, instances where some training is required (Prysby and Oberhauser, 2004; Xerces Society for Invertebrate Conservation, 2017) with only one project requiring expert knowledge (Jue and Daniels, 2014). The Big Butterfly Count, eButterfly and LepiMAP do not require a specific participant selection process other than users having to register on the platform (Dennis *et al.*, 2017; Loftie-Eaton., 2019; Prudic *et al.*, 2017). The Pieris Project and The Journey North do not have a selection process for participants either (Davis and Howard, 2005; Howard and Davis, 2009; Howard *et al.*, 2010; Ryan *et al.,* 2019). While the Western Monarch Thanksgiving Count and The Monarch Larva Monitoring Project (MLMP) do not have specific participant selection procedures, they do require that participants undergo training prior to undertaking data collection (Prysby and Oberhauser, 2004; Xerces Society for Invertebrate Conservation, 2017). The MLMP project involved youth and adult participants that were recruited via various websites, including other citizen science projects (Prysby and Oberhauser, 2004). The Peninsular Malaysia Butterfly Count recruited participants from Facebook with no selection process but provided the participants with a butterfly count guide (Wilson *et al.*, 2015). Similarly, participants for the Big Butterfly Count were provided a downloadable butterfly identification chart or had access to a mobile application to identify butterflies and record the total numbers of observations (Big Butterfly Count, 2020). One of the only studies with specific participant selection criteria is FNAI where coordinators, who had expert skills and a reputation in butterfly identification, were recruited to manage the surveys (Jue and Daniels, 2014). These coordinators were responsible for recruiting and training volunteers to conduct the survey (Jue and Daniels, 2014).

As the projects differ in nature, so too do the quality checks used by researchers in the studies. The data submitted by participants for both eButterfly and LepiMAP is vetted by regional experts to ensure a high standard of data quality (Loftie-Eaton, 2019; Prudic *et al.*, 2017). A similar vetting process is conducted by staff members at the Journey North project (Howard and Davis, 2009), and for MLMP (Prysby and Oberhauser, 2004) whereby the researchers verified the data before input into the project database. Minimal verification is undertaken for the Big Butterfly Count (Dennis *et al.*, 2017). FNAI's approach to ensuring quality data requires photographs to be taken of rare butterfly species that were investigated as part of this study (Jue and Daniels, 2014). It is not apparent what quality checks are in place for the Western Monarch Thanksgiving Count citizen science project. However, Shultz *et al.* (2017), who used this dataset as a source for their study, conducted their own quality check to test if the variance in observation error is constant over time. The Pieris Project focuses on the integrity of the specimens supplied to the researcher by providing guidelines on handling butterfly specimens on their website (Pieris Project, 2020). In terms of quality of the data being analysed, the researcher ensured that they used multiple simulations and identified estimates that were most common (Ryan *et al.*, 2019). The Peninsular Malaysia Butterfly Count had a butterfly count guide document which was the only resource given to participants to ensure quality data was collected (Wilson *et al.*, 2015).

Even with this guide document, the project experienced poor species identification and poor sample handling (Wilson *et al.*, 2015).

To determine the success of each of these butterfly-related citizen science projects, scientific goals are used as a criterion to determine a citizen science project's success, as detailed by Golumbic's (2015) 15 characteristics of a successful citizen science project. In this review, the success of a citizen science project lies in its ability to provide data that can contribute to meaningful academic understanding of the butterflies in question. For example, the results of Shultz *et al.* (2017) on the Western Monarch Thanksgiving Count suggests that the data shows a dramatic decline in Monarch populations over the 36-year period and therefore the project was deemed successful. Prudic *et al*., (2017), in contrast with the other types of quantitative research discussed, was more a descriptive study describing the data being collected by the eButterfly programme.

In terms of its contribution on focusing conservation effects, the Peninsular Malaysia Butterfly Count, Journey North and FNAI were deemed successful (Howard and Davis, 2009; Jue and Daniels, 2014; Wilson *et al.*, 2015). In addition, the data from Journey North facilitated an understanding of the flyways that monarchs use during their migration to Mexico (Howard and Davis, 2009), the collective understanding of the overwintering stage of the monarch life cycle (Howard *et al.*, 2010) as well as insights into the temporal patterns of the monarch migrations over time (Davis and Howard 2005). The data from MLMP successfully assisted in understanding the spatial and temporal variability of monarch egg and larvae across several parts of the United States of America (Prysby and Oberhauser, 2004). Project Pieris facilitated the understanding of how the European and Asian subspecies of *Pieris rapae* have diverged across the globe both in terms of timing and spread (Ryan *et al.*, 2019). Ryan *et al.* (2019) further suggest that the divergence was facilitated by certain crops, human movement, and trade of the past century. Ryan *et al.* (2019) assert that understanding of the species is important as the species is regarded as a destructive pest to some crops. The Peninsular Malaysia Butterfly Count provided insights into the species composition and distribution across the region (Wilson *et al.*, 2015). The data generated by the project also provided a baseline from which future monitoring initiative can be conducted (Wilson *et al.*, 2015). From the results obtained by FNAI, the study was able to identify the most at-risk butterflies in the Florida region (Jue and Daniels, 2014). The Big Butterfly Count was shown to produce results comparable to the traditional monitoring methods used by the UK Butterfly Monitoring Scheme (Dennis *et al.,* 2017).

None of the data collection methodologies discussed in the above-mentioned studies are comparable to traditional data collected by scientists. Data collected by scientists as part of monitoring schemes include counting butterflies as part of transects (Dennis *et al*., 2017; van Swaay *et al.,* 2008). The UK Butterfly Monitoring Scheme (UKBMS), for instance, requires that over a thousand transects are monitored weekly during April and September each year (Dennis *et al.,* 2017). These traditional data collection methodologies are labour-intensive processes (Dennis *et al.,* 2017).

2.3 Social Media as a Data Source

Several applications exist for the usage of social media data in research. There is an abundance of studies using Twitter and Instagram data. Examples of these studies include its use to predict S&P 500 stocks (Mao *et al.*, 2012), to forecast migration trends of people (Zagheni *et al.,* 2014) and to track transportation incidents (Mai and Hranac, 2013). Data derived from social media is also used in psychological research (Ferwerda *et al.*, 2015; Lup and Rosenthal, 2015; Pittman and Reich, 2016; Reece and Danforth, 2017; Sheldon and Bryant, 2016; Sheldon and Newman, 2019) and in other medical research such as cardiovascular disease (Sinnenberg *et al.,* 2016), dermatology (Achrekar *et al*., 2011; Karimkhani *et al.*, 2014), healthy food references (Widener and Li, 2014) and flu predictions (Lampos and Cristianini, 2010). Data from social media can also be found in studies focusing on tourism (Fatanti and Suyadnya, 2015; Sotiriadis and van Zyl, 2013; Sin and He, 2018) and politics (Barberá, 2013; Jacobs *et al.,* 2020; Wang *et al.,* 2012).

There have been several applications of social media data in the earth and environmental sciences. From 2011 onward, Twitter has been more frequently utilised as a data source for earth science projects (Steiger *et al.*, 2015), in disaster and emergency management (Chae *et al.,* 2012; Crooks *et al*., 2013; De Longueville *et al.,* 2009; Demuth *et al*., 2018; Earle *et al.,* 2012; Gomide *et al.*, 2011; Kireyev *et al.,* 2009; Lacassin *et al*., 2020; Martín *et al.*, 2020; MacEachren *et al.*, 2011; Murthy and Longwell, 2013; Sakaki and Matsuo, 2010; Stefanidis *et al.,* 2011; Terpstra *et al.*, 2012), flood prevention (Brangbour *et al*., 2019; Kankanamge *et al.*, 2020; Smith *et al.,* 2017), monitoring protected areas (Tenkanen *et al.*, 2017), conservation (Baksh, 2019; Carew, 2014; Di Minin *et al.,* 2015) and citizen science (Daume and Galaz, 2016).

The remainder of this subsection will focus specifically on a few ecological studies which utilise social media as a data source. These studies are summarised in Table 2.3.

Table 2.3. Summary of ecological studies using social media as a data source.

Authors	Focus Area	Social Media Platform Used
ElQadi *et al.,* 2017	Bees and flowering plants	Flickr
Zipper, 2017	Agriculture	Twitter
Fitchett and Fani, 2018	Plant phenology	Twitter, Instagram, and others
Hart *et al.*, 2018	Animal phenology	Twitter
Silva *et al.,* 2018	Vegetation phenology	Twitter
Breckheimer *et al.,* 2019	Wildflower	Flickr
Marshall and Strine, 2019	Snake Distribution	Flickr

The academic articles reviewed in this subsection (Table 2.3) included data derived from three social media platforms including Flickr (Breckheimer *et al.*, 2019; ElQadi *et al.*, 2017; Marshall and Strine, 2019), Twitter (Fitchett and Fani, 2018; Hart *et al.*, 2018; Silva *et al.,* 2018; Zipper, 2017) and Instagram (Fitchett and Fani, 2018). Flickr has been used as a data source in Australian research focusing on bees, flowers, and wildflower observations in Mount Rainier National Park (Breckheimer *et al.,* 2019) and global snake distributions (Marshall and Strine, 2019). Twitter has been used for plant (Breckheimer *et al*., 2019; Fitchett and Fani, 2018; Silva *et al.,* 2018) and animal (Hart *et al.,* 2018) phenology and agricultural research (Zipper, 2017).

Data were collected predominantly utilising the Application Programming Interface (API) on the social media platforms for either Flickr (Breckheimer *et al.,* 2019; ElQadi *et al.*, 2017; Marshall and Strine, 2019) or Twitter (Hart *et al.*, 2018; Silva *et al.,* 2018; Zipper, 2017). The only difference between the data collection methods was the applications used to connect to the API. The tools to collect data from the Flickr API was Python-Flickr-API (ElQadi *et al.*, 2017), and the taxize package in R (Marshall and Strine, 2019). Breckheimer *et al.* (2019) does not specify what exact applications were used to connect to the Flickr API in their study. Zipper (2017) accessed the Twitter API using the Twitter package in R. Hart *et al.* (2018) used proprietary software by the company FollowtheHashtagcom to collect their data. Similarly, Breckheimer *et al.* (2019), Silva *et al.* (2018) do not mention which application was used to connect to the Twitter API for their study. The study by Fitchett and Fani (2018) was the only investigation to use manual methods to collect data from the Twitter and Instagram social media platforms.

Various analysis methods were used, based on the type of study that was being conducted. Silva *et al.* (2018) analysed images of national parks from Twitter from an image analysis technique using the mean greenness index. Marshall and Strine (2019) spatially analysed their data using the spatstat package in R. ElQadi *et al.* (2017) also used a form of spatial analysis by overlaying and buffering the data from social media in relation to those from the Global Biodiversity Information Facility (GBIF). Fitchett and Fani (2018) used a linear regression model to determine the phenological shift in the flowering phenology of jacarandas while Hart *et al.* (2018) used linear regression models for the spatial analysis of the winged ant emergence. Breckheimer *et al.* (2019) used a generalized non-linear mixed-effects model to analyse flower phenology and an additional regression to model visitors by using a binomial distribution and a log link. The most unique method of analysis was comparing cumulative tweets to that of traditionally collected data using the mean absolute difference (Zipper *et al.,* 2017).

The data compiled from these social media records are unlike many of the traditional sources used for this type of research. Planting data is usually collected using labour-intensive surveys or meetings (Zipper, 2017). Phenological research is also labour intensive as it is usually conducted through ground-based observations (Fitchett *et al.,* 2015). Information on species distributions is usually based on a combination of ground-based observations and statistical models (Chuine, 2010). Social media is said to be a cost-effective way to obtain large amounts of data at various spatial scales (Hart *et al.*, 2018; Zipper, 2017). Yet, the use of social media in these studies is not without risks as data may be unreliable (ElQadi *et al.,* 2017). The data is also prone to poor species identification (Marshall and Strine, 2019). Some data sources may not be suitable for all species of study given the differences in results for winged ants, spiders and starling murmurations (Hart *et al.*, 2018). There is generally a paucity of spatial data (Hart *et al.,* 2018), including that of geotagged imagery (Breckheimer *et al.*, 2019; Marshall and Strine, 2019).

This literature review has discussed three of the main themes related to this research topic. The phenology subsection discussed related research and provided an overview of phenological studies conducted in South Africa. The citizen science subsection broadly discussed citizen science research that is being conducted and focused specifically on butterfly-related citizen science projects. The final subsection focused on the use of social media as a data source. It provided details of some of the popular applications of social media data and applications to ecological research.

Chapter 3: Study Area

South Africa is the southernmost country on the African continent (Figure 3.1), spanning the coordinates of 22-35° S and 17-33° E (Fitchett and Hoogendoorn, 2018; Government of South Africa, 2020). The surface area is approximately 1,220,813 km². The topography consists of an interior plateau, surrounded by The Great Escarpment. The country comprises nine provinces which are home to 19 national parks (Government of South Africa, 2020). This study considered any mention of *Belenois aurota* on Twitter, Instagram and Facebook that was known to have originated in South Africa. Therefore, the entire country forms part of the study area.

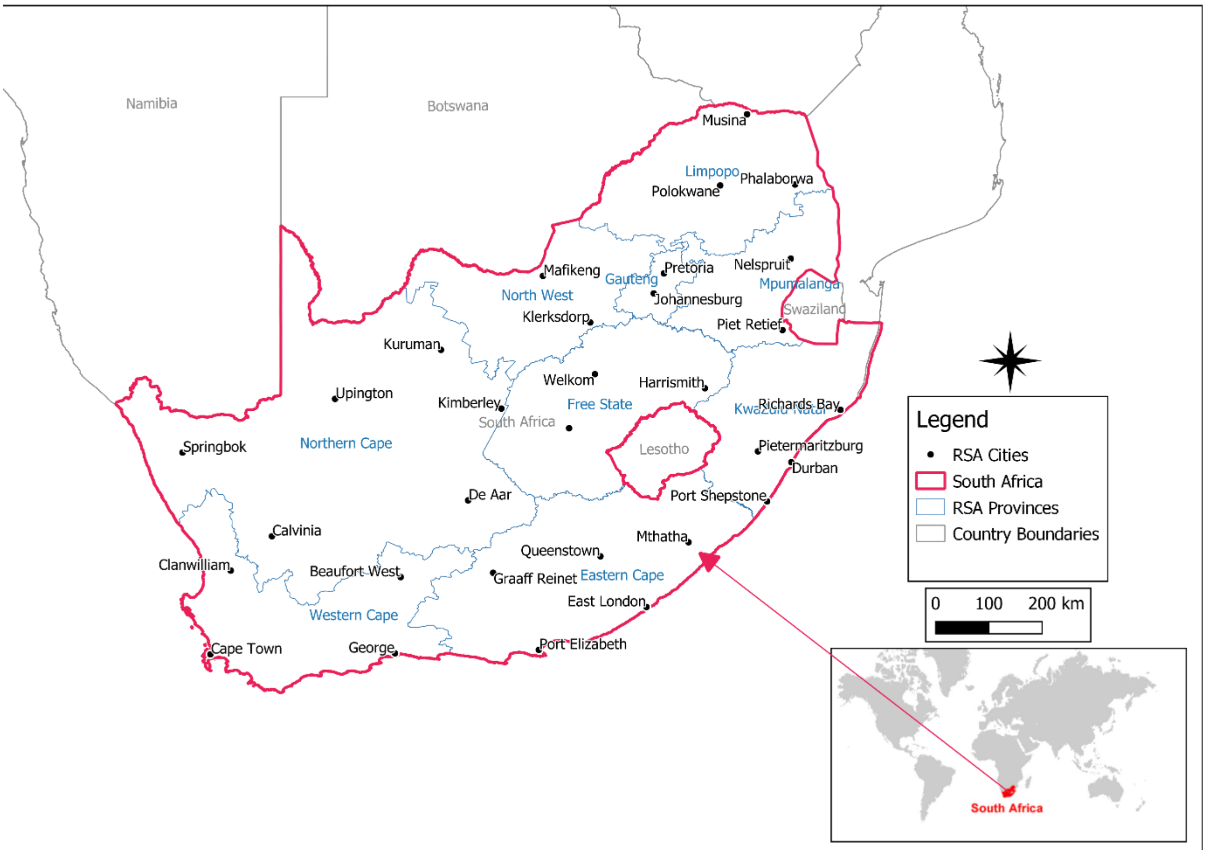


Figure 3.1: Map depicting the study area, South Africa (RSA).

South Africa comprises ten main biomes (Mucina and Rutherford, 2006): the Albany thicket, Azonal vegetation, Desert, Forests, Fynbos, Grassland, Indian Ocean coastal belt, Nama Karoo, Savanna and

Succulent Karoo. The delineations of these biomes are depicted in Figure 3.2. *Belenois aurota* are said to occur in all these biomes throughout the country (Woodhall, 2012; 2020).

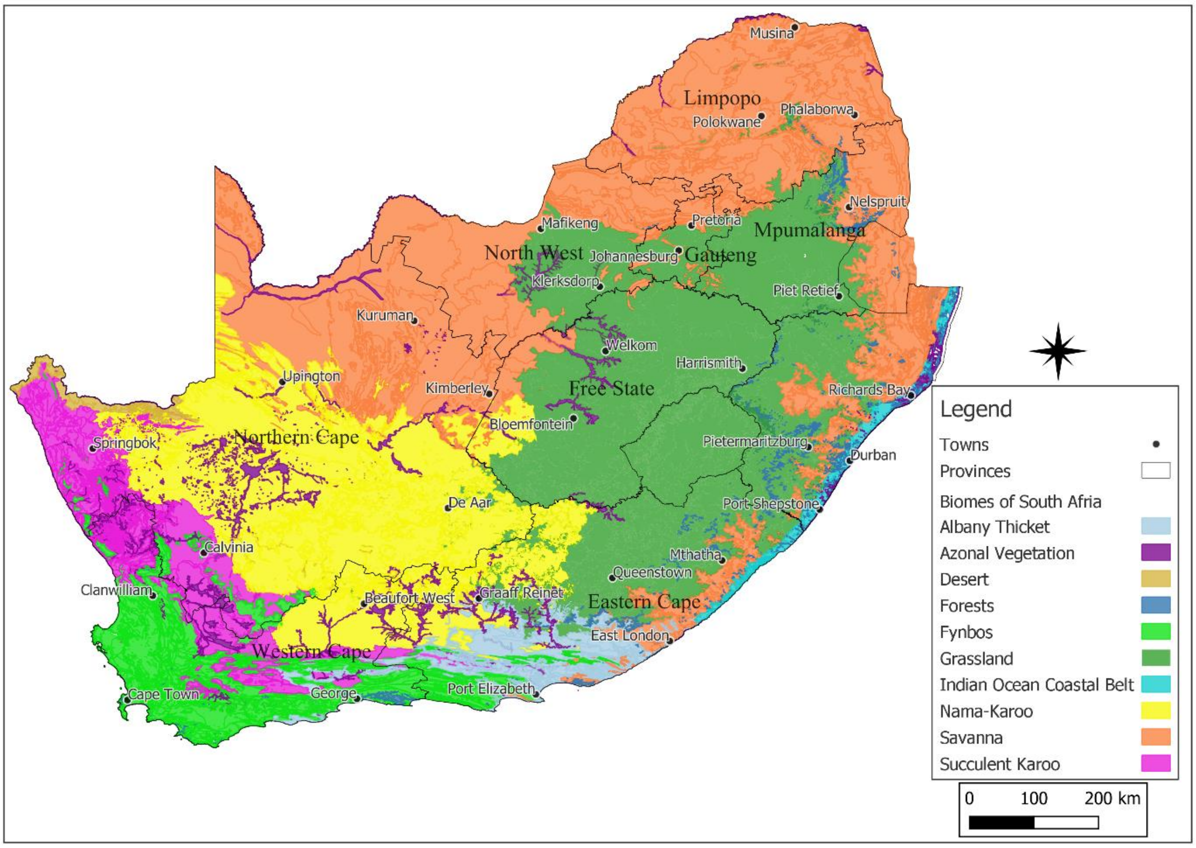


Figure 3.2: Study area showing the main biomes of South Africa (South African National Biodiversity Institute, 2006-2018).

3.1 Climate of South Africa

The climate of South Africa is influenced by its varied topography, subtropical latitude, urban heat islands and ocean currents (Fitchett and Hoogendoorn, 2018; Jury, 2018). The warm Agulhas current occurs in the east coast Indian Ocean while the cold Benguela current is found in the Atlantic Ocean on the west coast of the country (Fitchett and Hoogendoorn, 2018). South Africa has a mean annual rainfall of 464 mm (Government of South Africa, 2020). The southwestern portion of the country receives most of its rainfall in the winter because of the Inter-Tropical Convergence Zone shifts, while the south coast receives year-round rainfall (Roffe *et al.*, 2019, 2020). The remaining interior and eastern and western coasts of the country receives predominantly summer rainfall (Fitchett and Hoogendoorn, 2018; Roffe *et al.,* 2019, 2020). The temperatures on the interior plateau during summer remain below 30°C (Government of South Africa, 2020). Figure 3.3 depicts the climate of South Africa as classified by the Köppen-Geiger climate classification system. It shows the various arid and warm temperate climates that South Africa comprises (Kottek *et al.*, 2006; Engelbrecht and Engelbrecht, 2016). The arid parts of the country include all parts of Northern Cape (BWh, BWk, BSk, and BSh) (Engelbrecht and Engelbrecht, 2016; Kottek *et al.,* 2006). Warm temperate parts of the country include the eastern coastline part of KwaZulu-Natal and Eastern Cape (Cfa, Cfb) as well as the interior northwestern KwaZulu-Natal, eastern Free State, Gauteng, Mpumalanga and eastern parts of Limpopo and North West province (Cwa, Cwb) (Engelbrecht and Engelbrecht, 2016; Kottek *et al.,* 2006). The southwestern part of the Western Cape also has a warm temperate climate (Csa, Csb) (Engelbrecht and Engelbrecht, 2016; Kottek *et al.,* 2006).

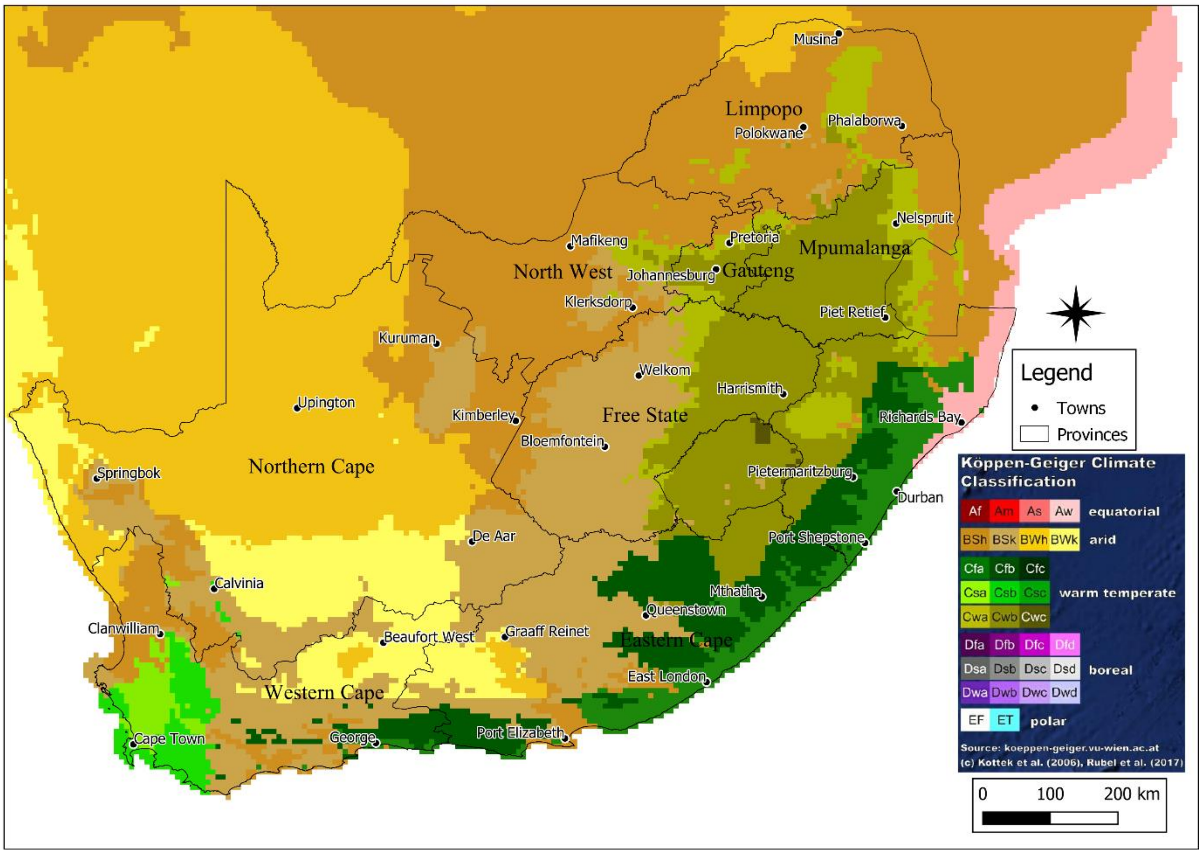


Figure 3.3: Köppen-Geiger climate classification map (Kottek *et al.,* 2006).

3.2 Social media usage in South Africa

With an increase in internet accessibility in the country over the years, more South Africans are making use of the various available social media platforms (Barr, 2020; Kemp, 2020). The internet penetration is approximately 37 million internet users across the country, increasing by 3.1% each year (Barr, 2020; Kemp, 2020). Of these internet users, 94% have access to a smartphone, 76% have access to a laptop or desktop computer and 43% have access to a tablet device (Kemp, 2020). Users accessing the internet via their mobile device amount to 96% with the entire 96% also report using social media mobile applications. 98% of active social media users are accessing social media via a mobile device (Figure 3.4). According to data from January 2020, users of all social media platforms amount to 22 million (Figure 3.4) from a population of 59 million (Barr, 2020; Kemp, 2020). On average, users have more than eight social media accounts across various platforms (Kemp, 2020). It is estimated that these users spend on average over three hours per day using social media (Kemp, 2020).

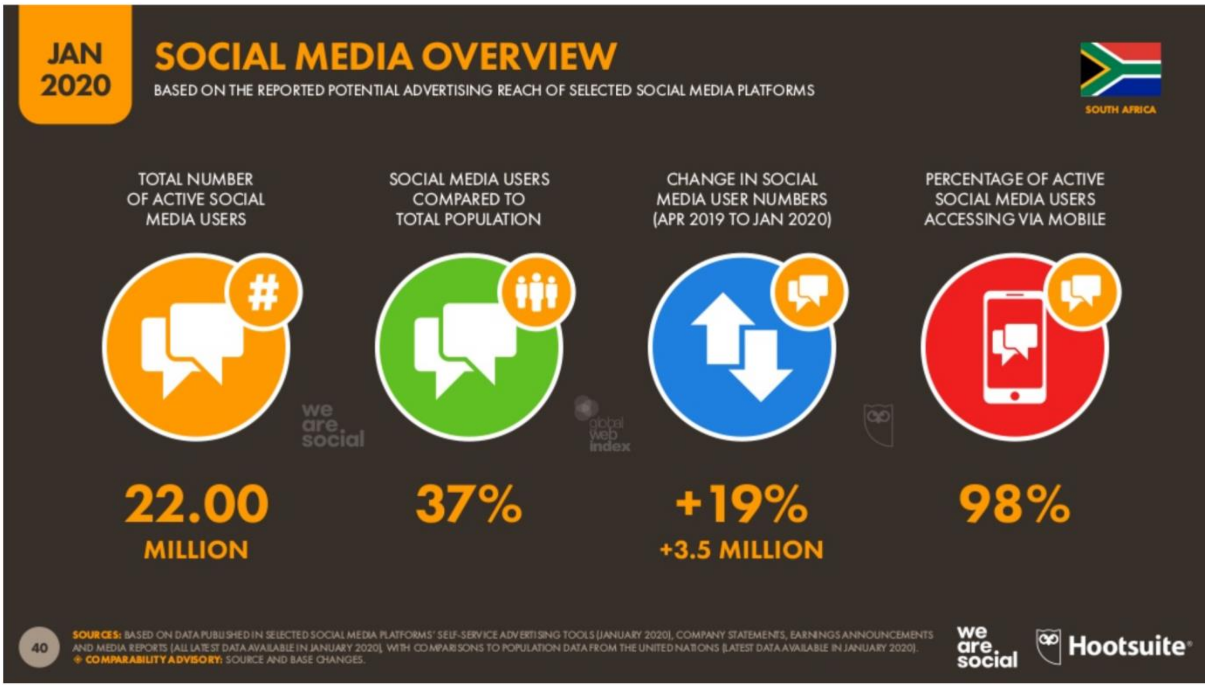


Figure 3.4: Social media overview for South Africa (Kemp, 2020).

16 popular social media platforms are currently being used in South Africa (Fig 3.5). WhatsApp, YouTube, and Facebook are the most popular platforms (Fig 3.5) used by South Africans (Kemp, 2020). Instagram and Twitter are also popular platforms but less so than Facebook (Kemp, 2020).

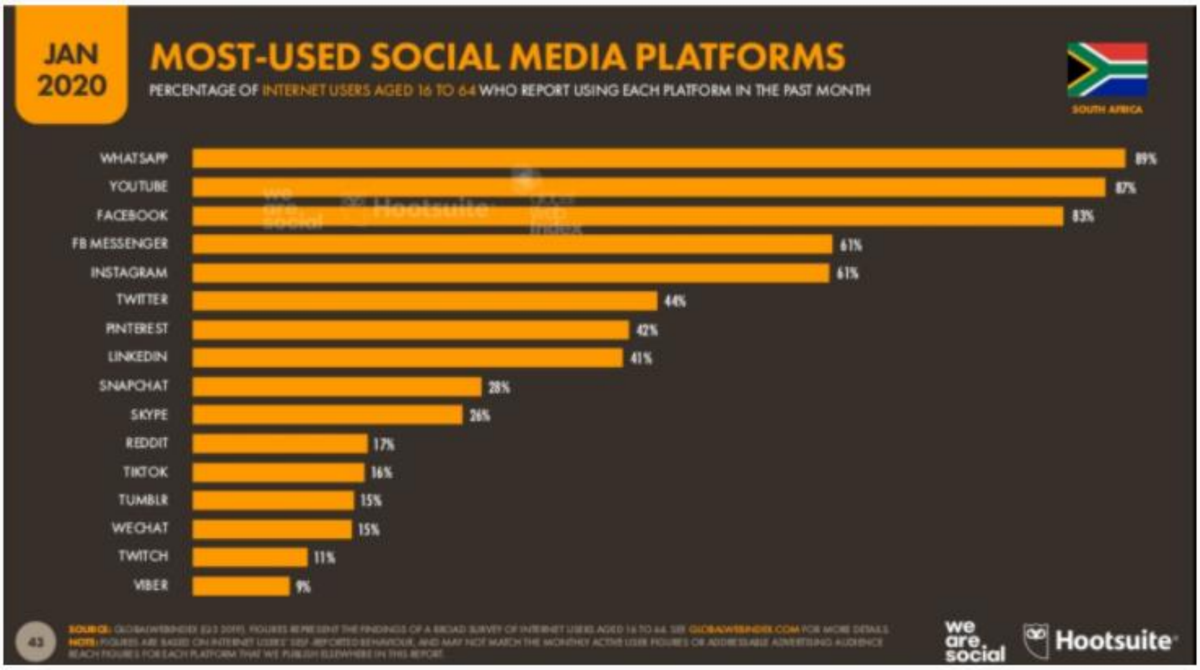


Figure 3.5: Most-used social media platforms in South Africa (Barr, 2020).

Many of the users on social media platforms are between the ages of 25 and 34 and are followed closely by the 18-24 age category (Figure 3.6). These two age groups also exhibit overall equal numbers of male and female users which is unlike any of the other groups where some variation in gender is seen (Barr, 2020).

When focused specifically on the platforms of interest for this study, Facebook usage is consistently higher than Twitter and Instagram between November 2019 and November 2020 (Fig. 3.7). Twitter and Instagram usage declined from April 2020 (Fig 3.7), which coincidentally, is when the nationwide lockdown began in South Africa. Facebook shows an increase in usage for the same period (Fig. 3.7)

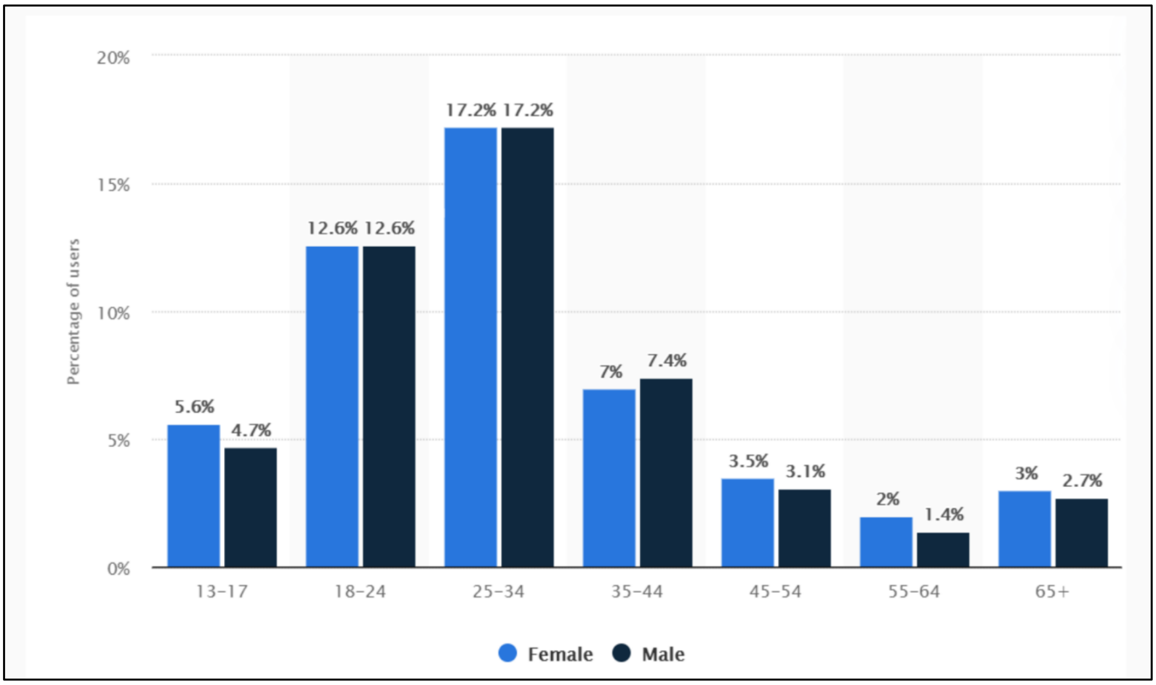


Figure 3.6: Composition of social media users in South Africa (Barr, 2020).

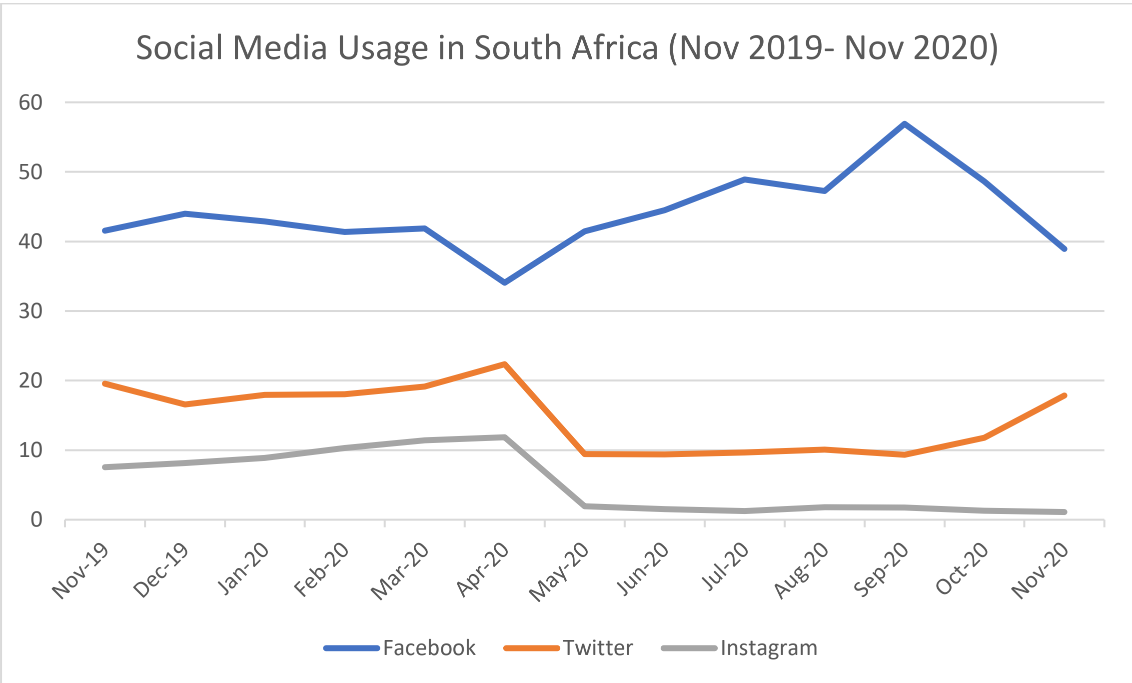


Figure 3.7: Social media usage in South Africa (Barr, 2020).

According to data from August 2020, Facebook had just under 22 million users in South Africa (Figure 3.8). Slightly more than half of the total Facebook users are female. The 25-34 age group (Fig. 3.8) dominates the usage, followed by the 18-24 and 35-44 age groups (Barr, 2020).

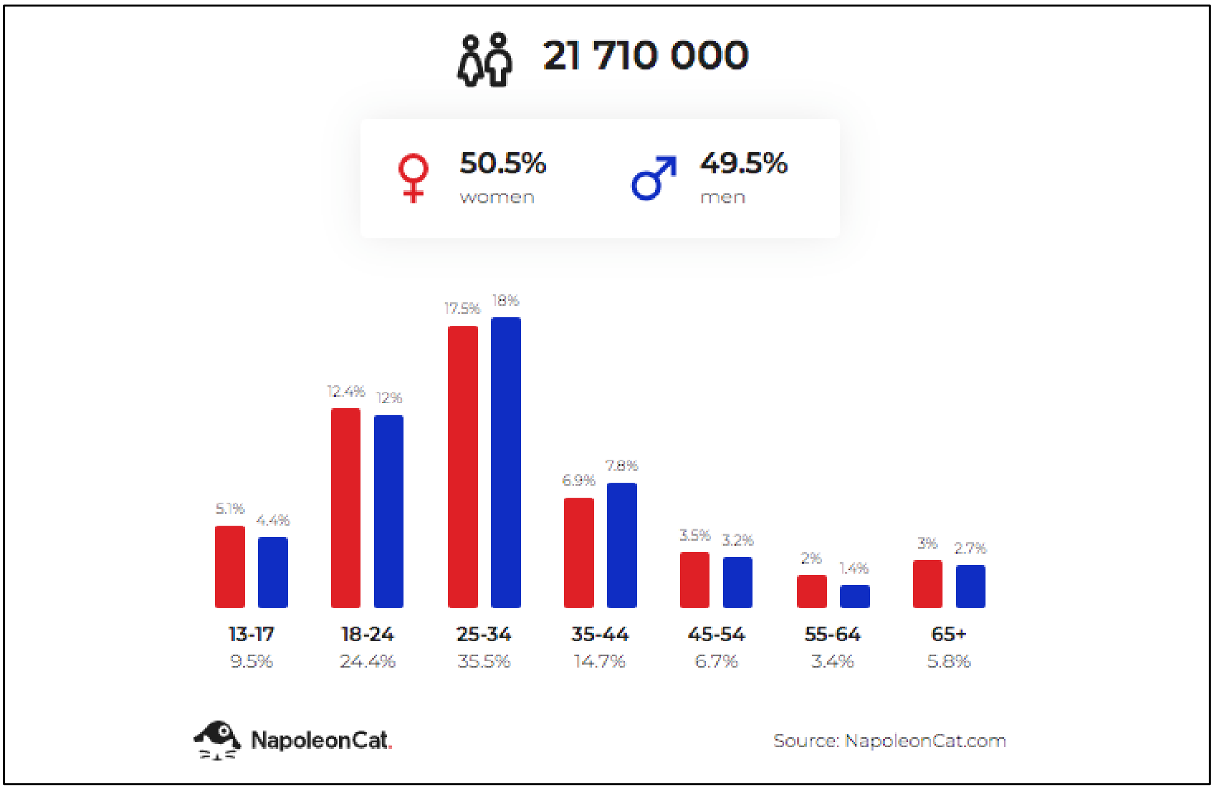


Figure 3.8: Composition of Facebook users in South Africa (Jansen van Rensburg, 2020).

According to the same data published in August 2020, there were over 5 million Instagram users in South Africa (Fig. 3.9), of which 53.1% are female (Fig. 3.9). The majority of Instagram users fall between the ages of 18-24 and 25-34 years old (Fig. 3.9).

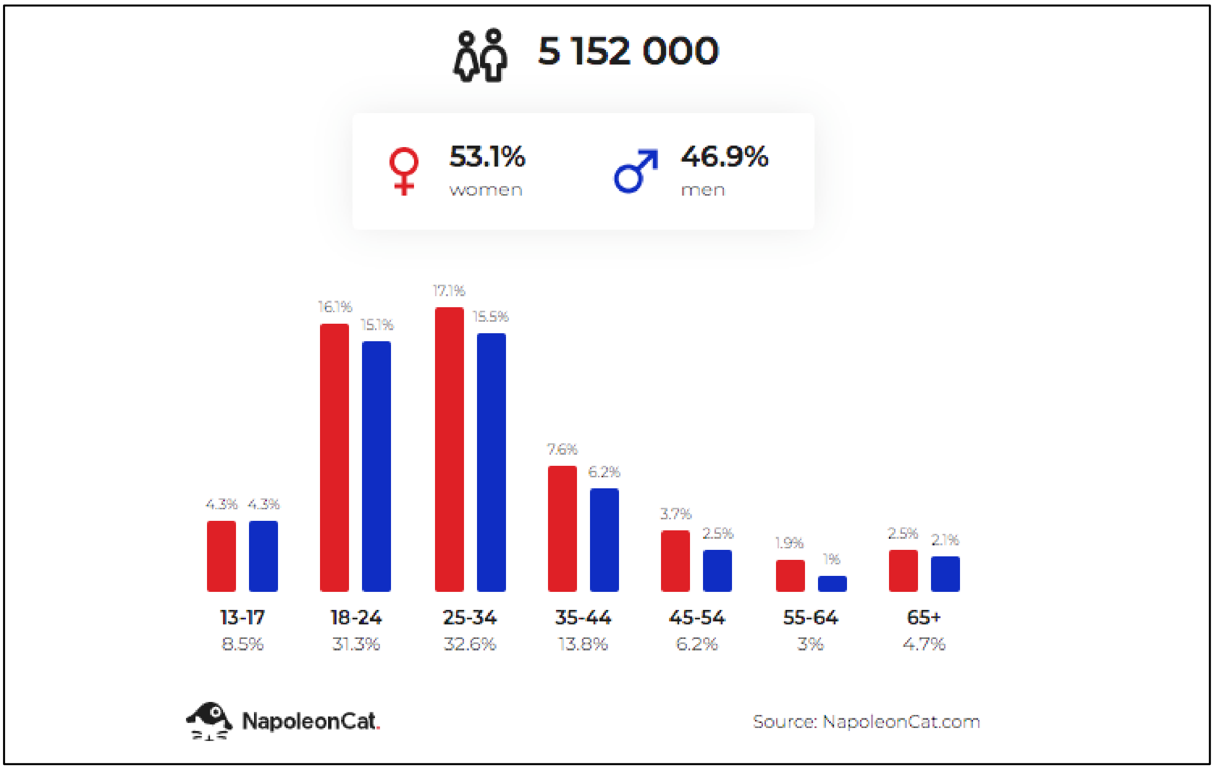


Figure 3.9: Composition of Instagram users in South Africa (Jansen van Rensburg, 2020).

No Twitter user data could be found for the same period as Facebook and Instagram. However, according to data published in January 2020, there were 2.3 million Twitter users in South Africa (Kemp, 2020). Users comprise 38.3% females and 61.7% males (Kemp, 2020).

3.3 Butterflies of South Africa

South Africa has 671 known species of butterflies (Woodhall, 2020). These butterflies are divided into five main families namely Nymphalidae, Hesperiidae, Lycaenidae, Pieridae and Papilionidae (Willis and Woodhall, 2010; Woodhall, 2012; 2020). These five families are also further broken down into several subfamilies (Willis and Woodhall, 2010; Woodhall, 2012; 2020). A large proportion of these butterflies are endemic to South Africa (Woodhall, 2012; 2020).

3.3.1 Main proponents of the Kgalagadi Butterfly Migration and other species

One of the most common species that occurs in all regions of the country and all national parks is *Belenois aurota* (Fig 3.10), which belongs to the Pieridae family (Terblanche, 2020; Willis and Woodhall, 2010; Woodhall, 2012; 2020). Similarly, *Catopsilia florella* (Fig 3.14) also belonging to the Pieridae family and can be found in all regions of the country and all national parks (Willis and Woodhall, 2010; Woodhall, 2012; 2020). Neither of these two butterflies are endemic to the country (Woodhall, 2012; 2020). The migration of *Belenois aurota* occurs annually between December and February, originating in the Northern Karoo and Kalahari (Terblanche, 2020; Woodhall, 2012) and is referred to as the 'Kgalagadi Butterfly Migration' (Terblanche, 2020). The two butterfly species are the main proponents of the Kgalagadi Butterfly Migration (Terblanche, 2020).

Belenois aurota is a white butterfly which occurs across South Africa and is particularly visible during migrations (Woodhall, 2012;2020). *Belenois aurota* is commonly referred to in South Africa as the Brown-veined White butterfly or Pioneer Caper White (Terblanche, 2020; Woodhall, 2012; 2020). Male specimens (Fig 3.10) have dark black or brown markings while female specimens (Fig 3.10) have dark marginal borders much broader than their male counterparts (Woodhall, 2012;2020). *Belenois aurota* originate in the northern parts of the Kalahari in the upper Nama Karoo (Terblanche, 2020).

Figure 3.10. Images showing male and female *Belenois aurota* (Willis and Woodhall, 2010).

As with other butterflies, *Belenois aurota* has 4 stages of its lifecycle that is egg, larva, pupa, and adult stages (Fig 3.11). The time spent in each of these stages is not well documented for this species. In the Kalahari, the adult female butterfly lays eggs on the Shepherds tree *(Boscia albitrunca)* which the caterpillars feed on in the larval stage. In other areas, several plants from the Caper family such as *B. oledides*, *Maerua angolensis*, *M. cafra*, *Capparis spp*, and *C. sepiaria,* are hosts on which the adult female butterfly lays eggs and caterpillars feed on (Terblanche, 2020; Woodhall, 2012;2020). Adult butterflies can be found feeding on flowers or damp mud (mud-puddling) or sand (Willis and Woodhall, 2010). Adult butterflies may rest during wet or chilly weather in tall grasses and this behaviour is called roosting (Terblanche, 2014).

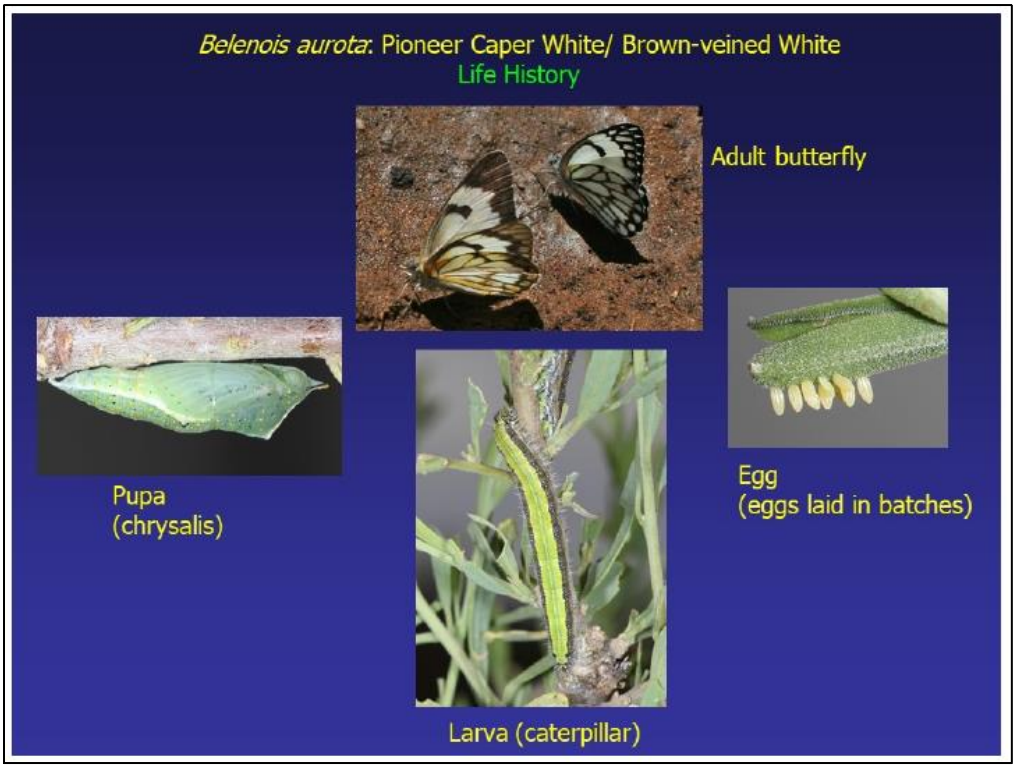


Figure 3.11. Images showing the life stages of *Belenois aurota* (Terblanche, 2020).

The species is also found over a wide area in Africa, the Middle East, India, and Sri Lanka (GBIF, 2020). Figure 3.12 illustrates the known global distribution of the species over these areas. There is a concentration of the species in southern and eastern Africa and occurrences have also been recorded in central and west Africa and Madagascar (Fig. 3.12). In South Africa, the butterfly is distributed across the entire country (Fig 3.12). The social media data that was derived for this study will follow a similar national distribution.

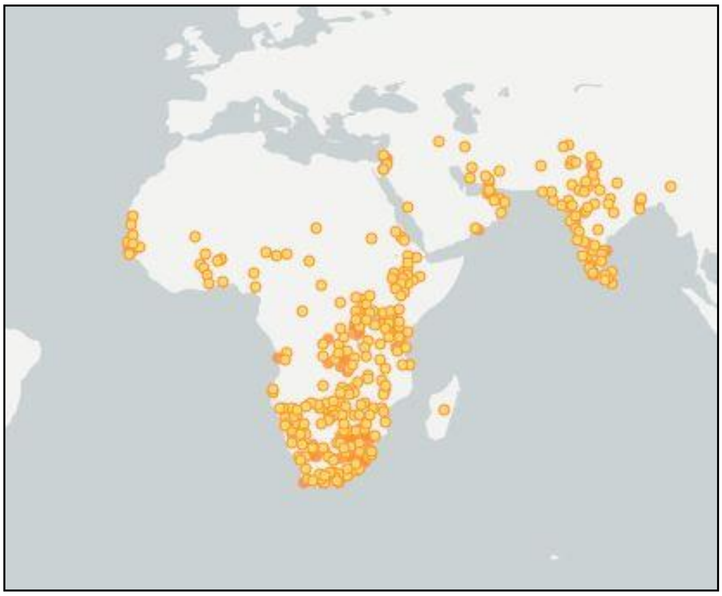


Figure 3.12: Distribution of *Belenois aurota* over Africa, Middle East, India, and Sri Lanka (GBIF, 2020).

Belenois aurota travel over long distances from the colonies where they originate (Terblanche, 2020; Woodhall, 2012; 2020). These butterflies, however, do not make multiple journeys between regions due to their short lifespans and therefore this is not a true migration but rather an emigration (Terblanche, 2020; Woodhall, 2012; 2020). The migration is triggered by overcrowding (Terblanche, 2020; Willis and Woodhall, 2010). This migration is triggered by increased rainfall in the Karoo and Kalahari, resulting in the growth and abundance of Shepherd's trees (Terblanche, 2014;2020; Willis and Woodhall, 2010). This facilitates the increase of the butterfly population as there are more plants on which to feed (Terblanche, 2014;2020; Willis and Woodhall, 2010). When the population grows beyond the capacity of food supply in the Kalahari, the butterflies are forced to migrate north-eastwards (Fig. 3.13) to find plants on which to feed (Terblanche, 2020; Willis and Woodhall, 2010; Woodhall, 2012).

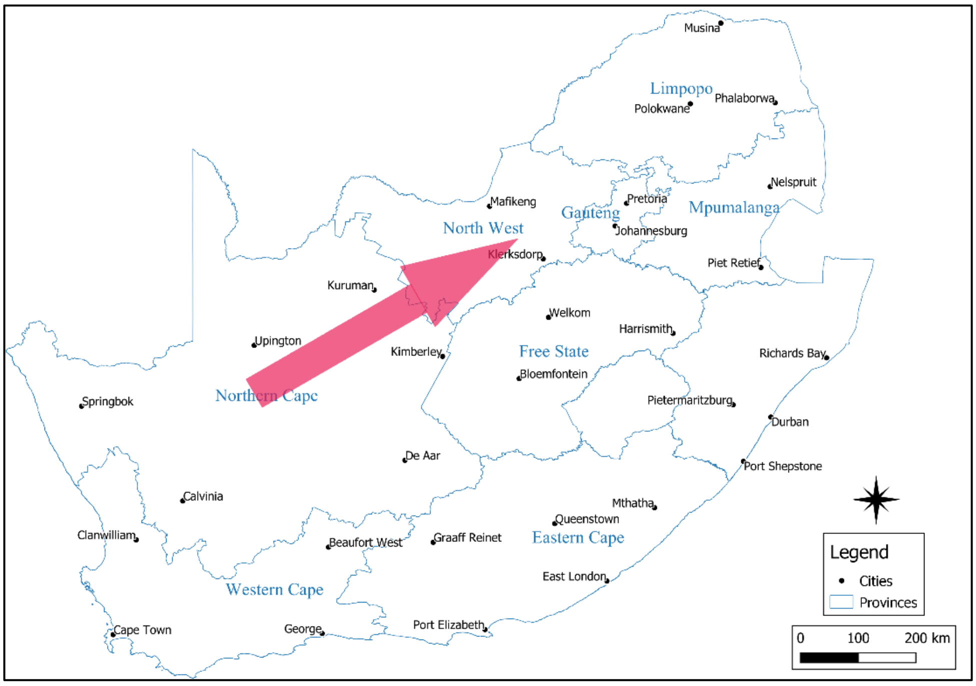


Figure 3.13: Map showing the direction of the Kgalagadi Butterfly Migration over the country (Terblanche, 2020).

Catopsilia florella (Fig 3.14) is a pale green (male) or deep yellow (female) butterfly that occurs throughout the country with their core distribution being in the Arid Savanna (Willis and Woodhall, 2010; Woodhall, 2012). The butterfly is part of the subfamily Coliadinae (Willis and Woodhall, 2010; Woodhall, 2012). Their most common times of migration is between November and June (Willis and Woodhall, 2010; Woodhall, 2012). While *Catopsilia florella* does not form part of this study, it is of interest since it commonly occurs with *Belenois aurota* as part of the annual migration.

Figure 3.14. Images showing male and female *Catopsilia florella* (Willis and Woodhall, 2010).

3.3.2 Other species of interest

Some butterflies are part of the same Genus that are commonly misidentified and may be confused with *Belenois aurota* based on their appearance. Two examples of such butterflies are *Belenois gidica abyssinica* (Figure 3.15) *and Belenois creona severina* (Figure 3.16). While these butterflies do not form part of the Kalagadi Butterfly Migration, they tend to fly year-round and may be confused with *Belenois aurota* during migrations (Willis and Woodhall, 2010; Woodhall, 2012; 2020).

Figure 3.15. Images showing male and female *Belenois gidica abyssinica* (Willis and Woodhall, 2010).

Figure 3.16. Images showing male and female *Belenois creona severina* (Willis and Woodhall, 2010)

Chapter 4: Methodology

This chapter gives an overview of the methods used to conduct this study. Firstly, the data sources, methods for data collection and pre-processing will be described. This section will briefly discuss the data sources that were used and why. The two methods used to collect data from each of the sources will be addressed. Finally, the section will discuss any pre-processing measures that were undertaken prior to the data analysis stage. Secondly, the chapter will conclude with the data analysis methods used to conduct the study.

4.1 Data Sources, Collection and Pre-processing

4.1.1 Data Sources

The database was compiled from five main data sources. Three of the sources are social media posts from Twitter, Instagram, and Facebook and the selection of these were guided by the literature (ElQadi *et al.,* 2017; Fitchett and Fani, 2018; Hart *et al.*, 2018; Lopez *et al.*, 2019; Zipper, 2017). The fourth data source was historical newspapers while the fifth source was LepiMAP data. This study investigates whether social media sources and historical newspapers can be utilised for butterfly migration research in South Africa.

Social media as a data source was utilised because there is a large quantity of easily accessible data points available (Lopez *et al.*, 2019). Gathering data from social media is cost-effective, requiring less resources (Lopez *et al.*, 2019) and is relatively easy to collect (Hart *et al.*, 2018; Silva *et al.*, 2018; Zipper, 2017). Several studies also show that social media is a valuable resource for obtaining phenological data and conducting ecological studies (Breckheimer *et al.,* 2019; Fitchett and Fani, 2018; Hart *et al.,* 2018; Silva *et al.*, 2018; Zipper, 2017). Based on the findings of these ecological studies, this study is therefore based on the assumption that social media is a reliable and credible source for butterfly sightings. Another important assumption in this study is that any mention of the butterfly on social media forms part of the Kgalagadi Butterfly Migration.

Social media platforms as data sources were chosen based on their popularity for ecological and protected area research (Lopez *et al.*, 2019; Wilkins *et al.*, 2020). Twitter and Instagram are two of the most used platforms for this type of research (Lopez *et al.*, 2019; Wilkins *et al.*, 2020). Facebook is also regarded as a popular platform to use but features less ecological research than the other two platforms (Lopez *et al.*, 2019; Wilkins *et al.*, 2020). Facebook was, however, still included for this study for its genre-specific groups, which did provide valuable information (Chamberlain, 2018; Liberatore *et al*., 2018; Superio *et al.*, 2020). Additionally, bias is reduced by spreading the sample across multiple social media platforms (Lopez *et al.,* 2019). Furthermore, the selection of sources is based on two South African phenological studies by Fitchett and Fani (2018) and Fitchett *et al.* (2019). The authors had a high level of success with the usage of social media as a data source in their study of the flowering phenology of jacarandas in South Africa (Fitchett and Fani, 2018) and the migration of sardines in the southwest Indian Ocean (Fitchett *et al.*, 2019). Historical newspaper articles were also utilised for its value in migration research on sardines in South Africa (Fitchett *et al.*, 2019). Moreover, the newspaper records also provide a continuous record over a longer period to elicit phenological data about the butterfly migration (Fitchett *et al.*, 2019).

The final source was LepiMAP data which was selected as it is the only butterfly-related citizen science project in the country with data being collected by citizen science groups. LepiMAP is an online database that consists of photographs and georeferenced location information for moths and butterflies in South Africa (Edge and Mecenero, 2015). This project will in time develop an atlas for Lepidoptera at a continental scale (Edge and Mecenero, 2015). The data set comprises records from the LepiMAP platform, Southern African Butterfly Conservation Assessment (SABCA) virtual museum, several contributions from the Lepidopterists' Society of Africa (LSA), and many citizen scientists (FitzPatrick Institute of African Ornithology, 2020). This dataset served as a vetted comparative dataset in which to compare the social media data set compiled in this study. Despite several data-related challenges experienced with citizen science data, LepiMAP is arguably still more accurate as each entry into the data set has been vetted by at least two butterfly experts (FitzPatrick Institute of African Ornithology, 2020).

Hart *et al.* (2018) use a similar approach in comparing data collected from a citizen science project to data collected from social media. This study, however, focuses primarily on Twitter as a source for data. Hart *et al.* (2018) argue that social media data is a form of indirect citizen science and reveal that there are opportunities for phenological studies of charismatic species and events relating to the species. This study determines if any parallels can be drawn between citizen science data (LepiMAP) and data collected from social media with regards to the Kalahari butterfly migration.

4.1.2 Data Collection

Two different data collection methods were used to obtain the data set about the butterfly migration (Fig 4.1). A total of 1363 posts from Facebook, Instagram and Twitter relating to the *Belenois aurota* migration were collected. Historical accounts of the butterfly were compiled from newspaper archives. The first method used automatic web scraping specifically for Twitter. Web scraping is a data science method for extracting data from websites via an internet browser to be stored in a database for later analysis (Krotov and Silva, 2018). Based on the various studies being conducted using Twitter as a data source in ecological research (Hart *et al.,* 2018; Lopez *et al.*, 2019), it was assumed that scraping would yield more results and speed up the process. Furthermore, the scraping of Twitter data is allowable under the platform's terms and conditions and privacy policy using the Twitter API (https://twitter.com/en/tos). The second method utilised manual data collection from Facebook and Instagram (Fig. 4.1). Facebook and Instagram have more stringent policies and frown upon the automatic scraping of data from these respective platforms (https://www.facebook.com/terms.php). More data was obtained through manual collection (887 posts) versus automatic scraping (150 tweets). Historical data was obtained from an honours student completing another study in butterfly phenology in the School of Geography, Archaeology and Environmental Studies (Pandazis, 2020). While traditionally the term historical may refer to data that are 50 years or older, for the purpose of this study all records including and prior to 1983 are referred to historical. The data was collected from The Rand Daily newspaper (68%) and Google News (32%) (Pandazis, 2020). Overall, a netnographic data collection approach was adopted. Netnographic data collection is a process whereby one obtains scientific information from unstructured information (Miguéns *et al.,* 2008).The aim was to extract scientific data from unstructured, self-reported data collected from individual accounts of people's observations (Miguéns *et al.*, 2008; McKenna *et al.,* 2017). The butterfly migration can be considered a distinguishable event that draws the attention of many citizens, thus the levels of accuracy in reported dates of the butterfly migration may be relatively high (Fitchett *et.al*., 2019). However, an inherent margin of error exists in this type of data source with regards to both the dates and reliability of annual records(Fitchett *et.al*., 2019).

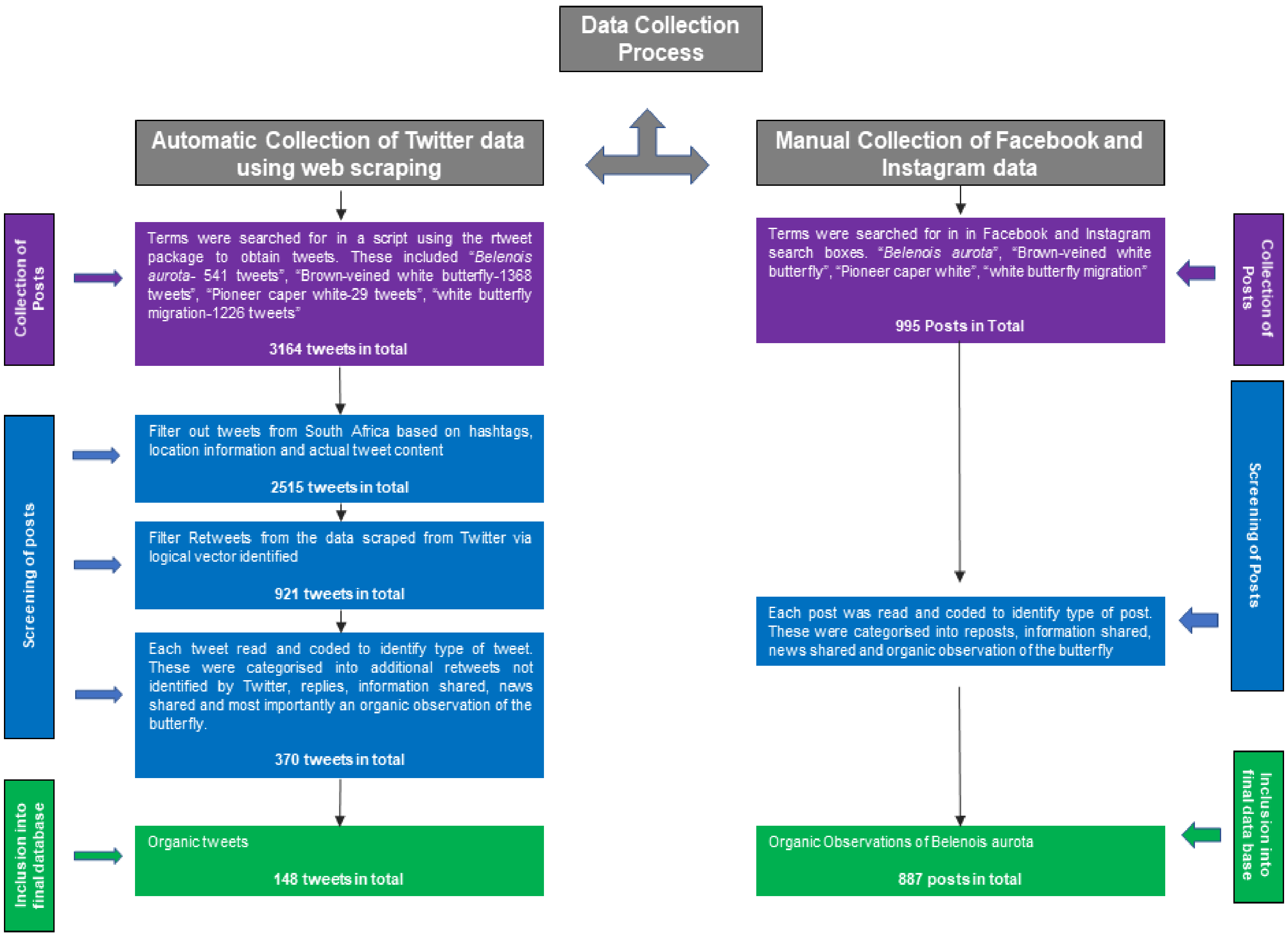


Figure 4.1: Data collection process used to collect data on *Belenois aurota* from social media.

The breakdown of posts per social media platform is outlined in Table 4.1. Most of the database is made up of posts from Facebook (703 posts), followed by Instagram (184 posts) and from the social media platforms Twitter (148 posts). The least number of records were taken from historical newspapers (28) which was focused exclusively on Gauteng..

Table 4.1: Social and print media sources of posts included in the final database.

Social Media Platform	**Number of posts included**
Facebook (2012-2020)	703
Instagram (2014-2020)	184
Twitter (2011-2020)	148
Total social media posts	**1,035**
Print Media (Historic Newspapers) (1914-1983)	28
Total of social and print media	**1,063**

Researchers who conduct research using Twitter data typically use the Twitter API (Zipper, 2017). This can be described as an interface that allows users access to Twitter's data via a developer portal (Zipper, 2017). The first step in collecting tweets from Twitter was to obtain authorisation from the platform. This process included an application to inform the platform about the intended purpose of the study and utilisation of the data collection process. Upon approval, the keys and tokens for the Twitter scraping process were provided by the platform for use in the scripting process. A script was written in R using the rtweet package (Kearney, 2019) to collect tweets for the 10-year period between the 2011 and 2020 butterfly migrations. The data was collected using a script that accessed the Twitter API using a similar process used by Zipper (2017) and Silva *et al.* (2018). The search terms that were used to scrape for tweets are captured in Table 4.2 below. As some discrepancies based on the usage of common names have arisen over time (Terblanche, 2020; Woodhall, 2020), this study assumes that names such Pioneer Caper White, Caper White, Pioneer White, and Brown-veined White butterfly refer to *Belenois aurota*.

The text-based search was used as opposed to location-based searches. While a location-based approach would be more accurate, the location information of Twitter posts is not widely available and therefore a text-based approach was obligatory. This lack of location information is consistent with the experiences of other researchers using social media as a data source (Hart *et al.,* 2018; Silva *et al.*, 2018). Although several posts were scraped via this method (3,164 tweets), only a small fraction of these posts were usable for the database (148 tweets) (Fig 4.1).

Table 4.2: Terms used to search for the data on Twitter.

Search Word	Search Begins	Search Ends	No of Entries Searched
Belenois aurota	31/05/2010	31/07/2020	5,000
African Caper White Butterfly			
Caper White Butterfly			
Pioneer White Butterfly			
Brown-veined White Butterfly			
Pioneer Caper White White Butterfly Migration			

The methods used for collecting data from Facebook and Instagram was derived from a study by Fitchett and Fani (2018). Data from Facebook and Instagram was collected manually by searching for posts using specific terms. Table 4.3 highlights the search terms utilised for the manual collection of data. Facebook posts were copied and pasted onto an Excel spreadsheet. Information such as the date, the text component of the post and where available, the locations of the post was collected. The Facebook posts were collected specifically from a page called the “The Great Little Butterfly Migration” (created 13/01/2014), a few posts from the pages “S.A. Butterflies, Bugs, Bees and other small things” (created 12/01/2012) and “African Butterflies & Moths LepSoc” (created 14/10/2013). The period for collection was from 2012 to 2020, based on when the group was created. The data set was then cleaned by filtering the text component of the social media posts and removing posts that referred to another butterfly or another country.

Table 4.3: Terms used to search for the data on Facebook and Instagram social media platforms.

Search Term	Related Hashtags
Africa Caper White	#africancaperwhite
	#africacaperwhite
Belenois aurota	#belenois_aurota
	#belenoisaurota
	#belenois
Brown-veined White Butterfly	#brownveined
	#brownveinedwhite
	#brownveinedwhites
	#brownveinedwhitebutterflies
	#brownveinedwhitebutterfly
Pioneer White Butterfly	#pioneerwhite
Pioneer Caper White Butterfly	#pioneercaperwhite

For Instagram, the hashtags listed in Table 4.3 were used to search for posts about *Belenois aurota* on the Instagram platform's search function. Each post was captured manually into a database in the form of a spreadsheet. Only those posts that referred specifically to *Belenois aurota* in the South African context were collected.

The historical newspaper dataset was obtained from a student in the School of Geography at the University of the Witwatersrand who was concurrently researching the same species. The student compiled their data using archived newspaper articles and social media (Pandazis, 2020). Only a portion (less than half) of the historical data set obtained was used in this study since many articles did not have important spatial and temporal information needed for this study. Data points from this historical dataset which had no definitive date regarding the migration were excluded from the final database. Some contemporary data from social media which was obtained from the student was, however, wholly excluded as to not duplicate any data in this study.

The data from all sources were combined into a single database in spreadsheet format. The combination of these records was done to allow for the analyses of all the observations to be conducted and imported into R. The structure of the database is described in Table 4.4 below.

Table 4.4: Description of the Data set containing data from all social media sources.

Column Name	Description
Date of Post	The date that the post was on the Social Media platform
Corrected Date	The corrected date based on the text portion of the post where a specific date of the observation was taken. If the post mentioned yesterday or this Monday, for instance, the corrected date was inserted in this column
Julian Date	This date was calculated using a modified function in R and was based on the date of the post and the corrected date where applicable
Text	The text portion of the post which included descriptive elements of the user's observation
Hashtag	This included hashtags that were associated with the Text column
Location	The location information of where the observation was made. This included separate columns for province and place

The LepiMAP dataset was obtained on request from the FitzPatrick Institute of African Ornithology (FitzPatrick Institute of African Ornithology, 2020). Data in this collection spanned over a period of 1887 to 2020 but only the years 1914 to 2020 were used for this study. Some of the data was further excluded based on missing and incorrect information. A total of 4,095 observations were included in the study. For disclosure of the limitations related to the data sources used please refer to section 6.5.3.

4.1.3 Data Pre-processing

Replies, reposts and retweets were removed from the dataset to remove any duplicate observations. These duplicate social media posts were identified using the excel duplicate values functionality. Posts on Facebook and Instagram were regarded as duplicate when users reposted the same text without acknowledging the original post. The original posts were determined by choosing the post with the earliest date on social media. Any duplicates of the original post were removed from the database. Retweets and replies were also present in the Twitter component of the dataset. The process used for removing these is outlined in Namugera *et al.* (2019). Retweets and replies were filtered out by using the pre-embedded fields composed of logical vectors called "is_retweet" and "reply_to_user_id" that formed part of the automatically scraped data set. These posts were removed from the database.

The dataset required cleaning to be compiled into a final dataset. Dates had to be corrected where observations were backdated. For example, if a user mentioned in a social media post today that they saw the butterfly migration yesterday, the date of the observation would be the previous day's date and not the date of the post on social media. This methodology of using the dates to record the observation of the butterfly is based on the studies by Fitchett and Fani (2018) and Fitchett *et al.* (2019). Irrelevant social media posts that included other butterfly species or posts from other countries were removed from the dataset.

The dates were then converted into modified Julian dates for analysis (Fitchett *et al.*, 2019). The lubridate package (Grolemund and Wickham, 2011) in R was utilised and the Julian date was then modified to calculate from the first of July of every year instead of the first of January. This was accomplished using a script that modified the dates in the dataset.

To conduct the text analysis on the content of the social media posts, the text and hashtag field needed some modification to be analysed in R. Special characters and emojis were removed by searching for and removing each of these characters from the database. This was done to ensure readability in the software, as R does not recognize these characters.

Only 882 were of the 1,035 total social media posts were used for the mapping component due to the fact that not all posts had associated location information. Many of the locations were inferred from the context of the text or the visible biographical information of the user. This follows the same approach as Hart *et al. (*2018). These records amounted to 294 unique locations. The location information was altered to give as much information as possible for each post. For example, if a city was mentioned, then the relevant province and country were added to a separate column for the entry. When a location was not clear, it was marked as unknown and removed from the spatial dataset. For instance: where specific reserves and lodges were mentioned, the specific coordinates for these were plotted. The package ggmap (Kahle and Wickham, 2013) was used to obtain spatial information using the geocode function for the plotting of locations where the butterfly was observed. Prior to being able to use this package, an application was made to obtain an API key for using some of the functionality including the geocoding of the locations. The coordinates were obtained through geocoding and were obtained in decimal degrees in a geographic coordinate system. The locations were structured to have the place, province and country concatenated in one column to avoid any confusion with ambiguous place names after geocoding.

4.2. Data Analysis

Summary tables were produced to describe the content contained in the dataset. These were created to address objectives 1 and 4 in Chapter 1 of this research project. The tables were created through pivot tables in Excel and then summarised using charts. The text analysis was conducted using R and content analysis. Both the social media and the LepiMAP datasets were statistically analysed using R. Maps plotting the locations on the butterfly sightings were created in QGIS.

The dataset from the previous step was imported into R for analysis. A text analysis was conducted on the text and hashtag fields using the tm package (Feinerer, 2013) to find the most used terms and hashtags in the dataset. The ggplot2 package (Wickham, 2011) was used to create figures. The text was also analysed using content analysis (Fitchett *et al.*, 2019).

To conduct the statistical analysis, the median dates were calculated for each of the years where dates of butterfly sightings were available using the modified Julian date. This analysis includes both the data from the period of this study and the historical data from the newspaper archives. Box and whiskers plots were created for observations in Julian days for all the years. These were created to compare the distribution of the dates across the years. This methodology utilises the median Julian data in contrast to Fitchett and Fani (2018), who uses the mean Julian date. The median was selected here as the distribution of data for this study was very skewed. A linear regression model using univariate linear regression analysis in the form y=mx+c was run in base R to calculate the rate of change in the timing of the butterfly migration (Fitchett and Fani, 2018; Fitchett *et al.*, 2019; Grab and Craparo, 2011; Hart *et al.,* 2018). Median Julian dates were used as the dependent variable and year as the independent variable. The same statistical analysis was conducted with the LepiMAP dataset for objective 4 stated in Chapter 1.

The coordinates obtained from the geocoding process were imported into QGIS 3. The coordinates were plotted to create a proportional symbol map for each of the years of the study, as well as a map that combined all the years (ElQadi *et al.,* 2017). These were analysed visually for each of the years of the study (Long and Robertson, 2018).

4.3. Ethical considerations

As this research used information indirectly from human participants, university-level ethical clearance was required to conduct this research. The application went through the Human Research Ethics Committee (HREC Non-Medical) of the University of the Witwatersrand. The project was deemed “Minimal -risk” and was approved unconditionally in July 2020. Proof of the ethical clearance certificate is in Appendix A.

The database containing the social media information was stored on a password-protected computer during and after research. The database was in a digital format with all identifying features of the social media participants removed. Only the research supervisors and the honours student who required access to the dataset for her research report had access to a limited version of this dataset.

Formal written consent was not practically obtainable from the social media users. Consent was addressed through the relevant permissions from Twitter through the app development application process and also by the adherence to the terms and conditions of Facebook and Instagram. The usernames associated with the posts regarding *Belenois aurota* observations were discarded prior to the time of the analysis and therefore, no usernames were published in the final report. In addition, the posts taken from Twitter, Facebook and Instagram do not form part of any closed group and are publicly available.

Chapter 5: Results

In this chapter, a description of the database collected from social media regarding the butterfly migration over the period 2011-2020 is provided. The interannual variability of the timing of both social media and historical data will be assessed. The distribution of social media posts in a spatial context will also be discussed. Finally, the data collected from social media will be compared to that of the LepiMAP dataset for the same period of the study.

5.1 Description of the database with dates and locations of posts relating to the butterflies

The original number of data points, before any data cleaning and removal of duplicates, was 2,472. Duplicate entries and retweets were created as part of the data scraping process. When these were removed, the number of data points was significantly reduced.

A total of 1,363 posts relating to the *Belenois aurota* migration were collected from social media platforms, particularly Facebook, Instagram, and Twitter. Historical accounts of the butterfly were compiled from newspaper archives. Table 5.1 depicts the number of posts that were collected from each of the social media platforms. The posts collected from Facebook accounted for 795 of the 1,363 records in the database. This is due to the data being collected specifically from butterfly-related groups. The data collected from Twitter and Instagram account for smaller numbers with 369 and 199 records, respectively. 28 data points were from historical newspapers (Table 4.1), which is significantly less compared to social media posts.

Table 5.1: Breakdown of the number of posts about the butterfly migration collected across three social media platforms.

Social media platform	**Number of posts collected**
Facebook	795
Instagram	199
Twitter	369
Total records	1,363

The number of social media posts collected from each of the platforms, and the various search terms used to find data related to this study, are summarised in Table 5.2. A breakdown of records in terms of the search words used to collect the posts from each of the social media platforms is reflected in Table 5.2. The term ‘brown-veined white butterfly’ yielded the most results across the social media platforms combined.

Table 5.2: Search terms used to collect the data about the butterfly migration.

<table>
<tr><th>Search Term to collect Instagram Posts</th><th>Instagram</th><th>Twitter</th><th>Facebook</th></tr>
<tr><td>Belenois aurota</td><td>55</td><td>38</td><td>170</td></tr>
<tr><td>Brown-veined White butterfly</td><td>133</td><td>123</td><td>625</td></tr>
<tr><td>Pioneer White Butterfly</td><td rowspan="2">11</td><td rowspan="2">8</td><td rowspan="2">0</td></tr>
<tr><td>Pioneer Caper White Butterfly</td></tr>
</table>

During the data collection process, various forms of media relating to B*elenois aurota* appeared on social media platforms. A summary of the type of media that was shared with the posts is seen in Table 5.3. While text-only posts were the most common, many posts were accompanied by photographs with more than half of the records having an associated photograph attached. 525 photographs of sightings of the Brown-veined White butterfly were shared (Table 5.3). On average, two photographs per post were shared. The second most popular shared medium was video footage of the butterfly. These amounted to 70 videos with an average of 1 video being shared per post. It must be stated that there would have been an increase in the types of social media shared but many of the older tweets did not have operational links and are thus not accounted for in this study. Some text-only observations did include a location pin of where the butterfly was observed. Approximately 1% of posts also included URLs to content relating to the butterflies (Table 5.3).

Table 5.3: Breakdown of the type of social media posts about the butterfly migration.

Type of associated media content	**No. of posts**
Only Text Description	752
Photographs	524
Video	70
Web URL	17

A total of 218 posts on social media made use of hashtags, of which 176 were from Instagram and 43 from Twitter. There was no hashtag usage on the Facebook social media posts. The most frequently used hashtags are depicted in Figure 5.1. The most frequently used hashtag is “butterfly”, which is expected for this study. The second most used hashtag is the “brown-veined white butterfly”. This can also be expected as this project and the data set focuses specifically on this species of butterfly. It is interesting to note that the only city that shows up in the hashtags is Johannesburg and its colloquial name “Joburg”. There are, however, many other cities and places that are mentioned - just not as frequently as Johannesburg. These echo the summary in Figure 5.1 that shows many of the records originate from Gauteng.

“Photo” is one of the 14 most used hashtags in the database. Upon closer inspection, however, there are several single-use or less frequently used photo-related hashtags. These were related to the type of photography such as wildlife and nature photography. There were also several hashtags on the type of the camera or lenses used to capture these photographs such as Nikon and Canon. There were also several hashtags relating to macro photography both in terms of the type of photography and the lenses used in the photography.

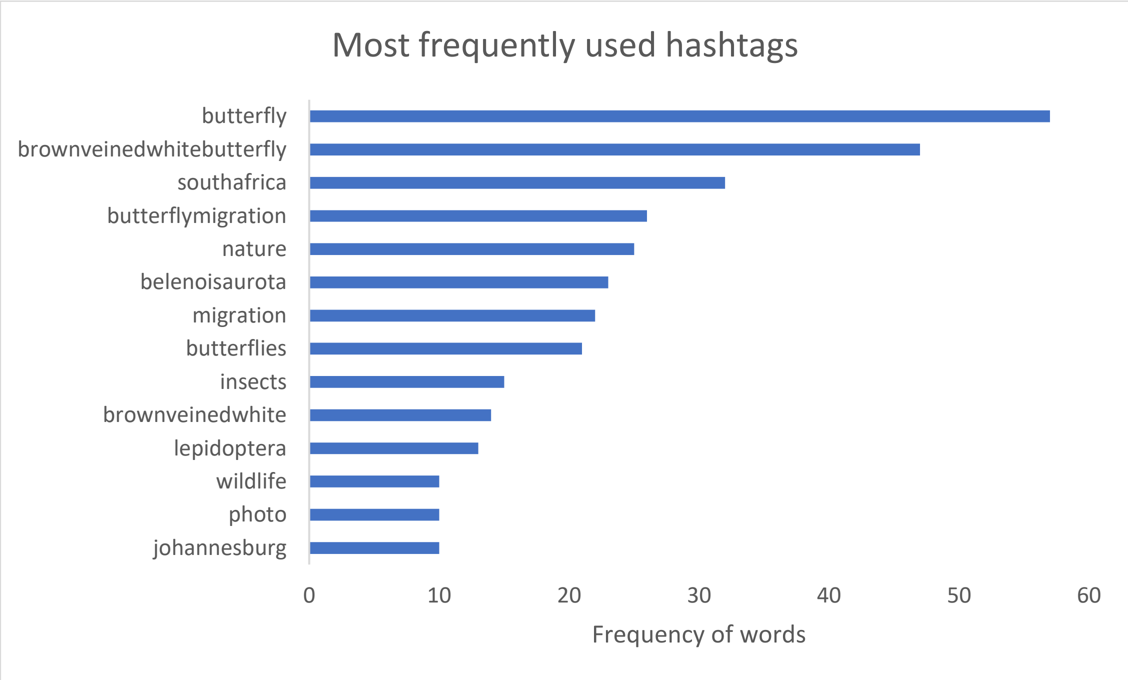


Figure 5.1: Number of records of frequently used hashtags in the social media dataset (Beginning 2011- mid 2020).

Only 50 users made use of emojis in the text portions of their posts. Appreciation and excitement for the butterflies were shown through the type of emojis used. Common emojis used included praying and clapping hands(🙏,👏,🙌), thumbs up (👍), excited faces (😄,😍,🥰,😃,😊,😲,😊,😯,🤣), hearts (❤️,💖, 💕), and camera (📷). However, the butterfly (🦋) emoji was most frequently used. Several posts mentioned what activities the butterflies were involved in when the users witnessed the butterflies. In some instances, multiple activities were mentioned in one post. “Flight” was mentioned 185 times while “migration” was mentioned 47 times. Other mentions include “mud-puddling” in “mud” or “dung” (12 times), “nectaring” (48 times) and “roosting” (5 times).

The most frequently used terms in the text portions of the posts in the database are illustrated in Figure 5.2. “White”, “brown-veined white”, “*Belenois*” and “*aurota*” feature as the most frequent terms relating to the butterfly in this study. The term “migration” and “hundreds” are also featured, which suggest that multiple butterflies were observed travelling across the country as part of a migration. Closer inspection of these terms also reveals words such as “few”, “thousands” and “millions” also indicating that users are attempting to quantify the number of butterflies witnessed. Some records included cardinal directions, which imply that observers determined the migration direction. When the word “today” was posted, it suggests that observers posted on social media platforms on the same day as the observation. Several users were observing butterflies in the gardens of their homes (95 posts) while a few users also mentioned witnessing them in gardens of nurseries, national botanical gardens, and nature reserves.

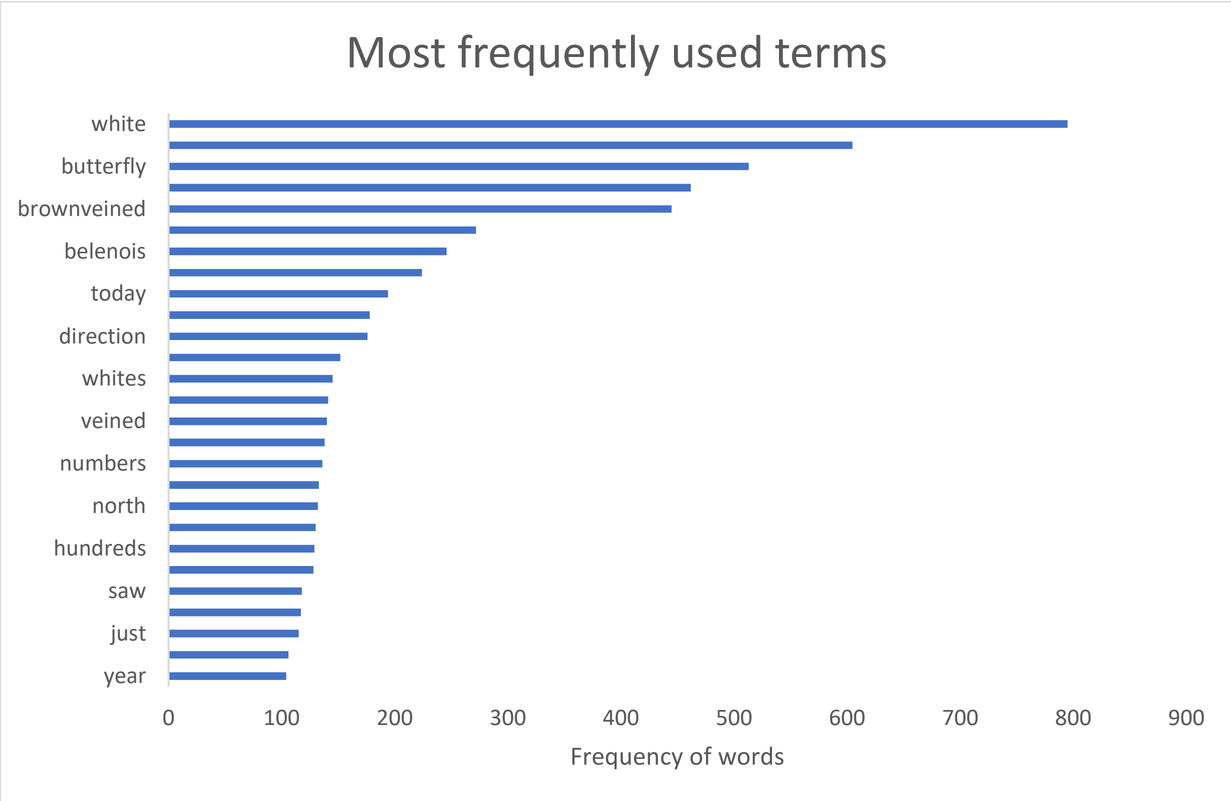


Figure 5.2: Number of records of frequently used terms used in the text portion of social media dataset (Beginning 2011- mid 2020).

A few examples of *Belenois aurota*-related posts that were collected are discussed below and these do not form part of the organic posts that were used in the analysis. The group administrators of “The Great Little White Butterfly Migration” Facebook group posted guidelines for the type of data they want users to contribute. This information is summarised in a post in Figure 5.3.

[redacted]
January 13, 2014 ·

This is an experiment in doing some real scientific research on Facebook. The question we are asking is: "What is the extent of the current migration of white butterflies across southern Africa?" The English common name for this species is Brown-veined White, the Afrikaans common name is Witgatwitjie, and the scientific name is Belenois aurota.

The most important information we need is "Where did you make your observations?" and "What was the date?" and "What did you see?" For the first question, if you can gives us geographical coordinates (from a GPS unit, from your phone, or from GoogleEarth) that would be great, but if you need to describe it in words, that is also fine (eg "10km north of Deneysville on the N14"). You can expand your answer to the third question by answering these questions: "How many butterflies were passing by?" (none, just a few, tens, 100s, 1000s, 10000s or more?). "In which direction were they going?" "When did the migration start in your area?" If there are no swarms of little white butterflies moving across the landscape in your area, please report that too – negative information is valuable. And you may contribute more than once. If we need to, we might send you a message on Facebook to ask you for some more information.

This project is led by LepiMAP, a joint initiative of the Animal Demography Unit at the University of Cape Town, and LepSoc, the Lepidopterists' Society of Africa. Data collected will be shared with Reinier Terblanche, who is doing his PhD on this topic.

Figure 5.3: Post informing users on the guidelines for posting to the "The Great Little White Butterfly Migration" Facebook group.

The information shared by scientists and lepidopterists included guidelines to identify butterflies that form part of the Kgalagadi Butterfly Migration (Figure 5.4). Some researchers also shared specific information about the migration that was occurring at the time of their posting on social media.

[redacted]

March 26, 2014 ·

Great to see how aware of things people on this facebook!

African Migrants (Catopsilia florella) are now really taking part in the autumn migrations. Even in flight you will be able to distinguish between Brown-Veined Whites (Belenois aurota) and African Migrants (Catopsilia florella). Do not underestimate your "instinctive eye"!

In flight the African Migrants are most of the time visibly larger than the Brown-Veined Whites. In flight the African Migrants appear "clean" white or yellow, whereas the Brown-Veined Whites appear a bit more dirty, especially the female (no correspondence about this look of these females please :-)). African Migrants have a high directed flight but will dive down suddenly for flowers. Flight of Brown-Veined White can also be high, and fairly directed. This directed flight distinguishes the migrating species above from Meadow Whites, Pontia helice, in the field. Pontia helice, the Meadow White, tends to fly lower down, a bit more at random and has a rounder wing-shape, than the above migrating species and the female of Pontia helice appears even more dirty in flight.

Sequence of pictures below: Male Belenois aurota, Female Belenois aurota, White form of Catopsilia florella, Yellow form of Catopsilia florella and Female Pontia helice, all from the same locality.

Figure 5.4: Post informing users how to identify butterfly species.

Researchers were predominantly active on various butterfly-related Facebook groups such as "The Great Little Butterfly Migration" Facebook group. The researchers working on the butterfly migration are made known through posts on research updates.

Citizen scientists shared posts which depicted the entire life cycle of a butterfly, including dead butterflies that could not finish the migration (Fig. 5.5) or were targeted by other animal species. This study only considers live adult butterflies.

Figure 5.5: Indicative comment showing posts about dead butterflies.

Citizens also shared accounts where they expected to see butterflies in a specific locality but there were none. Retrospective information was also shared by users of about migrations witnessed in the past (Figure 5.6) and, in a few instances, some old pictures of butterflies were shared.

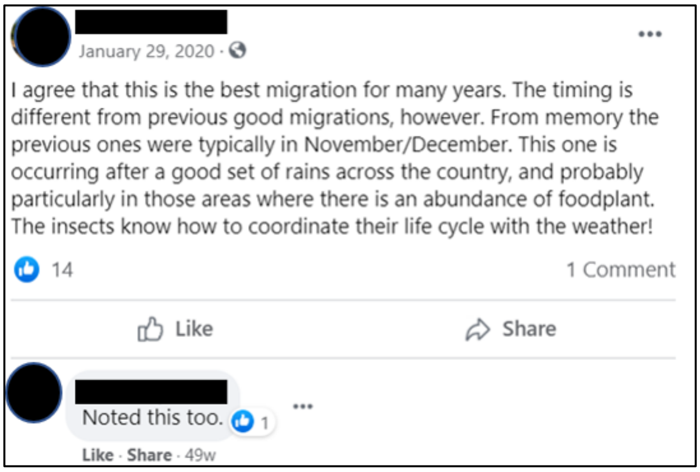
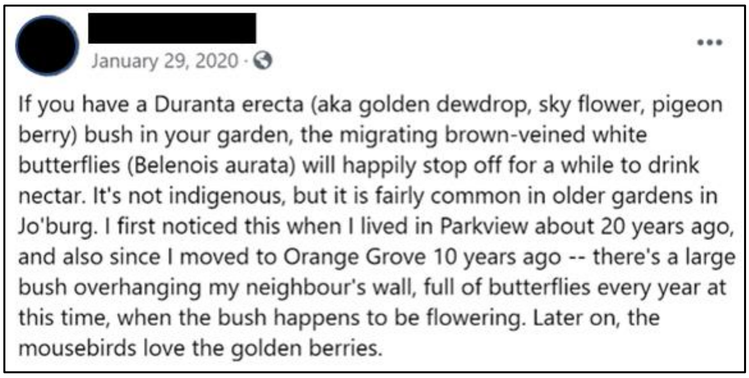


Figure 5.6: Indicative comment with user giving retrospective information about butterfly migrations.

Users and researchers shared tips for attracting butterflies to one's garden and tips to capture butterflies for research purposes. Advice on which plants attract butterflies was also shared (Figure 5.7). Users also shared tips on what to feed butterflies in one's garden (Figure 5.8).



[redacted]
January 29, 2020 ·

If you have a Duranta erecta (aka golden dewdrop, sky flower, pigeon berry) bush in your garden, the migrating brown-veined white butterflies (Belenois aurata) will happily stop off for a while to drink nectar. It's not indigenous, but it is fairly common in older gardens in Jo'burg. I first noticed this when I lived in Parkview about 20 years ago, and also since I moved to Orange Grove 10 years ago -- there's a large bush overhanging my neighbour's wall, full of butterflies every year at this time, when the bush happens to be flowering. Later on, the mousebirds love the golden berries.

Figure 5.7: Indicative comment with user giving plant advice on attracting butterflies to one's garden.

Figure 5.8: Indicative comment with user giving advice on attracting butterflies to one's garden.

Citizen scientists also used the social media platforms to pose questions to the researchers and lepidopterists that are active on the platforms, particularly on Facebook where these experts are part of butterfly-related groups. Companies such as BP South Africa and the Wild Bean Café referred to the butterfly migration in their advertising campaigns on social media (Figure 5.9).

Figure 5.9: Indicative comment with a company (BP South Africa) referring to the butterfly migration in an advert on social media.

There was a total of 525 unique users contributing to records across all the social media platforms. Most users (388) shared only one post (Table 5.4). The number of users that shared between two and ten posts amounted to 125 while the number of users that shared between 11 and 20 posts is 10 (Table 5.4). Only one individual shared over 20 posts on any of the platforms.

Table 5.4: No of posts users posted on social media platforms.

No of posts	**Users**
1 post	388
1 to 10	512
11 to 20	10
>20	1

The next part of the analysis determined if the posts were organic (original posts) or shared content from other sources. Not all data points of this dataset could be used in this next part of the analysis.

The division of posts between those that were organic posts from individual users' records of the butterfly migration and those from shared sources are summarised in Table 5.5. The total number of organic posts were 1,035, while 328 were from other sources. These 328 posts comprised duplicate replies and reposts - especially in the Twitter dataset - news articles from local newspapers sharing migrations witnessed at the time of publication and interviews with experts in the field.

Table 5.5: Summary of organic and other posts from social media.

Type of Post	Number
Other	328
Organic Observation	1,035

5.2 The interannual variability in the timing of posts relating to the butterflies

In this section, a total of 1,035, organic social media posts about the butterfly were utilised (Table 5.5). The interannual variation depicts the changes in migration dates over 1914-2020.

The number of social media posts per year for 2011-2020 are illustrated in Figure 5.10. The largest number of records from the social media platforms are in the years 2014, 2016 and 2020. The number of posts about the butterfly migration for these years are 223, 143 and 403, respectively. Facebook accounts for the most records for all the years from 2014 onwards (Figure 5.10). An increase in posts is observed for all platforms in 2020. The year 2020 also has the highest records for all platforms for the period 2011-2020 (Figure 5.10). The number of records in 2020 amounts to 40% of the total data set and is almost double that of 2014 (22%) which has the second-largest number of records per year (Fig. 5.10).

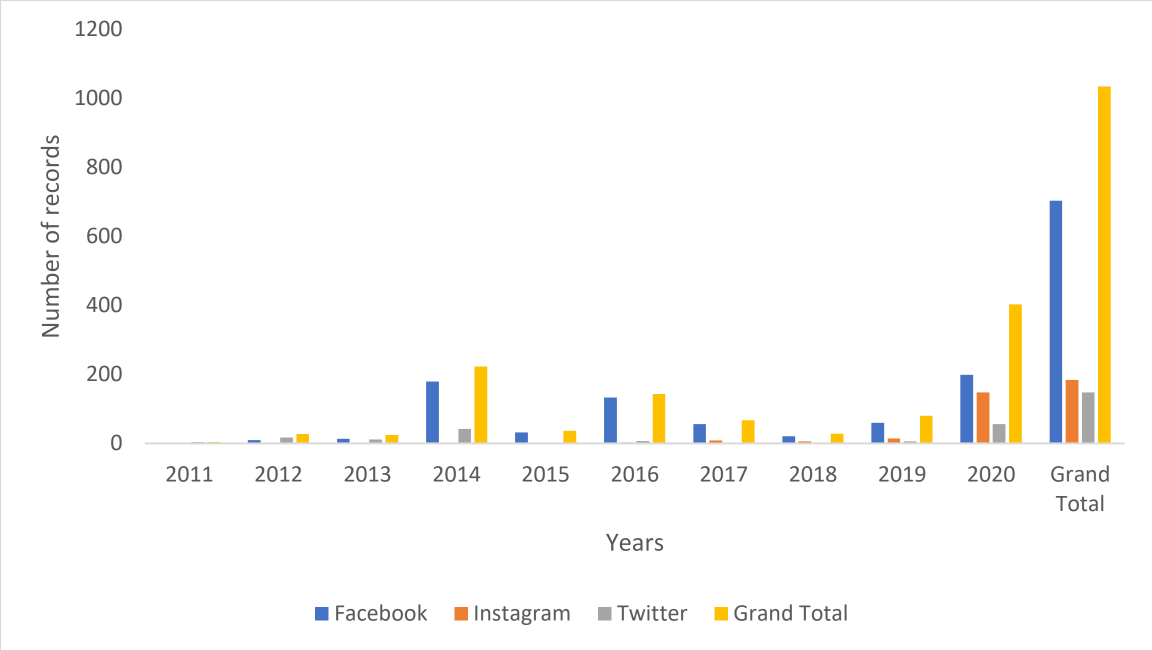


Figure 5.10: Number of social media posts mentioning *Belenois aurota* per year per social media platform.

The total number of records sorted by month is depicted in Figure 5.11. The greatest number of posts start in December and end in March. The majority of the records and posts on social media were posted in January and February. The records for 2020 account for most of the total January and February records. The records from 2016 primarily account for the total March and December records. The records for 2014 account for the second-highest number of records in January and predominantly the total number of records in November.

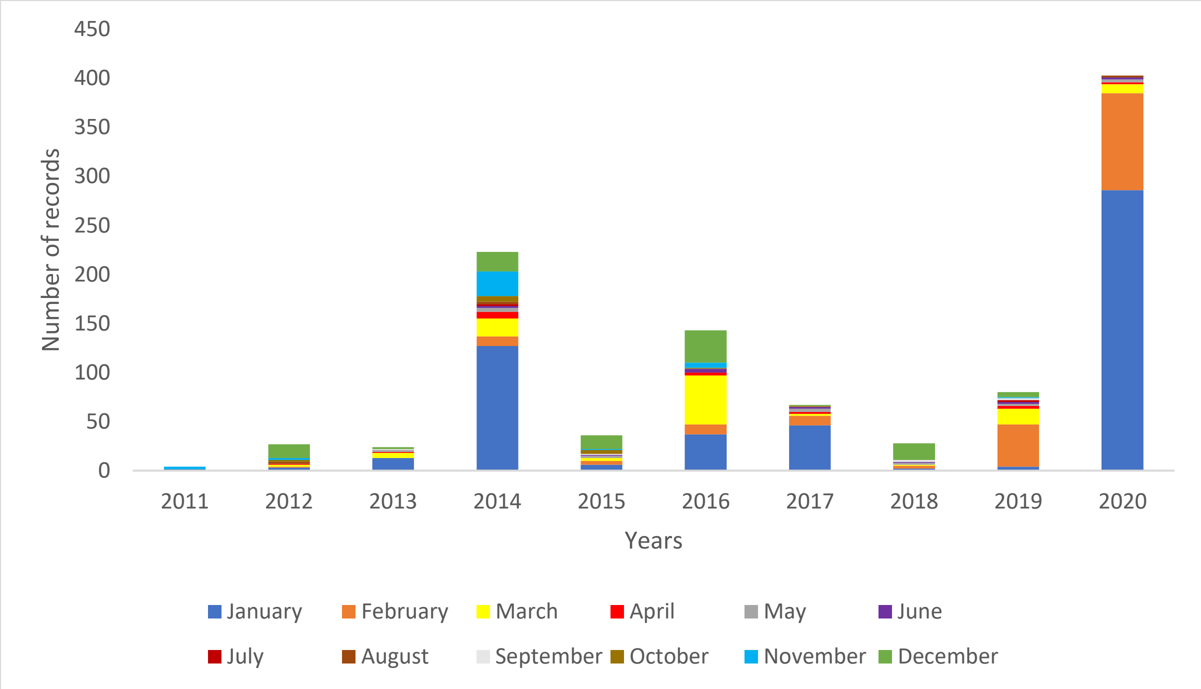


Figure 5.11: Total number of social media posts per month.

The distribution spread of the data from 1914-2020, including historical newspapers in the total records, is summarised in Fig 5.12. The dataset is not continuous across decades; a break in the records occurs between 1983 and 2011. The years 1914, 1915, 1930, 1933, 1935, 1952, 1962, 1974, 1975 and 1983 have only one observation per year. Overall there is much variation in the median Julian date across the years as the range is between 161 (~8th December) and 269 (~25th March) (Fig. 5.12). The range of the mean Julian dates is between 161 (~8th December) and 269 (~25th March). The range in minimum values is between 1 (~1st July) and 269 (~25th March), while the range in maximum values is between 161 (~8th December) and 269 (~25th March). There is also variation in the interquartile ranges across the years, especially during the last decade. Several outliers are present during the last decade. Many of the distributions for the years do not exhibit symmetrical distributions.

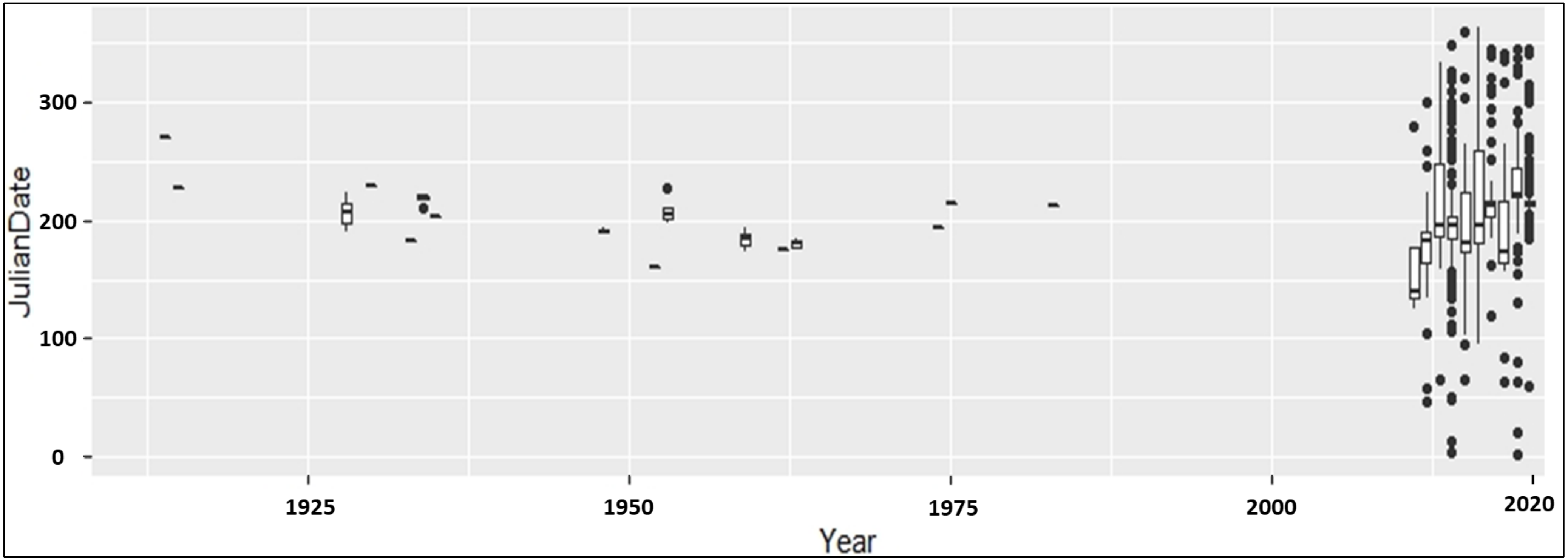


Figure 5.12: Box and whisker plots for social media data for the years 1914-2020.

A linear regression model was run to determine the change in migration dates over time from print and social media. The *Belenois aurota* migration dates range from 140-269 days (~ November to March). The median date for the butterfly migration was 275. The median Julian dates for each year of the study is depicted in Figure 5.13. The regression line is shown in red. This diagram shows that there is an advance in the migration as the median Julian date is earlier in the year for the most recent years. From 1914-2020, the median Julian migration date is from 269 (~26 March) in 1914 and 214 in 2020 (~29 January). The year 1914 seems to have a migration date that is inconsistent with the other years since the range for all other years is between December and February, with one other exception in 2011 with the Julian day 140 (~17 November).

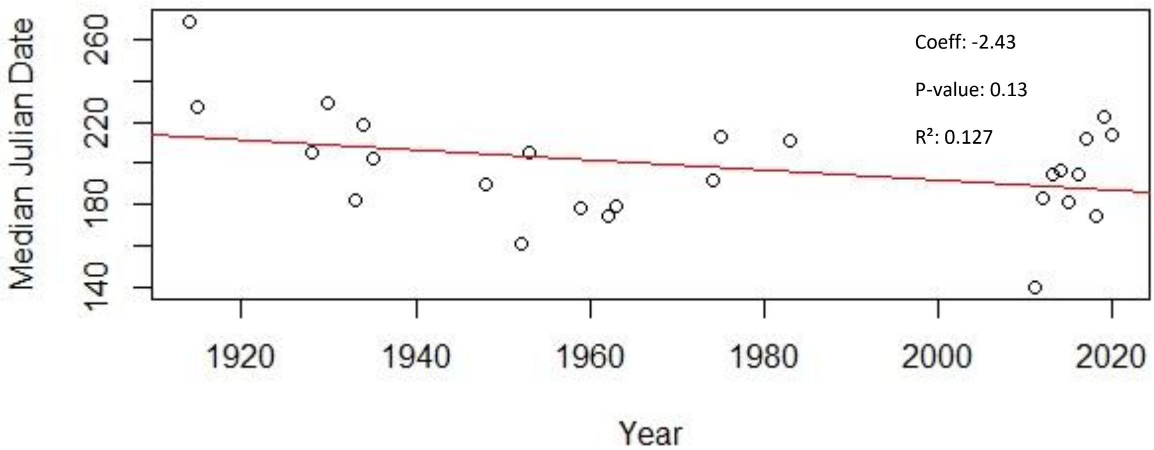


Figure 5.13: *Belenois aurota* migration phenological shifts derived from social media for the period 1914-2020.

The estimates (Figure 5.13) suggest that 1 year will have a decrease of 0.243 Julian days and hence a 2.4-day advance in migration of these butterflies per decade. The regression model shows that the estimates are not statistically significant as $p > 0.05$. The R^2 is very low (0.127) suggesting that the model does not explain any of the variation in the response variable around its mean. There is also a poor negative bivariate association between the median and year variables ($r = 0.36$).

5.3 Compilation of a spatial inventory of butterfly sightings which are differentiated by year.

In total there have been records of *Belenois aurota* across the entire South Africa, with most of the records originating from Gauteng, the second largest number of posts are from KwaZulu-Natal and third-largest from Limpopo (Fig 5.14). The least number of records are from the Free State province and Western Cape province (Fig 5.14).

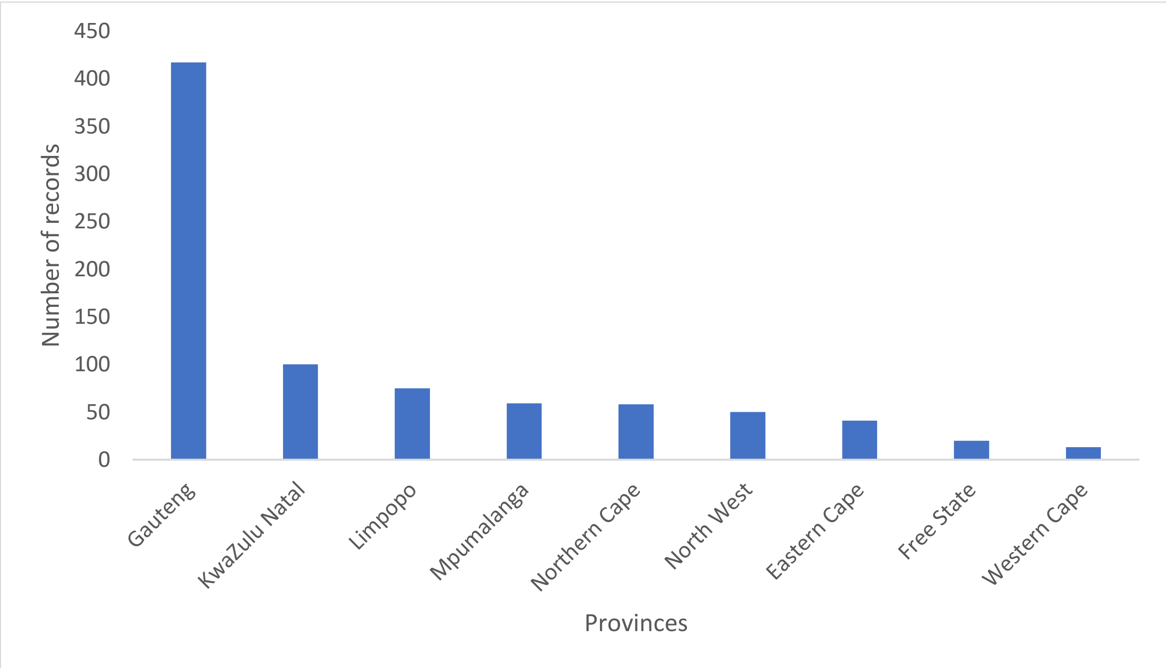


Figure 5.14: Number of records per province where *Belenois aurota* were observed (Beginning 2011-mid 2020).

A few records were removed from the data set, as discussed in the previous subsection. Only 834 posts had an associated spatial location and posts with unknown locations were removed. The following figures show the distribution of these butterfly sightings across the country for each of the year's data collected on social media.

The records for the years 2020 (309 records), 2014 (190 records) and 2016 (124 records) account for most of the records in total. The year with the least number of records is 2011 (Fig 5.15).

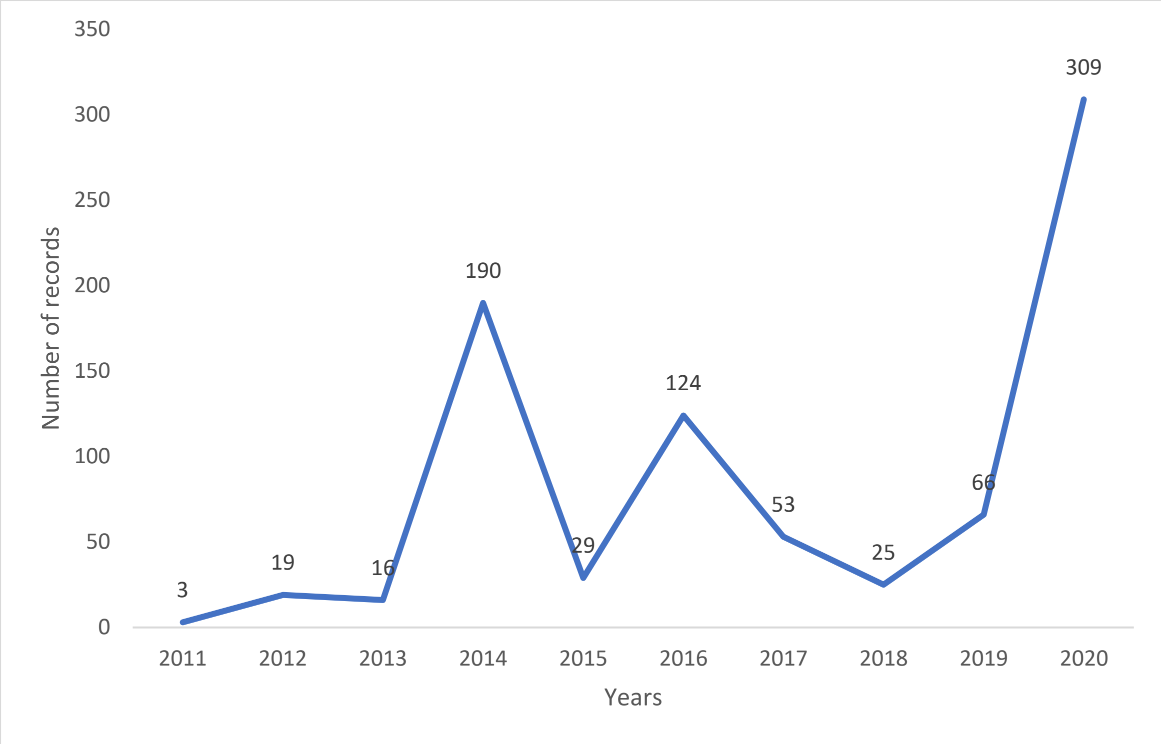


Figure 5.15: Number of spatial records per year.

The records in Figure 5.16 show that the sightings of the butterfly are scattered throughout the country with every province accounting for some records for many of the years. Gauteng accounts for the greatest number of records, specifically in the Johannesburg and Pretoria areas, with some records scattered throughout the province. Furthermore, there are more records in North West Province, Limpopo, Mpumalanga, KwaZulu-Natal, and Eastern Cape compared to Western Cape, Northern Cape and Free State (Fig. 5.16). Additional maps showing the observations per year can be found in Appendix C.

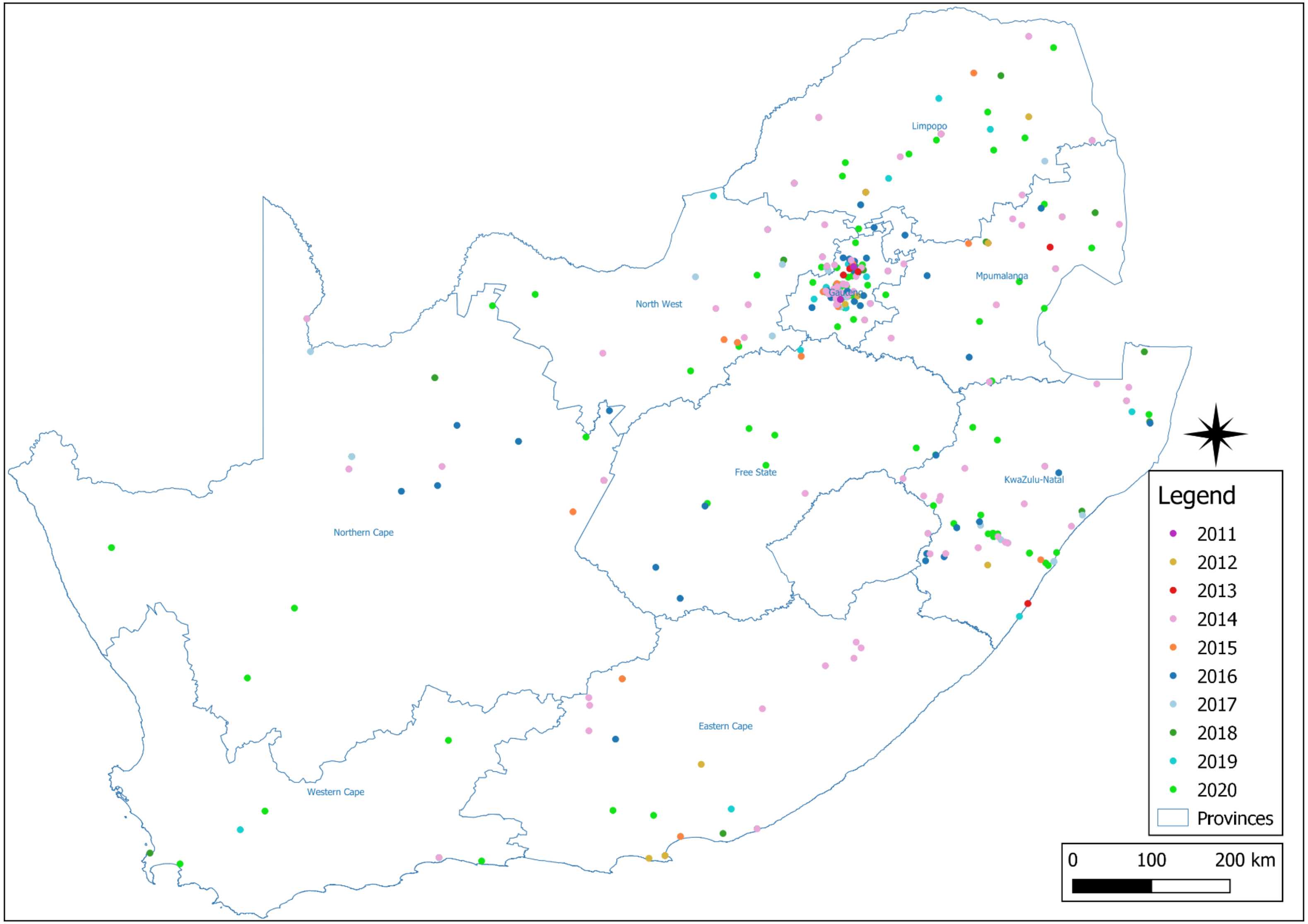


Figure 5.16: Map of locations from which recorded sightings of *Belenois aurota* were reported on social media.

5.4 Comparison of data from social media with that of the citizen science project LepiMAP in terms of interannual variability

As LepiMAP is regarded as a citizen science project and has no relation to the social media data collected in the previous section, these data were handled separately from the social media data. LepiMAP comprises data that is vetted by researchers in the field and it served as a comparative data set from which to compare social media data, which in most cases had no verification.

The number of sightings of *Belenois aurota* on LepiMAP increased drastically in the past decade (Fig. 5.17). There is a marked increase in records after 2007. The year 2009 has the most records (349), followed by the year 2020 (327). It is important to note, however, that the data for 2020 is not complete and does not include data for the summer of 2020.

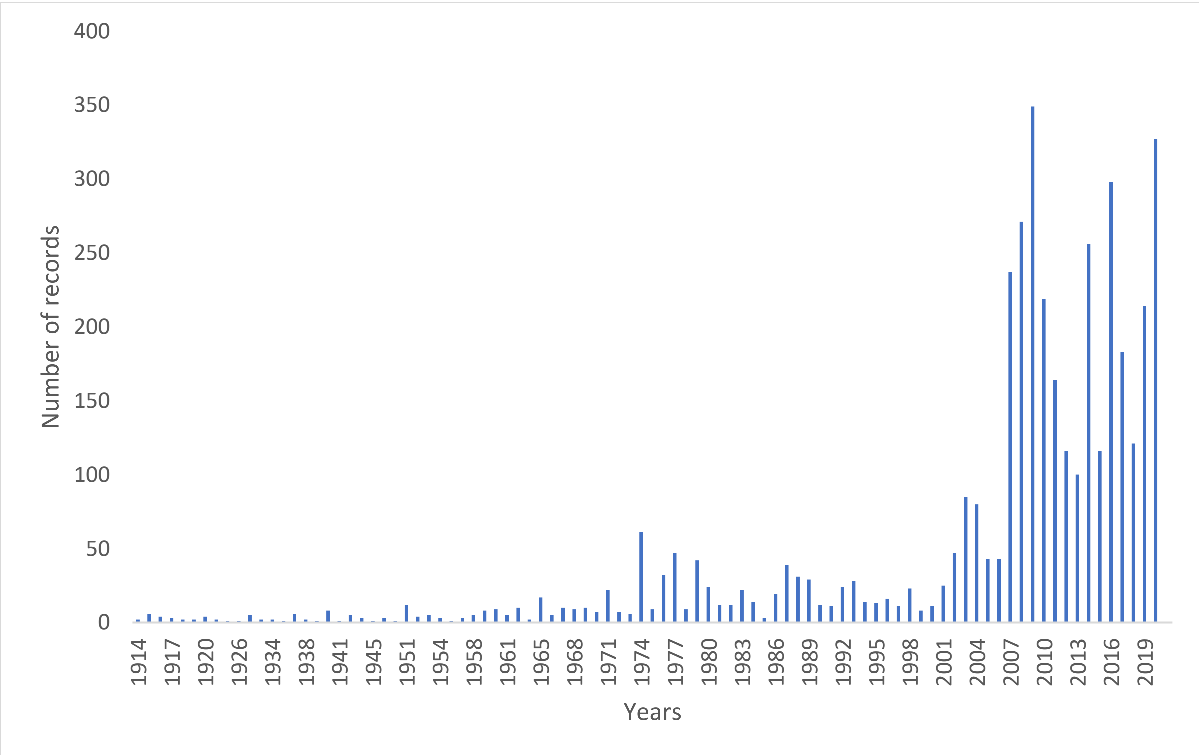


Figure 5.17: Number of records of *Belenois aurota* recorded on LepiMAP over the period 1914-2020.

Figure 5.18 depicts the entire data set available on LepiMAP for *Belenois aurota* in months. The peak period in which citizen scientists post on LepiMAP is December (804) and January (824), with January being the month with the highest number of records. Fewer records can be seen throughout the year. The months that have the least number of sightings made by citizen scientists are July (85) and August (49).

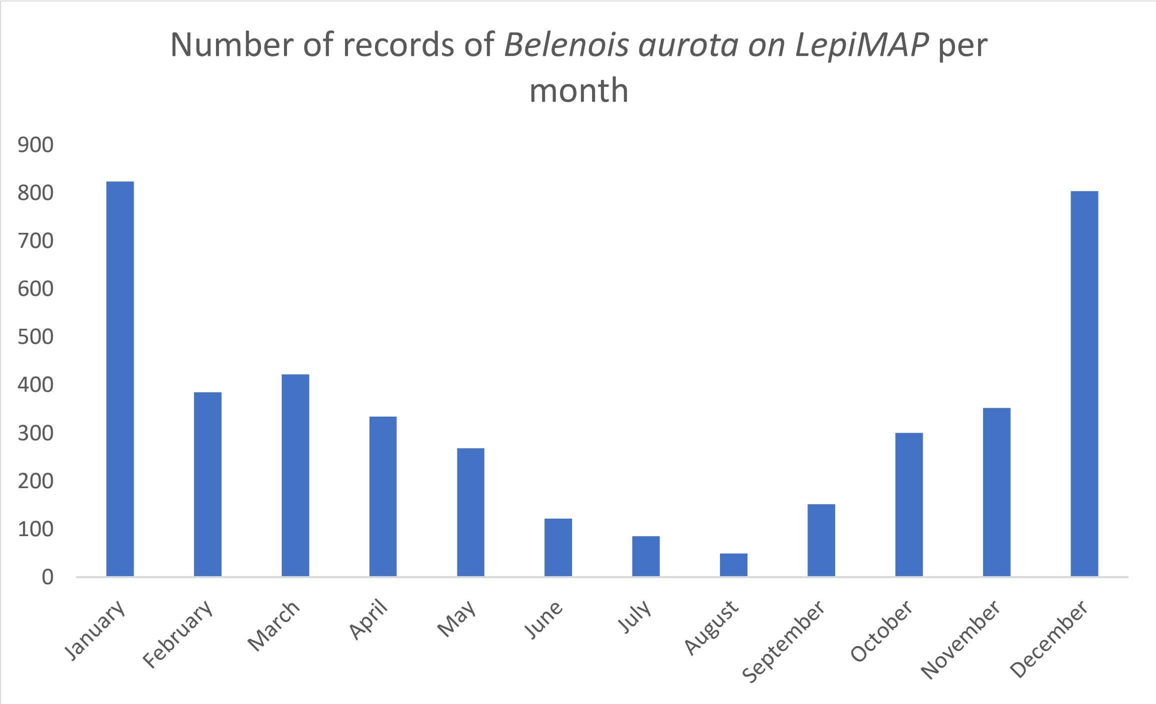


Figure 5.18: Number of records per month of *Belenois aurota* within LepiMAP over the period 1914-2020.

The greatest number of sightings of *Belenois aurota* come from KwaZulu-Natal (990). Limpopo has the second-highest number of sightings and Gauteng has only the third number of sightings (Fig. 5.19). The least number of sightings come from the Free-State and Western Cape provinces.

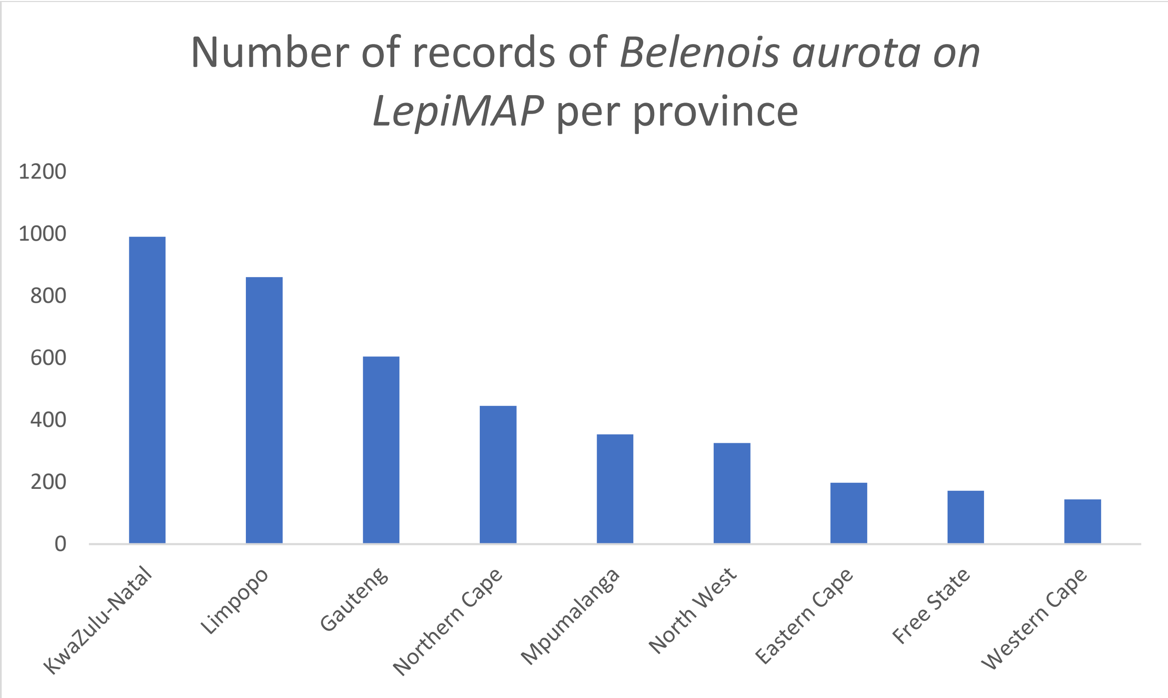


Figure 5.19: Number of records of *Belenois aurota* within LepiMAP over the period 1914-2020 per province.

The box plots below (Figure 5.20) show the spread of the Julian dates of the LepiMAP data from 1914 to parts of 2020. Figure 5.20 illustrates much variation in the median Julian dates across the years. There is also a great deal of variation in the interquartile ranges across the years. Also noticeable are several outliers per year, especially for the last decade of the study. In addition, many of the years do not exhibit symmetrical distributions.

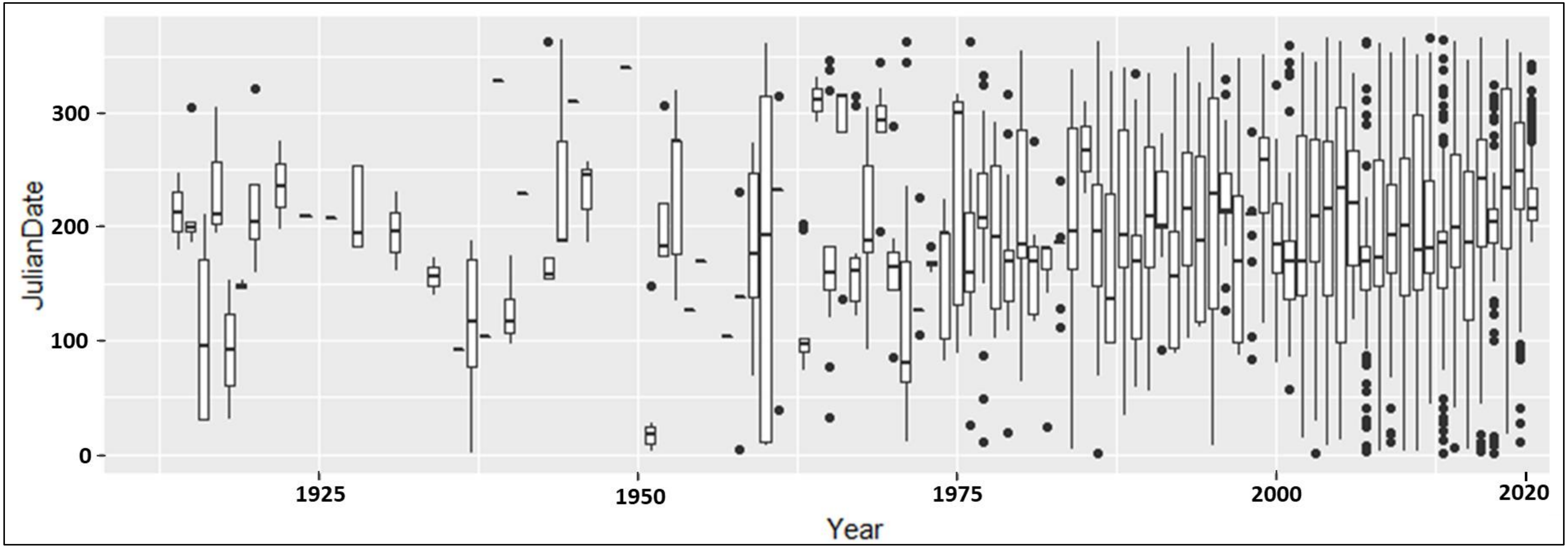


Figure 5.20: Box and whiskers plots of the LepiMAP data from 1914 to 2020.

To determine the change over time, a regression model was run. The scatter plot shows the median Julian dates per year with the regression depicted by the red line (Figure 5.21). The trend shown by the regression illustrates that there is a delay in the timing of the butterfly migration and that migrations are happening later in the year by approximately 2.3 days per decade. The median Julian dates range between 104 (~12th October) and 275 (~31st March) over the period 1914-2020. The Julian dates 212-214 (~28-30 January) account for 20% of the total number of reports. Julian day 214 (~30 January) accounts for the most reports overall with 8% of sightings being reported on this day.

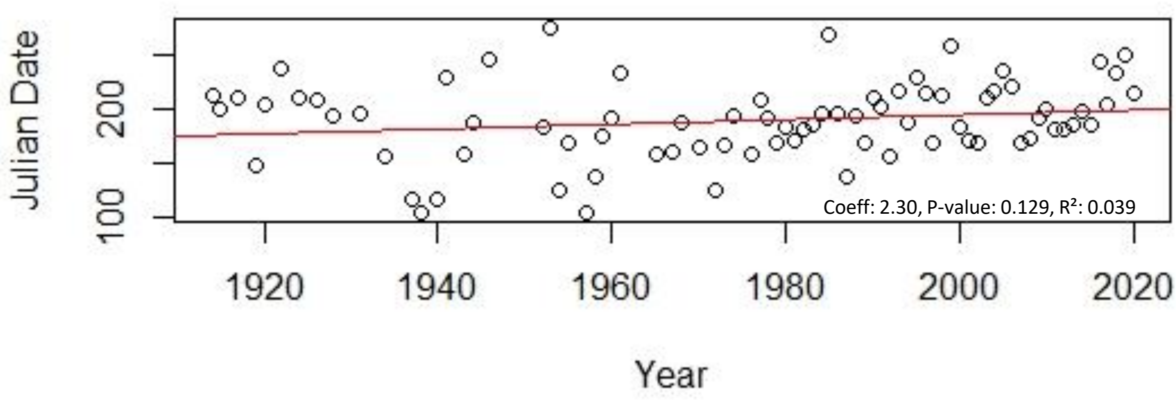


Figure 5.21: *Belenois aurota* migration phenological shifts for the period 1914-2020.

The model shows that the coefficient for Year is not statistically significant as $p > 0.05$ (Figure 5.21). The R^2 is very low (0.04) suggesting that the model does not explain any of the variation in the response variable around its mean. There is also poor bivariate association between the median and year variables (r=0.2).

The average Julian days compared for social media (Facebook, Twitter, and Instagram) and LepiMAP is summarised in Figure 5.22. The range of Julian dates between social media data and the LepiMAP data is between 64 and 240. (Figure 5.22). Twitter shows the most variation among platforms with the biggest difference for the year 2018, where LepiMAP has a mean Julian date of 231 (~16th February) and Twitter has a mean Julian date of 64 (~2nd September). Since there is a decline in posts between January to March in 2018 compared to other years (Appendix B), the mean Julian date for the year is skewed (Figure 5.22). Another smaller instance of this variation is seen with Instagram in 2016 where a mean Julian date of 177 (~24th December) is observed while the other platforms range between 217 (~2nd February) and 227 (12th February).

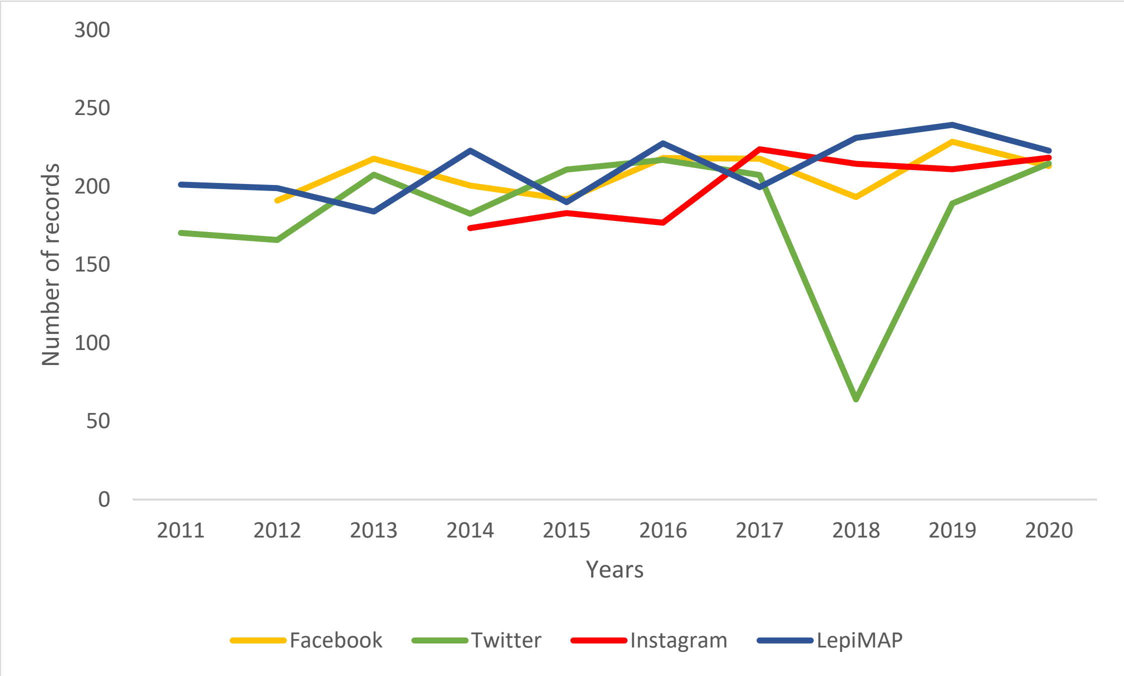


Figure 5.22: Mean Julian Dates for Facebook, Twitter, Instagram and LepiMAP for period 2011-2020.

In this chapter, a description of the database, constructed from social media records, was provided for the data collected on social media for the period 2011-2020. The interannual variability of the timing of both social media and historical data was assessed for the period 1914-2020. The spatial distribution of the social media records was discussed per province. A map showing spatial records derived from social media for the period 2011-2020 was provided. Finally, the interannual variability of the LepiMAP data was determined for the period 1914-2020. The data collected from social media was also compared to that of the LepiMAP dataset for the same period of the study.

Chapter 6: Discussion

This study aimed to explore the timing of the sightings of *Belenois aurota* from a database constructed from social media and newspaper records. The use of social media as a data source will be discussed. The spatial and temporal trends in the study will also be addressed. In addition, the use of social media as a source for phenological data and the phenological trends identified in this study will be discussed. This chapter will discuss the results in more detail comparing and interpreting the results to findings from other studies, particularly in terms of phenological studies conducted within the South African context and in particular those that specifically determined rates of change as part of their study. Finally, the limitations identified in this study will be provided.

6.1 Social Media Data Source

The number of records of *Belenois aurota* on social media had increased from 2011-2020; this may be attributed to the growing popularity and use of the social media platforms over time (Kemp, 2020; Statista, 2018). In January 2020, the most used social platform was Facebook (Figure 3.5), followed by Instagram and Twitter respectively (Kemp, 2020). A 2019 South African study determined similar preference trends in platforms used (Budree *et al.,* 2019). Facebook was shown to be the most used social media platform across all age groups in South Africa (Budree *et al.,* 2019). The composition of the data set collected in this study correlates to the popularity of the social media platforms, as described in Figure 3.5. Facebook yielded the most records, followed by Instagram and Twitter. The increases may also be due to the size of the migrations itself having caught the attention of the social media users.

While the total number of records collected in this study may not appear to be significant for a social media study, when compared to the citizen science project LepiMAP, which has only 1,896 observations in total for this period 2011-2020 (FitzPatrick Institute of African Ornithology, 2020), the number is notable. When compared to other social media studies however, the sample size is small. In the study exploring snake occurrences (Marshall and Strine, 2019), over 300,000 records were collected from social media. This study is, however, less comparable as it was conducted on a global scale. Hart *et al.* (2018) collected tens of thousands of records for flying ant and house spiders and approximately 4,000 for starling murmurations. This study was, however, conducted on a regional scale (Hart *et al.,* 2018). Zipper (2017) conducted agricultural research on planting periods on a national scale. Over 11,000 records were collected with only 3,000 being relevant to their study (Zipper, 2017). ElQadi *et al.* (2017) had under 2,000 posts. However, this study focused entirely on images and did not consider any other components of the social media post. Similarly, Breckheimer *et al.* (2019) focused entirely on photographs with a collection of over 10,000. Both Silva *et al.* (2018) and Fitchett and Fani (2018) did not explicitly declare their number of records obtained from social media. A common challenge experienced in this and other studies is that the collection of several records is pruned significantly based on the relevance to the study before analyses are conducted (Breckheimer *et al.,* 2019; Hart *et al.*, 2018; Zipper, 2017). The records are minimised further when considering spatial locations as part of the analysis (Hart *et al.*, 2018; Zipper, 2017). In this project the same reduction was seen when spatial locations of social media posts were considered.

The species migration event needs to be charismatic (Hart *et al.*, 2018) or striking (Marshall and Strine, 2019) or the users need to be intrinsically interested in the species (ElQadi *et al.,* 2017) for people to post about their observations on social media. This is subjective to the user as everyone has unique interests. Fortunately for this project, *Belenois aurota* and the migration thereof has proven to capture the attention of social media users (Figure 5.3 – 5.10).

With Facebook being the most used platform, Liberatore *et al.* (2018) argue that it allows users to interact in a way that is unique and is uncommon on other platforms. Facebook caters for niche interests in the form of moderated groups (Chamberlain, 2018). The use of hashtags could potentially mimic these groups but would be much harder to moderate unless the researcher creates and heavily advertises very specific hashtags and request the tagging of users. The only other social platform that allows for similar niche interests is Reddit. However, although Reddit is older than Facebook, it has less than a quarter of the total number of users compared to Facebook (Figure 3.5).

There is one unique feature of the Facebook groups that does not exist on the other two platforms: the occurrence of species verification by researchers on the platform. Many of the users from the two main groups ("The Great Little White Butterfly Migration" and "S.A. Butterflies, Bugs, Bees and other small things") from which this data was collected are lepidopterists or have advanced Lepidoptera knowledge and are constantly correcting and verifying information being posted on the social media platforms. Therefore, data from this platform is most likely to be the most accurate of all social media platforms. This highlights the current importance of having experts on social media verifying species until machine learning processes can outperform the human aspect of species identification (Jarić *et al.*, 2020; Silva *et al.,* 2018).

6.2 Spatial and temporal trends in data

While butterflies have been observed throughout the country, with a small number of observations in the Western Cape, Eastern Cape and the western parts of the Northern Cape, butterfly sightings are primarily concentrated in the northeast of the country (Fig 5.16) (Terblanche, 2020; Woodhall, 2020). The detailed maps per year of the study also exhibit a similar concentration of records albeit much smaller numbers (Appendix C). Sightings across the (2011-2020) period of the study have been concentrated in the Gauteng, North West, Limpopo, Mpumalanga, and KwaZulu-Natal provinces. This also corresponds with the movement of the butterfly migration as described in Terblanche (2020). There are consistent records in the area from which the butterfly originates (Kalahari) from 2014 onwards. Similarly, Prudic *et al.* (2017), records are concentrated around urban areas of high population density. When compared to population density data, a correlation is found between the spatial locations of records and areas of high population density (Figure 6.1). The densely populated areas are depicted by dark hatching in Fig 6.1. Most of the records are concentrated in Gauteng and along the east coast which are both areas of high population density (Figure 6.1). There may be other factors to consider as to why Gauteng residents post more frequently regarding the butterfly. These include socioeconomic, the fact that Gauteng is small in comparison to the other provinces and that it is much more densely populated. The number of sightings from Gauteng may be attributed to more social media activity in the province. A 2019 Stats SA report revealed that 74% of individuals from the Gauteng province always have internet access. The percentages for the other provinces are much lower. In KwaZulu-Natal, 56% of individuals have constant internet access (Stats SA, 2019). There is a lack of consistency between LepiMAP and social media regarding the origin and amount of posts across the provinces. Many of the observations for LepiMAP are for sightings in KwaZulu-Natal (24%) and Limpopo (21%). The number of sightings in these provinces may be attributed to many hobbyists in the KwaZulu-Natal and Limpopo provinces. This is in contrast to the social media data where Gauteng has the most sightings (32%) with KwaZulu-Natal only accounting for 8% of the total social media observations.

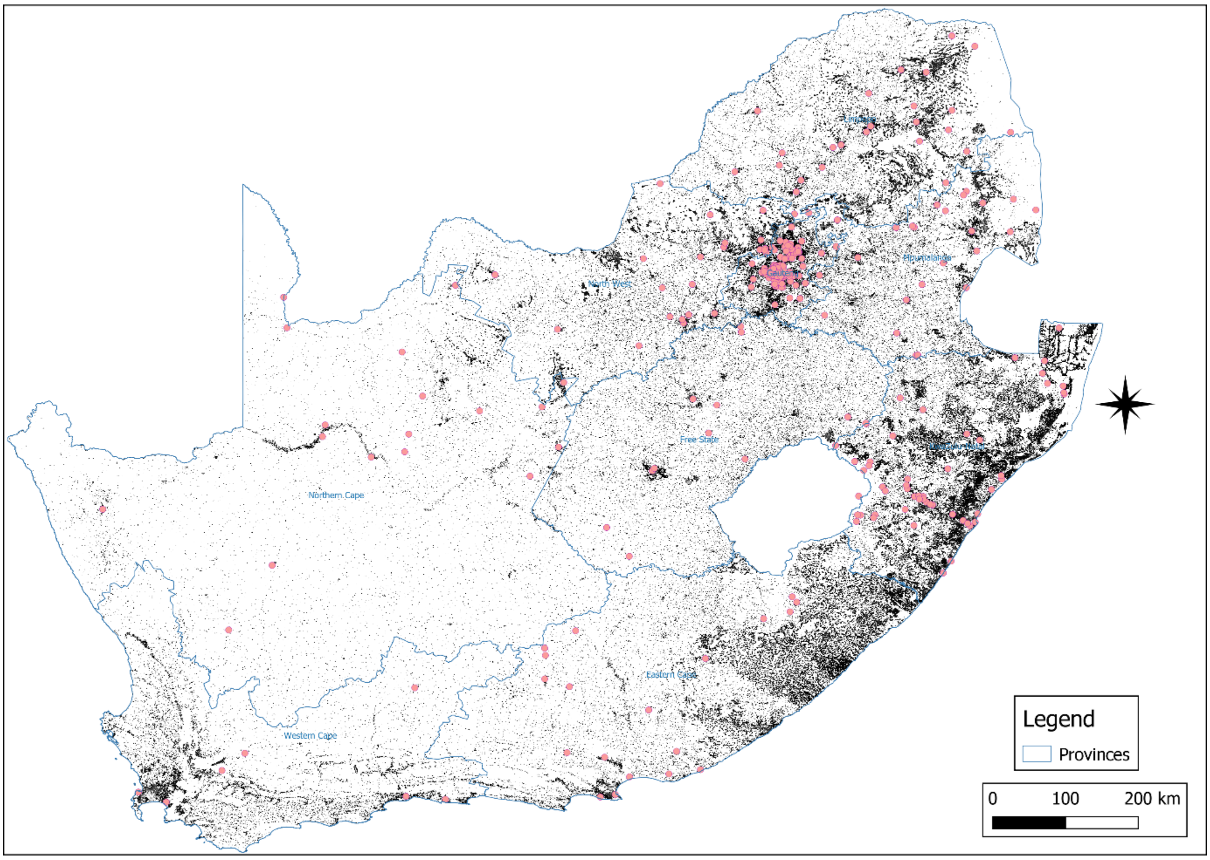


Figure 6.1 Map of *Belenois aurota* records in relation to population density.

The temporal pattern of the reporting of *Belenois aurota* was different across the years. The years with the most sightings on social media (Fig 5.10) are 2014 (21%), 2016 (13%), and 2020 (38%). These appear to correspond to big migrations (in terms of number of individuals sighted) reported in the media for those years (Burger, 2020; Mbhele, 2020; van der Walt, 2016). The migrations are the species responses to seasonal climatic variables and food supply (Terblanche, 2014;2020). In terms of the historic years from newspaper articles, 1934 and 1953 have more reports on the butterfly migration than on average for the same period. The same years with high volumes of reports on social media correspond to higher reports of *Belenois aurota* on LepiMAP (Fig 5.17), 2020 (8%). 2016 (7.3%) and 2014 (6.2%). There was a broad agreement of the monthly observations of *Belenois aurota* across both LepiMAP (Fig 5.18) and the social media data (Appendix B) showing that the largest percentage of sightings occurred in January. These January sightings account for 20% of the LepiMAP dataset and 51% of the social media dataset. In both cases these account for the highest proportions for all the months of the year. Several sightings also occurred in December, accounting for 20% of the sightings on LepiMAP, and 11% of the social media dataset. However, for observations occurring in February, those on social media account for 8% of the dataset while on LepiMAP this accounts for only 9%. The size of both the box and whisker plots for the social media data (Fig 5.12) and the LepiMAP data (Fig 5.12) clearly show that there are records of the butterfly year-round. This is more prominent in the LepiMAP dataset as there is a more complete dataset from which to compare.

For the social media and newspaper data set, 12 of the 26 years show a peak in the butterfly observations during January (1928, 1935, 1948, 1953, 1959, 1974, 1975, 1983, 2013, 2014, 2017 and 2020) (Appendix B). However, two of these years, 1928 and 1959, have reports of the migration over two months: January, and February for the year 1928 and January and December for 1959. The year 1928 shows that there is a continuation of a migration from January into February and 1959 is an example of two separate migrations during that year with one in January and another in December. The years 1933, 1952, 1959, 1962, 1963, 2012, 2015 and 2018 show peaks in December. The years 1915, 1928, 1930, 1934 and 2019 have peaks in February (Appendix B). Observations that fall within the months December – February correspond to summer season which is consistent with the seasons in which ground-truthing of *Belenois aurota* occurred (Terblanche, 2014; 2020). The peak months for 1914 and 2011 fall out of the summer season but according to Terblanche (2020) migrations can still occur in early autumn, therefore 2011 is the only true outlier from the data set. When comparing the years in the datasets to ENSO conditions (Table 6.1), 23% of the years in the social media dataset correlate to La Niña conditions compared to 16 % of years in the LepiMAP dataset (NOAA, 2020; Wolter and Timlin, 2011). El Niño conditions account for 27% of the years in the social media dataset compared to 29% in the LepiMAP dataset (NOAA, 2020; Wolter and Timlin, 2011). Most of the years in both the social media (50%) and LepiMAP (54%) datasets correlate to neutral ENSO conditions (Table 6.1) (NOAA, 2020; Wolter and Timlin, 2011).

Table 6.1: Data in relation to ENSO conditions

ENSO Conditions	**Social Media**	**LepiMAP**
La Niña	23.08	16.46
El Niño	26.92	29.11
Neutral	50.00	54.43

6.3 Opportunities for citizen science through social media

Social media does not require any registration and approval for users to contribute to the platform, apart from creating a user profile on the social media platform itself (Chamberlain, 2018; Hart *et al*., 2018). People may contribute to the platform as they observe the world without needing any knowledge of any citizen science projects that relate to their posts (Chamberlain, 2018; Hart *et al*., 2018). The limitation with a citizen science project such as LepiMAP is that platforms are not as popular as social media. The use of social media for data collection purposes is not without its own risks since researchers may not always confirm if posts on social media are true and correct (Chamberlain, 2018; Hart *et al*., 2018). However, advanced machine learning techniques may assist in species identification and the accuracy thereof can improve over time (Chamberlain, 2018; Jarić *et al.,* 2020).

Approximately 10% of posts specifically mentioned that they sighted these butterflies in their gardens. Cooper *et al.* (2007) argue that as urban sprawl increases, the need for the management of residential landscapes becomes critical. Citizen science initiative that includes property owners can contribute to positive impacts on biodiversity. The result of this study resonates with the findings of the Cooper *et al.* (2007) study in that it demonstrates the willingness of individuals to contribute intentionally and unintentionally to citizen science by sharing information about *Belenois aurota* on social media via observations made in their gardens. The data from the study was collected over a 10-year period at a national scale. The strength of citizen science is that projects can operate at broad spatial and temporal scales (Cooper *et al.*, 2007). Social media is an inherently ideal source of data for citizen science projects due to the way the platforms operate as well as the increase in reach and usage of these platforms over space and time (Hart *et al.*, 2018; Lopez *et al.*, 2019; Zipper, 2017). The collection of social media data also eliminates the need for coordinating costly (Breckheimer *et al.,* 2019; ElQadi *et al*., 2017) and complex (Dennis *et al.,* 2017) projects.

This study has demonstrated the opportunity to research many butterfly activities. Almost 3% of the posts in this study made mention of butterfly activities which include flight, migration, mud puddling, nectaring and roosting. With regards to roosting, the data on social media is very similar to the data collected regarding roosting monarch butterflies in the Journey North project (Davis and Howard, 2005; Howard and Davis, 2009; Howard *et al.*, 2010). The similarities in the data collected include the spatial and temporal aspects of these butterfly activities for example the date and location which would be important information also captured by citizen scientists. While posts about *Belenois aurota* larvae was not considered for this study, many posts did mention the larval phase. This type of information could facilitate larvae monitoring projects like MLMP (Prysby and Oberhauser, 2004) but focusing on *Belenois aurota* or other local butterflies. Hart *et al.* (2018) argue that mining data for ecological purposes from social media is a form of indirect citizen science. Obtaining data from non-expert sources - such as those on social media - about ecological phenomena is the same as non-expert sources for many citizen sciences projects (ElQadi *et al.,* 2017; Kankanamge *et al.*, 2020; Reed, 2020; Willis *et al.,* 2017). Jarić *et al.* (2020) outline all the new methodologies and techniques that can be used to obtain and interpret crowdsourced data from social media. These technologies include the methods used in the study such as automated web-scraping and natural language processing but also include technologies such as artificial intelligence (AI) and machine learning (Jarić *et al*., 2020). Jarić *et al.* (2020) demonstrates how data from social media in terms of text, images and online activity, combined with knowledge of phenology and occurrences of *Belenois aurota,* can produce ecological insights (Jarić *et al.,* 2020). Social media can therefore also form part of monitoring programmes for species where these programmes do not exist and can therefore contribute to adding of phenological data where a data paucity currently exists.

6.4 Phenological data and trends

In terms of the use of the social media as a data source for phenology, the dates and locations of the butterfly migration are important as these were integral to the data analysis (Fitchett and Fani, 2018; Hart *et al.,* 2018; Zipper, 2017), although this study shows that social media data can produce neither reliable nor precise data for either which is also confirmed by Hart *et al.* (2018). This is a function of the users ability to provide information in a certain format. Facebook posts revealed there was some certainty when butterflies were sighted. This was most likely due to the type of information shared on the butterfly Facebook pages (Chamberlain, 2018), particularly meeting the requirements for posting to the "The Great Little White Butterfly Migration". Posts from other platforms incurred some ambiguity regarding the date in which the butterfly migration was observed. In some instances, it appeared that users posted old pictures of the butterfly with no context on when the photograph was taken. This will most likely improve with the uptake of smart mobile phones where individuals are able to photograph butterflies and almost instantaneously upload the photographs to social media (Jarić *et al.,* 2020).

Obtaining quality spatial information from both the automatic and manual collection methods data was very challenging for this study. Similar spatial challenges were outlined in other studies as very few social media posts have associated location tags or coordinates (Zipper, 2017; Hart *et al.*, 2018). Locations were often based on the descriptions in social media posts or based on the user biographical information.

Photography may also be regarded as an important aspect when using social media for phenological studies. This need for photographs of the butterfly species becomes apparent when observations of butterflies are posted on the butterfly specific Facebook groups allowing experts and lepidopterists to verify that the correct species was identified. In this study, *Belenois aurota* was sometimes confused with *Belenois gidica abyssinica* (African veined White) and *Belenois creona* (African common White). Photography plays an important role in species identification and verification for the LepiMAP citizen science project. The importance of photography can be seen in other citizen science projects (Jue and Daniels, 2014) and ecological projects using social media as a data source (Marshall and Strine, 2019). In this study, many users across the three social media platforms have complained about the difficulty in capturing photographs of *Belenois aurota* because the butterfly moves so swiftly (ElQadi *et al.,* 2017). Instagram users have often posted about special lenses (either in hashtags or by posting details of equipment used) used in taking pictures of *Belenois aurota* to not disrupt the butterfly whilst taking photographs.

In terms of the changes in migration phenology over time, the results of the study show that there is an advance of 2.4-days/decade in the migration of the butterflies over the period 1914-2020 based on social media and historical data (Fig 5.13). The changes over time for social media contrast to that of LepiMAP which reveals a 2.3 days/decade delay (Fig 5.21) in migration phenology. The reasons for the differing rates of change may be due to the considerable gaps in the social media database (Fitchett *et al.*, 2015; Ledneva *et al.*, 2004), or that the social media records are unreliable (Lopez *et al.*, 2019). Based on the paucity of phenological records in South Africa, however, imperfect records cannot be wholly discounted (Fitchett *et al.*, 2015; Ledneva *et al.*, 2004).

The social media findings of this study are consistent with global butterfly projects that demonstrate that butterfly activity is advancing over time. In the Northern Hemisphere, advances in butterfly phenology have been reported in the Mediterranean (Colom *et al.,* 2020; Stefanescu *et al.*, 2003) and in the United Kingdom (Roy and Sparks, 2000). In the Mediterranean advances relating to temperature increases were found in many species of which only three showed statistically significant results (Colom *et al.,* 2020). In the United Kingdom study, which has data for 22 years, over 70% of butterflies exhibit advances in first appearance dates due to increasing temperatures (Parmesan, 2006; Roy and Sparks, 2000). Butterflies are said to advance their flight period by two days for a 1 °C increase in temperature. However, more extreme advances of up to 16 days were detected in United Kingdom studies (Roy and Sparks, 2000). Polgar (2013) also found that butterfly activity peaks as plants begin to flower whereas other studies suggest that these interactions between plants and insects are not always in synchrony (Parmesan, 2006).

In the South African context, this advance is consistent with the rates of advance observed in other studies conducted on flowering plants (Fitchett and Fani, 2018; Grab and Craparo, 2011; Snyman, 2020). This is most likely due to the butterfly's dependence on plants (Colom *et al.*, 2020; Stefanescu *et al.,* 2011; Wallis De Vries *et al.*, 2011). These advances compare to studies on the flowering of apple trees in the Western Cape province and that of jacaranda trees in the Gauteng provinces of South Africa. The study on apple trees reported a rate of 4.2 days/decade advance in the flowering of Golden Delicious apple trees (Grab and Craparo, 2011), and a 2.4 days/decade advance in the flowering of Granny Smith apple trees for the period 1973-2009 (Grab and Craparo, 2011). There is also consistency with the study on jacaranda trees which reported a 2.4 days/decade advance in the flowering of trees over the period 1924-2016 (Fitchett and Fani, 2018). That study contrasts with the study on the sardine run which reports a delay in the migration of 1.3 days/decade (Fitchett *et al.*, 2019).

6.5 Limitations of the Project

6.5.1 Methods

The 63 historical newspaper records that were from a concurrent study (Pandazis, 2020) were limited to Johannesburg and did not consider any other part of the country. Based on this study being conducted at a national level, ideally nation-wide historical records were needed to match the scale of the social media data collected.

Further analysis could have been conducted to determine spatial relationships such as the regression analysis conducted by Hart *et al.* (2018). The concern regarding this was that spatial locations were at varying levels of accuracy. Some social media records had exact coordinates, while some were at suburb level and a large majority were at city level. Based on this, spatial statistical analysis was not feasible (Hart *et al.*, 2018).

The limitation with the linear regression models used is that these models are not robust when using data that do not have strong linear correlations (Diez *et al.*, 2014). In the case of this study, the data exhibited poor correlation between the variable's year and the median Julian date. Primack *et al.* (2009) caution against the overgeneralisation of phenological shifts as linear relationships seldom exist (Primack *et al.*, 2009).

6.5.2 Biases in sample selected

The nature of the data derived from social media is inherently biased since individuals can choose when to post to social media (Hart *et al.*, 2018). Some reasons that people may have refrained from posting on social media may have been due to personal or cultural reasons. Some may have chosen to not post on social media due to a lack of interest in *Belenois aurota* (ElQadi *et al*., 2017). Another reason may have been users' socioeconomic conditions where some users may not have had internet access or did not have devices from which to interact with social media; not all individuals have access to these (Stats SA, 2019). Furthermore, individuals may have chosen not to contribute on social media because of the lack of ability to communicate in English or because of poor levels of literacy (Bidwell, 2016). There may have been several individuals who have seen the *Belenois aurota* migration and chose not to post about the event on any of the social media platforms. The reasons that people have chosen to contribute may be because they are affiliated with lepidopterist societies, they are butterfly enthusiasts or simply just nature lovers. Those who have interests in photography may also have chosen to share their pictures on social media.

6.5.3 Social and print media as a data source

As the average user has eight social media accounts (Kemp, 2020), there is a possibility that users may post multiple observations of *Belenois aurota* on multiple social media platforms and the citizen science platform LepiMAP. As users may use different names across platforms, duplication in records may have arisen. Duplicate records will result in an inaccurate database and the resulting trends derived from this database may be unreliable (Hart *et al*., 2018).

Upon inspection of comments of posts in the “African Butterflies & Moths – LepSoc” and “The Great Little White Butterfly Migration” Facebook pages, several people were incorrectly identifying *Belenois aurota* and mistaking other species as *Belenois aurota*. Other studies have also identified species identification as a challenge in using social media as a data source (ElQadi *et al*., 2017; Hart *et al.,* 2018). Chamberlain (2018), however, advocates that platforms such as Facebook groups can assist with correct identification of species. This was confirmed in this study of *Belenois aurota*. Species identification is a skill that should not be underestimated as it may affect the outcomes of studies such as this by having databases compiled with records of incorrect species (ElQadi *et al*., 2017; Hart *et al.*, 2018). For this reason, photography and video content in posts are important for verification purposes (ElQadi *et al*., 2017; Hart *et al.*, 2018). Several records in this study did not have any media uploaded and therefore verification was not possible.

Many posts may have been lost due to poor language, lack of hashtag or poor use of the keywords (Chamberlain, 2018). Several posts feature just media with no text component or captions. Another reason for the loss of posts may be due to the language used. South Africa has 11 official languages and posts may have been in one of these languages (Government of South Africa, 2020). A study by Ligidima and Makanansie in 2020 suggested that social media is becoming a platform to promote indigenous languages. There is a strong possibility that some posts may have been lost due to keywords and hashtags being in languages that were not considered part of this study. Examples of words may have been used to identify the butterfly is ‘Witgatwitjie’ which is the Afrikaans word for the species. A word that may have been used more generally is ‘serurubele’ which is the Sotho word for butterflies and moths. The usage of both ’Witgatwitjie’ and ‘serurubele’ were identified in the English social media posts collected as part of this study.

The spatial data obtained in this study was poor. Several posts did not have associated geolocation data. In many cases, the locations were based on descriptions in the posts (Hart *et al.,* 2018), but some posts did not have any specific location. Additionally, the fact that users may have been travelling while posting on social media may further complicate deriving accurate spatial information. This lack of information limited the understanding of where records were posted spatially. However, Marshall and Strine (2019) argue that this will be less problematic over time with the usage of GPS enabled cameras. Another spatial limitation is that this project was conducted at a national scale, but the butterflies are mobile and may move across borders. Future studies may extend the study area to a regional or continental scale as posts from other countries were identified during the data collection phase of this study. This will not only yield more records but also contribute to a holistic understanding of butterfly movements.

It is likely that only a sample of data was obtained for the period of the study as the free version of the data available on Twitter API was used (Lopez *et al.,* 2019). Lopez *et al.* (2019) further assert that studies using social media as a data source are challenging to replicate. Researchers may use the same query to obtain data but may obtain non-identical samples (Lopez *et al.,* 2019). Additional sample limitations include the inconsistency of data saturation across the years. The data coverage for newspapers were intermittent while the social media data were very saturated. This would have implications for the decadal changes calculated as part of this project as the accuracy of these inferences may be compromised.

It is important to consider why users contribute to social media when using these platforms as a source for data (Chamberlain, 2018; Hart *et al*., 2018). Some users may want to share 'pretty pictures' on social media while others may want to contribute to the understanding of a species by posting their observations on social media. While both are valid uses of social media, the latter users may post meaningful content for ecological research. Chamberlain (2018) asserts that only a small percentage of users contribute meaningful information while others just consume site resources by only viewing content. There is agreement with Chamberlain (2018) in this study as a small number of users contributed to most of the content.

It cannot be guaranteed that social media platforms will be available indefinitely (Chamberlain, 2018). Social media platforms such as Panoramio have shut down in the past (Wilkins *et al.,* 2020), and platforms such as Flickr have decreased in popularity (Wilkins *et al.,* 2020). Should this occur in the future, the continuity of data that is available will be compromised (Chamberlain, 2018).

In this chapter, the use of social media as a data source was explained. The spatial and temporal trends in the study were discussed. Opportunities for social media in citizen science studies were then provided. Social media as a source for phenological trends and identifying the change in these trends was examined. Finally, the limitations of this study were provided.

Chapter 7: Conclusion

This is the first study to explore the social media records of the *Belenois aurota* migration. This chapter will examine the extent to which the aims of this study have been achieved. The chapter will also provide a summary of outcomes. Finally, avenues for future research will be provided.

7.1 Achievement of the study aims and objectives

This study aimed to explore the timing of *Belenois aurota* sightings in the annual butterfly migration from a database constructed from social media records. A database of 1,063 records of *Belenois aurota* was constructed from Facebook, Instagram, Twitter, and historical newspapers. The social media component comprised 1,035 records and the historical newspapers comprised of only 28 records. The achievement of the objectives is given below.

1. To compile a database of dates and locations of social media posts relating to the butterfly

A total of 1,063 records on *Belenois aurota* were collected for the period 1914-2020 (Table 4.1). Of these records 703 were from Facebook, 184 were from Instagram, 148 were from Twitter and 28 from historical newspapers. The newspaper records were however were limited to Gauteng. Spatial data was not easily obtained, however, the spatial information this project collected revealed interesting spatial patterns.

2. To assess the interannual variability in the timing of the social media posts on the butterfly migration

The peak posting period on social media and historical newspaper records ranged between November and March each year. Most sightings of the butterfly were reported on Julian days 212-214 (~28-30 January) The range of median Julian dates was 140-269 (~November to March). An advance in the migration phenology of 2.4 days per decade was calculated using a linear regression model. There are however, questions of accuracy regarding the date information of some of the data collected. This will impact the overall accuracy of the decadal changes calculated.

3. To compile a spatial inventory of butterfly sightings which were differentiated by year

Of the 1,035 records collected on social media, only 834 records had a spatial component. The remaining records had either only a reference from which province the post originated or had no spatial information at all. The concentration of records was found consistently in the northern and eastern parts of the country for the period 2011-2020. In contrast with the data from social media, differences were present in the spatial observations of LepiMAP where KwaZulu-Natal accounted for most of the records and Limpopo accounted for the second most.

4. To determine how the interannual variability of the social media and newspaper posts compare to that of the LepiMAP data set for the same period

The peak posting period of LepiMAP differs from the social media dataset. The range of peak observation dates on LepiMAP fall between December and January each year. The range of median Julian dates was 104-275 (~ November to March) for LepiMAP. The migration phenology differed between the social media dataset and LepiMAP. The social medial data shows an advance in migration phenology while LepiMAP shows a delay in migration phenology. A delay of 2.3 days per decade was calculated using a linear regression model on the LepiMAP dataset.

7.2 Summary of Outcomes

The study shows promise in using social media as a data source for *Belenois aurota* phenological records. The data from social media provided this study with spatial and temporal trends in the species that in some instances are comparable to LepiMAP. Trends with global consistency were found when analysing trends using data derived from social media. South Africa has a paucity of phenological data. A study such as this proves that social media can be used as a data source to contribute to the country's phenological database. This methodology also shows potential for researching other migratory butterflies and even other migratory species. This, however, is not without its limitations as several considerations need to be made when using these records for phenological studies, most notably the presence of outliers when analysing data and the users' motivation for posting (Hart *et al.*, 2018). One also needs to consider the popularity of the species as this may determine whether the species gets coverage on social media at all. Jarić *et al.* (2020) assert that this form of research using social media as a data source is likely to increase with time. The low cost and abundance in records outweigh any limitations in the usage of social media as a data source (Breckheimer *et al.,* 2019; Chamberlain, 2018; ElQadi *et al*., 2017; Jarić *et al.,* 2020).

7.3 Avenues for future research

This methodology can form part of a makeshift monitoring system for *Belenois aurota* migration dates. The database can be updated regularly using automatic or manual data collection to obtain the latest annual information from social media. Future research could implement the same methodology for other butterfly species. This could be especially interesting for researching butterflies that form part of the Kalahari Butterfly Migration such as the African Migrant (*Catopsilia florella*).

Reference List

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.H. and Liu, B., 2011. Predicting flu trends using twitter data. In *2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS)*, 702-707.

Allan, S. and Redden, J., 2017. Making citizen science newsworthy in the era of big data. *Journal of Science Communication*, *16*(2), DOI: 10.22323/2.16020305.

Altermatt, F., 2010. Climatic warming increases voltinism in European butterflies and moths. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1685), 1281-1287.

Altermatt, F., 2012. Temperature-related shifts in butterfly phenology depend on the habitat. *Global Change Biology*, *18*(8), 2429-2438.

Altwegg, R., Broms, K., Erni, B., Barnard, P., Midgley, G.F. and Underhill, L.G., 2012. Novel methods reveal shifts in migration phenology of barn swallows in South Africa. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1733), 1485-1490.

Araújo, M.B. and Rahbek, C., 2006. How does climate change affect biodiversity? *Science*, *313*(5792), 1396-1397.

Badeck, F.W., Bondeau, A., Böttcher, K., Doktor, D., Lucht, W., Schaber, J. and Sitch, S., 2004. Responses of spring phenology to climate change. *New Phytologist*, *162*(2), 295-309.

Baksh, M., 2019. *Natural Places & Digital Spaces: Challenges and Opportunities for Instagram in Biodiversity Conservation.* Masters research report submitted to York University, Toronto, Onstario, Canada.

Barberá, P., 2013. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, *23*(1), 76-91.

Balfour, N.J., Ollerton, J., Castellanos, M.C. and Ratnieks, F.L., 2018. British phenological records indicate high diversity and extinction rates among late-summer-flying pollinators. *Biological Conservation*, *222*, 278-283.

Barnard, P. and De Villiers, M., 2012. Biodiversity early warning systems: South African citizen scientists monitoring change. *South African National Biodiversity Institute and Department of Environmental Affairs*, Pretoria, South Africa.

Barr, L., 2020. Social media statistics and usage in South Africa. Available from: https://www.talkwalker.com/blog/social-media-stats-south-africa#:~:text=There%20is%20a%2062%25%20Internet,19%25%20increase%20compared%20to%202019.&text=51%25%20of%20the%20population%20in,is%20female%20and%2049%25%20male, accessed 9 October 2020.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. and Courchamp, F., 2012. Impacts of climate change on the future of biodiversity. *Ecology Letters*, *15*(4), 365-377.

Bidwell, N.J., 2016. Moving the centre to design social media in rural Africa. *AI & Society*, *31*(1), 51-77.

Big Butterfly Count, 2020, About Page. Available from: https://bigbutterflycount.butterfly-conservation.org/about, accessed 9 October 2020.

Bison, M., Yoccoz, N.G., Carlson, B.Z. and Delestrade, A., 2019. Comparison of budburst phenology trends and precision among participants in a citizen science program. *International Journal of Biometeorology*, *63*(1), 61-72.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V. and Shirk, J., 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, *59*(11), 977-984.

Bonney, R., Shirk, J.L., Phillips, T.B., Wiggins, A., Ballard, H.L., Miller-Rushing, A.J. and Parrish, J.K., 2014. Next steps for citizen science. *Science*, *343*(6178), 1436-1437.

Brangbour, E., Bruneau, P., Marchand-Maillet, S., Hostache, R., Matgen, P., Chini, M. and Tamisier, T., 2019. Extracting localized information from a Twitter corpus for flood prevention. *arXiv preprint arXiv:1903.04748*.

Breckheimer, I.K., Theobald, E.J., Cristea, N.C., Wilson, A.K., Lundquist, J.D., Rochefort, R.M. and HilleRisLambers, J., 2019. Crowd-sourced data reveal social–ecological mismatches in phenology driven by climate. *Frontiers in Ecology and the Environment*, *18*(2), 76-82.

Broekhuysen, G.J. and Broekhuysen, J.G.P., 1960. Butterfly migration over Johannesburg. *South African Journal of Science*, *56*(10), 237-239.

Budree, A., Fietkiewicz, K.J. and Lins, E., 2019. Investigating usage of social media platforms in South Africa. *African Journal of Information Systems*, *11*(4), 314-336.

Burgess, H.K., DeBey, L.B., Froehlich, H.E., Schmidt, N., Theobald, E.J., Ettinger, A.K., HilleRisLambers, J., Tewksbury, J., and Parrish, J.K., 2017. The science of citizen science: exploring barriers to use as a primary research tool. *Biological Conservation*, *208*, 113-120.

Burger, A., 2020. Where did all these butterflies come from? Sandton residents all aflutter on Twitter. Available from: https://www.news24.com/news24/southafrica/news/where-did-all-these-butterflies-come-from-sandton-residents-all-aflutter-on-twitter-20200128, accessed 11 March 2020.

Burman, M.S., 2016. *Citizen science reveals complex changes in barn swallow phenology in South Africa over three decades*. Doctoral dissertation submitted to the University of Cape Town.

Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T., Bayne, E. and Boutin, S., 2015. Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675-685.

Callaghan, C.T., Rowley, J.J., Cornwell, W.K., Poore, A.G. and Major, R.E., 2019. Improving big citizen science data: moving beyond haphazard sampling. *PLoS Biology*, *17*(6), DOI: 10.1371/journal.pbio.3000357.

Carew, J., 2014. Online environmental activism in South Africa: A case study of the# IAM4RHINOS Twitter campaign. *Global Media Journal-African Edition*, *8*(2), 207-230.

Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D.S. and Ertl, T., 2012. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *2012 IEEE conference on visual analytics science and technology (VAST)*, 143-152.

Chamberlain, J., 2018. Chapter Five - Using Social Media for Biomonitoring: How Facebook, Twitter, Flickr and Other Social Networking Platforms Can Provide Large-Scale Biodiversity Data. In: Bohan, D.A., Dumbrell, A.J., Woodward, G., Jackson, M. (Eds.), *Advances in Ecological Research, Next Generation Biomonitoring: Part 2*. Academic Press, pp. 133–168.

Chambers, L.E., Altwegg, R., Barbraud, C., Barnard, P., Beaumont, L.J., Crawford, R.J., Durant, J.M., Hughes, L., Keatley, M.R., Low, M. and Morellato, P.C., 2013. Phenological changes in the southern hemisphere. *PloS One*, *8*(10), DOI: 10.1371/journal.pone.0075514.

Chambers, L.E., Plotz, R.D., Dossis, T., Hiriasia, D.H., Malsale, P., Martin, D.J., Mitiepo, R., Tahera, K. and Tofaeono, T.I., 2017. A database for traditional knowledge of weather and climate in the Pacific. *Meteorological Applications*, *24*(3), 491-502.

Chmielewski, F.M. and Rötzer, T., 2002. Annual and spatial variability of the beginning of growing season in Europe in relation to air temperature changes. *Climate Research*, *19*(3), 257-264.

Chuine, I., 2010. Why does phenology drive species distribution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1555), 3149-3160.

Cohn, J.P., 2008. Citizen science: Can volunteers do real research? BioScience, 58(3), 192-197.

Colom, P., Traveset, A., Carreras, D. and Stefanescu, C., 2020. Spatio-temporal responses of butterflies to global warming on a Mediterranean island over two decades. *bioRxiv*, DOI: 10.1111/een.12958.

Cooper, C.B., Dickinson, J., Phillips, T. and Bonney, R., 2007. Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society*, *12*(2), 11.

Crooks, A., Croitoru, A., Stefanidis, A. and Radzikowski, J., 2013. # Earthquake: Twitter as a distributed sensor system. *Transactions in GIS*, *17*(1), 124-147.

Daume, S. and Galaz, V., 2016. "Anyone Know What Species This Is?"–Twitter Conversations as Embryonic Citizen Science Communities. *PloS One*, *11*(3), DOI: 10.1371/journal.pone.0151387.

Davis, A.K. and Howard, E., 2005. Spring recolonization rate of monarch butterflies in eastern North America: new estimates from citizen-science data. *Journal of the Lepidopterists' Society*, *59*(1), 1-5.

De Longueville, B., Smith, R.S. and Luraschi, G., 2009. " OMG, from here, I can see the flames!" a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the 2009 international workshop on location based social networks*, pp. 73-80.

Devictor, V., Van Swaay, C., Brereton, T., Brotons, L., Chamberlain, D., Heliölä, J., Herrando, S., Julliard, R., Kuussaari, M., Lindström, Å. and Reif, J., 2012. Differences in the climatic debts of birds and butterflies at a continental scale. *Nature Climate Change*, *2*(2), 121-124.

Diamond, S.E., Frame, A.M., Martin, R.A., and Buckley, L.B., 2011. Species' traits predict phenological responses to climate change in butterflies. *Ecology*, *92*(5), 1005-1012.

Dickinson, J.L., Zuckerberg, B. and Bonter, D.N., 2010. Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 149-172.

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T. and Purcell, K., 2012. The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291-297.

Dickson, C. and Schofield, T., 1958. Observations on the migration of Belenois aurota (F.) (Lepidoptera: Pieridae). *Journal of the Entomological Society of Southern Africa*, *21*, 427-428.

Dennis, E.B., Morgan, B.J., Brereton, T.M., Roy, D.B. and Fox, R., 2017. Using citizen science butterfly counts to predict species population trends. *Conservation Biology*, *31*(6), 1350-1361.

Demuth, J.L., Morss, R.E., Palen, L., Anderson, K.M., Anderson, J., Kogan, M., Stowe, K., Bica, M., Lazrus, H., Wilhelmi, O. and Henderson, J., 2018. "Sometimes da# beachlife ain't always da wave": Understanding people's evolving hurricane risk communication, risk assessments, and responses using Twitter narratives. *Weather, Climate, and Society*, *10*(3), 537-560.

Diez, D.M., Barr, C.D. and Çetinkaya-Rundel, M., 2014. *Introductory statistics with randomization and simulation*. OpenIntro. Available from: https://www.openintro.org/, accessed 12 March 2019.

Di Minin, E., Tenkanen, H. and Toivonen, T., 2015. Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, *3*, DOI: 10.3389/fenvs.2015.00063.

Dreyer, L., Esler, K.J. and Zietsman, J., 2006. Flowering phenology of South African Oxalis—possible indicator of climate change? *South African Journal of Botany*, *72*(1), 150-156.

Earle, P.S., Bowden, D.C. and Guy, M., 2012. Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, *54*(6), 708-715.

eButterfly, 2020. eButterfly Quick Start Guide. Available from: https://help.e-butterfly.org/en/support/solutions/articles/44000493156-ebutterfly-quick-start-guide, accessed 9 October 2020.

Edge, D.A. and Mecenero, S., 2015. Butterfly conservation in southern Africa. *Journal of Insect Conservation*, *19*(2), 325-339.

Eisenhauer, N., Herrmann, S., Hines, J., Buscot, F., Siebert, J. and Thakur, M.P., 2018. The dark side of animal phenology. *Trends in Ecology & Evolution*, *33*(12), 898-901.

ElQadi, M.M., Dorin, A., Dyer, A., Burd, M., Bukovac, Z. and Shrestha, M., 2017. Mapping species distributions with social media geo-tagged images: case studies of bees and flowering plants in Australia. *Ecological Informatics*, *39*, 23-31.

Engelbrecht, C.J. and Engelbrecht, F.A., 2016. Shifts in Köppen-Geiger climate zones over southern Africa in relation to key global temperature goals. *Theoretical and Applied Climatology*, *123*(1-2), 247-261.

Fatanti, M.N. and Suyadnya, I.W., 2015. Beyond user gaze: How Instagram creates tourism destination brand? *Procedia - Social and Behavioral Sciences*, *211*, 1089-1095.

Feinerer, I., 2013. Introduction to the tm Package Text Mining in R. Available from: http://cran. r-project. org/web/packages/tm/vignettes/tm. Pdf, accessed 9 October 2020.

Ferwerda, B., Schedl, M. and Tkalcic, M., 2015, September. Predicting personality traits with Instagram pictures. In *Proceedings of the 3rd workshop on emotions and personality in personalized systems 2015*, pp. 7-10.

Feldman, R.E., Žemaitė, I. and Miller-Rushing, A.J., 2018. How training citizen scientists affects the accuracy and precision of phenological data. *International Journal of Biometeorology*, *62*(8), 1421-1435.

Fitchett, J.M., Grab, S.W., Thompson, D.I. and Roshan, G., 2014. Spatio-temporal variation in phenological response of citrus to climate change in Iran: 1960–2010. *Agricultural and Forest Meteorology*, *198*, 285-293.

Fitchett, J.M., Grab, S.W., and Thompson, D.I., 2015. Plant phenology and climate change: Progress in methodological approaches and application. *Progress in Physical Geography*, *39*(4), 460-482.

Fitchett, J.M. and Fani, H.A., 2018. Advance In Jacaranda Blossom Phenology In The South African Highveld Interior: A Preliminary Analysis. *South African Geographers*, *1*, 467-486.

Fitchett, J. and Hoogendoorn, G., 2018. An analysis of factors affecting tourists' accounts of weather in South Africa. *International Journal of Biometeorology*, *62*(12), 2161-2172.

Fitchett, J.M., Grab, S.W. and Portwig, H., 2019. Progressive delays in the timing of sardine migration in the southwest Indian Ocean. *South African Journal of Science*, *115*(7-8), 1-6.

FitzPatrick Institute of African Ornithology, 2020. *Virtual Museum Projects*. Available from: http://vmus.adu.org.za/vm_projects.php?database=&prj_acronym=&db=&URL=&Logo=&Headline=&Use_main_filter=0&User_id=&Full_name=&serve_sp_list=0&drop_down_list=0&assessment=0&query_id=0&Vm_number=0&recNo=0&numRows=0&start=0, accessed 27 July 2020.

Follett, R. and Strezov, V., 2015. An analysis of citizen science-based research: usage and publication patterns. *PloS One*, *10*(11), DOI: 10.1371/journal.pone.0143687.

Forister, M.L. and Shapiro, A.M., 2003. Climatic trends and advancing spring flight of butterflies in lowland California. *Global Change Biology*, *9*(7), 1130-1135.

Global Biodiversity Information Facility (GBIF), 2020. Occurrence Map of Belenois aurota, Available from: Global Biodiversity Information Facility, https://www.gbif.org/occurrence/map?has_coordinate=true&has_geospatial_issue=false&taxon_key=1919181, accessed 28 February 2020.

Global Biodiversity Information Facility (GBIF), 2020b. Pieris rapae (Linnaeus, 1758). Available from: https://www.gbif.org/species/1920496, accessed 20 October 2020.

Golumbic, Y.N., 2015. *What makes citizen science projects successful, and what can we learn from them for future projects*. Technion Citizen Science Project (TCSP) report, pp. 1-56.

Gomide, J., Veloso, A., Meira Jr, W., Almeida, V., Benevenuto, F., Ferraz, F. and Teixeira, M., 2011, June. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In *Proceedings of the 3rd international web science conference*, pp.1-8.

Government of South Africa, 2020. About SA. Available from: https://www.gov.za/about-sa, accessed 9 October 2020.

Grab, S. and Craparo, A., 2011. Advance of apple and pear tree full bloom dates in response to climate change in the southwestern Cape, South Africa: 1973–2009. *Agricultural and Forest Meteorology*, *151*(3), 406-413.

Graham, E.A., Henderson, S. and Schloss, A., 2011. Using mobile phones to engage citizen scientists in research. *Eos, Transactions American Geophysical Union*, *92*(38), 313-315.

Greenwood, J. J. D. (2007). Citizens, science, and bird conservation. *Journal of Ornithology*, *148*, DOI: 10.1007/s10336-007-0239-9.

Grolemund, G. and Wickham, H., 2011. Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1-25.

Hart, A.G., Carpenter, W.S., Hlustik-Smith, E., Reed, M., and Goodenough, A.E., 2018. Testing the potential of Twitter mining methods for data acquisition: Evaluating novel opportunities for ecological research in multiple taxa. *Methods in Ecology and Evolution*, *9*(11), 2194-2205.

Hegland, S.J., Nielsen, A., Lázaro, A., Bjerknes, A.L. and Totland, Ø., 2009. How does climate warming affect plant-pollinator interactions? *Ecology Letters*, *12*(2), 184-195.

Hofmeester, T.R., Young, S., Juthberg, S., Singh, N.J., Widemo, F., Andrén, H., Linnell, J.D. and Cromsigt, J.P., 2019. Using by-catch data from wildlife surveys to quantify climatic parameters and timing of phenology for plants and animals using camera traps. *Remote Sensing in Ecology and Conservation*, *6*(2), 129-140.

Hofmeyr, S.D., 2012. *Impacts of environmental change on large terrestrial bird species in South Africa: insights from citizen science data*. Doctoral dissertation submitted to the University of Cape Town.

Hofmeyr, S.D., Symes, C.T. and Underhill, L.G., 2014. Secretarybird Sagittarius serpentarius population trends and ecology: insights from South African citizen science data. *PloS One*, *9*(5), DOI: 10.1371/journal.pone.0096772.

Howard, E., and Davis, A.K., 2009. The fall migration flyways of monarch butterflies in eastern North America revealed by citizen scientists. *Journal of Insect Conservation*, *13*(3), 279-286.

Howard, E., Aschen, H. and Davis, A.K., 2010. Citizen science observations of monarch butterfly overwintering in the southern United States. *Psyche*, DOI: 10.1155/2010/689301.

Illán, J.G., Gutierrez, D., Diez, S.B. and Wilson, R.J., 2012. Elevational trends in butterfly phenology: implications for species responses to climate change. *Ecological Entomology*, *37*(2), 134-144.

Isaac, N.J., van Strien, A.J., August, T.A., de Zeeuw, M.P. and Roy, D.B., 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, *5*(10), 1052-1060.

Jacobs, K., Sandberg, L. and Spierings, N., 2020. Twitter and Facebook: Populists' double-barreled gun? *New Media & Society*, *22*(4), 611-633.

Jansen van Rensburg, T.,2020. South African social media stats for 2020. Available from: https://www.mediaupdate.co.za/social/149345/south-african-social-media-stats-for-2020, accessed 11 March 2020.

Jarić, I., Correia, R.A., Brook, B.W., Buettel, J.C., Courchamp, F., Di Minin, E., Firth, J.A., Gaston, K.J., Jepson, P., Kalinkat, G. and Ladle, R., 2020. iEcology: Harnessing Large Online Resources to Generate Ecological Insights. *Trends in Ecology & Evolution, 35*(10), 630-639.

Journey North, 2020. Citizen Science. Available from: https://journeynorth.org/tm/monarch/indexFall.html, accessed 9 October 2020.

Jue, D.K. and Daniels, J.C., 2014. A successful model for citizen scientist involvement in building a statewide at-risk butterfly database. *Journal of Insect Conservation*, *19*(2), 421-431.

Jury, M.R., 2018. Climate trends across South Africa since 1980. *Water SA*, *44*(2), 297-307.

Kahle, D. and Wickham, H., 2013. ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144-161.

Kankanamge, N., Yigitcanlar, T., Goonetilleke, A. and Kamruzzaman, M., 2020. Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets. *International journal of disaster risk reduction*, *42*, DOI: 10.1016/j.ijdrr.2019.101360.

Karimkhani, C., Connett, J., Boyers, L., Quest, T. and Dellavalle, R.P., 2014. Dermatology on instagram. *Dermatology Online Journal*, *20*(7), 13030-qt71g178w9.

Kearney, M.W., 2019. rtweet: Collecting and analyzing Twitter data. *Journal of Open-Source Software*, *4*(42), DOI: 10.21105/joss.01829.

Kelling, S., Johnston, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., Bonn, A., Fernandez, M., Hochachka, W.M., Julliard, R., Kraemer, R. and Guralnick, R., 2018. Finding the signal in the Noise of Citizen Science Observations. *bioRxiv*, DOI: 10.1093/biosci/biz010.

Kemp, S.,2020. Digital 2020: South Africa. Available from: https://datareportal.com/reports/digital-2020-south-africa?rq=digital%202020%20south%20africa, accessed 9 October 2020.

Kireyev, K., Palen, L. and Anderson, K., 2009, December. Applications of topics models to analysis of disaster-related twitter data. In *NIPS workshop on applications for topic models: text and beyond*, *1*. Canada: Whistler.

Kirilenko, A.P., Molodtsova, T. and Stepchenkova, S.O., 2015. People as sensors: Mass media and local temperature influence climate change discussion on Twitter. *Global Environmental Change*, *30*, 92-100.

Kosmala, M., Crall, A., Cheng, R., Hufkens, K., Henderson, S. and Richardson, A.D., 2016. Season Spotter: Using citizen science to validate and scale plant phenology from near-surface remote sensing. *Remote Sensing*, *8*(9), DOI: 10.3390/rs8090726.

Kountoupes, D.L. and Oberhauser, K., 2008. Citizen science and youth audiences: educational outcomes of the Monarch Larva Monitoring Project. *Journal of Community Engagement and Scholarship*, *1*(1), 10-20.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. and Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259-263.

Krotov, V. and Silva, L., 2018. Legality and Ethics of Web Scraping. AMCIS 2018

Proceedings.

Lacassin, R., Devès, M., Hicks, S.P., Ampuero, J.P., Bossu, R., Bruhat, L., Wibisono, D.F., Fallou, L., Fielding, E.J., Gabriel, A.A. and Gurney, J., 2020. Rapid collaborative knowledge building via Twitter after significant geohazard events. *Geoscience Communication*, *3*(1), 129-146.

Lampos, V. and Cristianini, N., 2010, June. Tracking the flu pandemic by monitoring the social web. In *2010 2nd International workshop on cognitive information processing*, 411-416.

Ledneva, A., Miller-Rushing, A.J., Primack, R.B. and Imbres, C., 2004. Climate change as reflected in a naturalist's diary, Middleborough, Massachusetts. *The Wilson Journal of Ornithology*, *116*(3), 224-231.

Lee, A.T.K. and Nel, H., 2020. BirdLasser: The influence of a mobile app on a citizen science project. *African Zoology*, DOI: 10.1080/15627020.2020.1717376.

Leocadio, J., Veiga, A., Barbosa, M., Albertini, B., Lohmann, L. and Saraiva, A., 2018. A multi-platform mobile application to collect citizen science data for Bignoniaceae phenological research. *Biodiversity Information Science and Standards, 2*, DOI: 10.3897/biss.2.25582.

Liberatore, A., Bowkett, E., MacLeod, C.J., Spurr, E. and Longnecker, N., 2018. Social media as a platform for a citizen science community of practice. *Citizen Science: Theory and Practice*, *3*(1) DOI: 10.5334/cstp.108.

Liebenberg, L., Steventon, J., Brahman, N., Benadie, K., Minye, J. and Langwane, H.K., 2016. Smartphone icon user interface design for non-literate trackers and its implications for an inclusive citizen science. *Biological Conservation*, *208*, 155-162.

Ligidima, M. and Makananise, F.O., 2020. Social media as a communicative platform to promote indigenous African languages by youth students at a rural based university, South Africa. *Gender & Behaviour*, *18*(2), 15824-15832.

Loftie-Eaton, M., 2016. Report on the Spring LepiBASH for LepiMAP, 15-23 October 2016. *Biodiversity Observations*, 1-11.

Loftie-Eaton, M., 2019. How to submit records to the Virtual Museum. Available from: http://thebdi.org/2019/01/30/how-to-biomap/, accessed 30 October 2020.

Long, J. and Robertson, C., 2018. Comparing spatial patterns. *Geography Compass*, *12*(2), DOI: 10.1111/gec3.12356.

Lopez, B.E., Magliocca, N.R. and Crooks, A.T., 2019. Challenges and opportunities of social media data for socio-environmental systems research. *Land*, *8*(7), DOI: 10.3390/land8070107.

Lup, K., Trub, L. and Rosenthal, L., 2015. Instagram# instasad? exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking*, *18*(5), 247-252.

MacKenzie, C.M., Murray, G., Primack, R. and Weihrauch, D., 2017. Lessons from citizen science: assessing volunteer-collected plant phenology data with Mountain Watch. *Biological Conservation*, *208*, 121-126.

MacEachren, A.M., Robinson, A.C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J. and Mitra, P., 2011, July. Geo-twitter analytics: Applications in crisis management. In *25th International cartographic conference*, pp. 3-8.

Mahlangu, M.N., and Fitchett, J.M., 2019. Climate change threats to a floral wedding: Threats of shifting phenology to the emerging South African wedding industry. *Bulletin of Geography. Socio-economic Series*, *45*(45), 7-23.

Mai, E. and Hranac, R., 2013, January. Twitter interactions as a data source for transportation incidents. In *Proc. Transportation Research Board 92nd Ann. Meeting*, *13*, p. 1636.

Mao, Y., Wei, W., Wang, B. and Liu, B., 2012, August. Correlating S&P 500 stocks with Twitter data. In *Proceedings of the first ACM nternational workshop on hot topics on interdisciplinary social networks research*, pp. 69-72.

Marshall, B.M. and Strine, C.T., 2019. Exploring snake occurrence records: Spatial biases and marginal gains from accessible social media. *PeerJ*, *7*, DOI: 10.7717/peerj.8059.

Martín, Y., Cutter, S.L. and Li, Z., 2020. Bridging twitter and survey data for evacuation assessment of Hurricane Matthew and Hurricane Irma. *Natural Hazards Review*, *21*(2), DOI: 10.1061/(ASCE)NH.1527-6996.0000354.

Mbhele, S., 2020. White butterfly migration has Twitter abuzz – not just in Sandton. Available from https://sandtonchronicle.co.za/247276/white-butterfly-migration-twitter-abuzz-not-just-sandton/, accessed 9 May 2020.

McKenna, B., Myers, M.D. and Newman, M., 2017. Social media in qualitative research: Challenges and recommendations. *Information and Organization*, *27*(2), 87-99.

Menzel, A., Sparks, T.H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kübler, K., Bissolli, P., Braslavská, O.G., Briede, A. and Chmielewski, F.M., 2006. European phenological response to climate change matches the warming pattern. *Global Change Biology*, *12*(10), 1969-1976.

Miguéns, J., Baggio, R. and Costa, C. (2008). Social media and tourism destinations: TripAdvisor Case Study. *Proceedings of the advances in tourism research conference 2008*, Aveiro, 26-28 May.

Milne, B.G., 1964. A migration of Belenois aurota F., and Catopsilia florella L. (Lepidoptera: Pieridae) in the Transvaal. *Journal of the Entomological Society of Southern Africa*, *27*(1), 37-42.

Møller, A.P., Nuttall, R., Piper, S.E., Szép, T. and Vickers, E.J., 2011. Migration, moult, and climate change in barn swallows Hirundo rustica in South Africa. *Climate Research*, *47*(3), 201-205.

Mouritsen, H., Derbyshire, R., Stalleicken, J., Mouritsen, O.Ø., Frost, B.J. and Norris, D.R., 2013. An experimental displacement and over 50 years of tag-recoveries show that monarch butterflies are not true navigators. *Proceedings of the National Academy of Sciences*, *110*(18), 7348-7353.

Moyo, H., Scholes, M.C. and Twine, W., 2015. Effects of water and nutrient additions on the timing and duration of phenological stages of resprouting Terminalia sericea. *South African Journal of Botany*, *96*, 85-90.

Moyo, M.S., 2018. *The phenology of the Enkangala Grasslands*. Master of Science dissertation submitted to the University of the Witwatersrand: Johannesburg.

Murthy, D. and Longwell, S.A., 2013. Twitter and disasters: The uses of Twitter during the 2010 Pakistan floods. *Information, Communication & Society*, *16*(6), 837-855.

Mucina, L., Rutherford, M.C. and Powrie, L.W., 2006. Vegetation Atlas of South Africa, Lesotho, and Swaziland. *The Vegetation of South Africa, Lesotho and Swaziland'.* Edited by Mucina, L. and Rutherford, M.C. pp.748-789.

Namugera, F., Wesonga, R. and Jehopio, P., 2019. Text mining and determinants of sentiments: Twitter social media usage by traditional media houses in Uganda. *Computational Social Networks*, *6*(1), DOI: 10.1186/s40649-019-0063-4.

Newson, S.E., Moran, N.J., Musgrove, A.J., Pearce-Higgins, J.W., Gillings, S., Atkinson, P.W., Miller, R., Grantham, M.J. and Baillie, S.R., 2016. Long-term changes in the migration phenology of UK breeding birds detected by large-scale citizen science recording schemes. *Ibis*, *158*(3), 481-495.

Newman, G., Zimmerman, D., Crall, A., Laituri, M., Graham, J. and Stapel, L., 2010. User-friendly web mapping: lessons from a citizen science website. *International Journal of Geographical Information Science*, *24*(12), 1851-1869.

National Oceanic and Atmospheric Administration(NOAA), 2020. Cold & Warm Episodes by Season. Available from: https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php, accessed 20 January 2021.

Panchariya, N.S., DeStefano, A.J., Nimbagal, V., Ragupathy, R., Yavuz, S., Herbert, K.G., Hill, E. and Fails, J.A., 2015, March. Current developments in big data and sustainability sciences in mobile citizen science applications. In *2015 IEEE First international conference on big data computing service and applications*, 202-212.

Pandazis, A., (2020). *Exploring temporal variation in the Brown-Veined White Butterfly dispersal phenology within South Africa over the past century*. Honours research report submitted to the University of the Witwatersrand: Johannesburg.

Parmesan, C. and Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421,* 37-42.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.*, *37*, 637-669.

Périquet, S., Roxburgh, L., le Roux, A., and Collinson, W.J., 2018. Testing the value of citizen science for roadkill studies: a case study from South Africa. *Frontiers in Ecology and Evolution*, *6*, DOI: 10.3389/fevo.2018.00015.

Pittman, M. and Reich, B., 2016. Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155-167.

Pieris Project. 2020. Project protocol. Available from: https://www.pierisproject.org/participate.html, accessed 10 October 2020.

Pretorius, M., Broders, H., Seamark, E. and Keith, M., 2020. Climatic correlates of migrant Natal long-fingered bat (Miniopterus natalensis) phenology in north-eastern South Africa. *Wildlife Research*, *47*(5), 404-414.

Primack, R.B., Ibáñez, I., Higuchi, H., Lee, S.D., Miller-Rushing, A.J., Wilson, A.M. and Silander Jr, J.A., 2009. Spatial and interspecific variability in phenological responses to warming temperatures. *Biological Conservation*, *142*(11), 2569-2577.

Prudic, K.L., McFarland, K.P., Oliver, J.C., Hutchinson, R.A., Long, E.C., Kerr, J.T. and Larrivée, M., 2017. eButterfly: leveraging massive online citizen science for butterfly conservation. *Insects*, *8*(2), DOI: 10.3390/insects8020053.

Prysby, M.D. and Oberhauser, K.S., 2004. Temporal and geographic variation in monarch densities: citizen scientists document monarch population patterns. *The monarch butterfly: Biology and Conservation*, 9-20.

Radchuk, V., Reed, T., Teplitsky, C., Van De Pol, M., Charmantier, A., Hassall, C., Adamík, P., Adriaensen, F., Ahola, M.P., Arcese, P. and Avilés, J.M., 2019. Adaptive responses of animals to climate change are most likely insufficient. *Nature Communications*, *10*(1), DOI: 10.1038/s41467-019-10924-4.

Rafferty, N.E., and Ives, A.R., 2011. Effects of experimental shifts in flowering phenology on plant-pollinator interactions. *Ecology Letters*, *14*(1), 69-74.

Reece, A.G. and Danforth, C.M., 2017. Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(1), DOI: 10.1140/epjds/s13688-017-0110-z.

Reed, M., 2020. Scientific citizens, smartphones and social media–reshaping the socio-spatial networks of participation: Insects, soil and food. *Moravian Geographical Reports*, *28*(1), 61-67.

Roffe, S.J., Fitchett, J.M. and Curtis, C.J., 2019. Classifying and mapping rainfall seasonality in South Africa: A review. *South African Geographical Journal*, *101*(2), 158-174.

Roffe, S.J., Fitchett, J.M. and Curtis, C.J., 2020. Investigating changes in rainfall seasonality across South Africa: 1987–2016. *International Journal of Climatology*, DOI: 10.1002/joc.6830.

Root, T.L., Price, J.T., Hall, K.R., Schneider, S.H., Rosenzweig, C. and Pounds, J.A., 2003. Fingerprints of global warming on wild animals and plants. *Nature*, *421*(6918), 57-60.

Rose, S., Suri, J., Brooks, M., and Ryan, P.G., 2020. COVID-19 and citizen science: lessons learned from southern Africa. *Ostrich*, *91*(2), 188-191.

Roy, D.B. and Sparks, T.H., 2000. Phenology of British butterflies and climate change. *Global Change Biology*, *6*(4), 407-416.

Roy, D.B., Rothery, P., Moss, D., Pollard, E. and Thomas, J.A., 2001. Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. *Journal of Animal Ecology*, *70*(2), 201-217.

Ryan, S.F., Lombaert, E., Espeset, A., Vila, R., Talavera, G., Dincă, V., Doellman, M.M., Renshaw, M.A., Eng, M.W., Hornett, E.A. and Li, Y., 2019. Global invasion history of the agricultural pest butterfly Pieris rapae revealed with genomics and citizen science. *Proceedings of the National Academy of Sciences*, *116*(40), DOI: 10.1073/pnas.1907492116.

Sachs, J. and Finin, T., 2011, August. Social and semantic computing in support of citizen science. In *Proceedings of the Third Canadian Semantic Web Symposium*, pp. 46-51.

Sakaki, T., Okazaki, M. and Matsuo, Y., 2010, April. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pp. 851-860.

South African National Biodiversity Institute (2006-2018). *The Vegetation Map of South Africa, Lesotho, and Swaziland*. Edited by Mucina, L., Rutherford, M.C. and Powrie, L.W. Available from: http://bgis.sanbi.org/SpatialDataset/Detail/18, Version 2018.

Schmucki, R., Pe'Er, G., Roy, D.B., Stefanescu, C., Van Swaay, C.A., Oliver, T.H., Kuussaari, M., Van Strien, A.J., Ries, L., Settele, J. and Musche, M., 2016. A regionally informed abundance index for supporting integrative analyses across butterfly monitoring schemes. *Journal of Applied Ecology*, *53*(2), 501-510.

Schleuning, M., Neuschulz, E.L., Albrecht, J., Bender, I.M., Bowler, D.E., Dehling, D.M., Fritz, S.A., Hof, C., Mueller, T., Nowak, L. and Sorensen, M.C., 2020. Trait-based assessments of climate-change impacts on interacting species. *Trends in Ecology & Evolution*, *35*(4), 319-328.

Schultz, C.B., Brown, L.M., Pelton, E., and Crone, E.E., 2017. Citizen science monitoring demonstrates dramatic declines of monarch butterflies in western North America. *Biological Conservation*, *214*, 343-346.

Schwartz, M.D., 1999. Advancing to full bloom: planning phenological research for the 21st century. *International Journal of Biometeorology*, *42*(3), 113-118.

Shackleton, C.M., 1999. Rainfall and topo-edaphic influences on woody community phenology in South African savannas. *Global Ecology and Biogeography*, *8*(2), 125-136.

Sheldon, P. and Bryant, K., 2016. Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, *58*, 89-97.

Sheldon, P. and Newman, M., 2019. Instagram and American teens: understanding motives for its use and relationship to excessive reassurance-seeking and interpersonal rejection. *The Journal of Social Media in Society*, *8*(1), 1-16.

Silva, S.J., Barbieri, L.K. and Thomer, A.K., 2018. Observing vegetation phenology through social media. *Plos One*, *13*(5), DOI: 10.1371/journal.pone.0197325.

Sin, H.L. and He, S., 2018. Voluntouring on Facebook and Instagram: Photography and social media in constructing the 'Third World 'experience. *Tourist Studies*, *19*(2), 215-237.

Sinnenberg, L., DiSilvestro, C.L., Mancheno, C., Dailey, K., Tufts, C., Buttenheim, A.M., Barg, F., Ungar, L., Schwartz, H., Brown, D. and Asch, D.A., 2016. Twitter as a potential data source for cardiovascular disease research. *JAMA Cardiology*, *1*(9), 1032-1036.

Smith, L., Liang, Q., James, P. and Lin, W., 2017. Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *Journal of Flood Risk Management*, *10*(3), 370-380.

Snyman, P. L., (2020*). Phenological advance in the South African Namaqualand Daisy Bloom over the past decades.* Master of Science dissertation submitted to the University of the Witwatersrand: Johannesburg.

Socolar, J.B., Epanchin, P.N., Beissinger, S.R. and Tingley, M.W., 2017. Phenological shifts conserve thermal niches in North American birds and reshape expectations for climate-driven range shifts. *Proceedings of the National Academy of Sciences*, *114*(49), 12976-12981.

Sotiriadis, M.D. and Van Zyl, C., 2013. Electronic word-of-mouth and online reviews in tourism services: the use of twitter by tourists. *Electronic Commerce Research*, *13*(1), 103-124.

Sparks, T.H. and Yates, T.J., 1997. The effect of spring temperature on the appearance dates of British butterflies 1883–1993. *Ecography*, *20*(4), 368-374.

Stats, S.A., 2019. *General Household Survey, Statistical Release P0318*. Statistics South Africa, Pretoria.

Steiger, E., De Albuquerque, J.P. and Zipf, A., 2015. An advanced systematic literature review on spatiotemporal analyses of twitter data. *Transactions in GIS*, *19*(6), 809-834.

Stefanescu, C., Peñuelas, J. and Filella, I., 2003. Effects of climatic change on the phenology of butterflies in the northwest Mediterranean Basin. *Global Change Biology*, *9*(10), 1494-1506.

Stefanescu, C., Carnicer, J. and Peñuelas, J., 2011. Determinants of species richness in generalist and specialist Mediterranean butterflies: the negative synergistic forces of climate and habitat change. *Ecography*, *34*(3), 353-363.

Stefanidis, A., Crooks, A. and Radzikowski, J., 2011. Harvesting ambient geospatial information from social media feeds. *GeoJournal*, *78*(2), 319-338.

Superio, D.L., Luceño, M.T., Bendalian, M.V.T., Yap-Zerudo, A.M.A., Ciriaco, J.A. and Estante-Superio, E.G., 2020. Pawikan on FB: Facebook as a Source of Information on the Current State of Sea Turtles in the Philippines with an Emphasis on Mortality. In: Living on the Edge: 45th IAMSLIC Annual Conference , 20-25 Oct 2019 ,Port Aransas, Texas, USA.

Taylor, J.S., 1959. A migratory flight of Belenois aurota (F.) Port Elizabeth. *Journal of the Entomological Society of Southern Africa*, *22*(2), 420-422.

Taylor, S.D., 2019. Estimating flowering transition dates from status-based phenological observations: a test of methods. *PeerJ*, *7*, DOI 10.7717/peerj.7720.

Taylor Jr, O.R., Pleasants, J.M., Grundel, R., Pecoraro, S.D., Lovett, J.P. and Ryan, A., 2020. Evaluating the migration mortality hypothesis using monarch tagging data. *Frontiers in Ecology and Evolution*, *8*, DOI: 10.3389/fevo.2020.00264.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L. and Toivonen, T., 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports*, *7*(1), DOI: 10.1038/s41598-017-18007-4.

Terblanche, R.F., 2014. Butterfly migrations gracing the face of Southern Africa in summer. *Babbel Blues, 1*, 16-20.

Terblanche, R.F., 2020.Current waves of a butterfly migration over southern Africa. *LepSoc Africa.*

Terpstra, T., Stronkman, R., de Vries, A. and Paradies, G.L., 2012, April. Towards a realtime Twitter analysis during crises for operational crisis management. In *ISCRAM*, pp. 1-9.

Thabethe, V. and Downs, C.T., 2018. Citizen science reveals widespread supplementary feeding of African woolly-necked storks in suburban areas of KwaZulu-Natal, South Africa. *Urban Ecosystems*, *21*(5), 965-973.

Thackeray, S.J., Henrys, P.A., Hemming, D., Bell, J.R., Botham, M.S., Burthe, S., Helaouet, P., Johns, D.G., Jones, I.D., Leech, D.I. and Mackay, E.B., 2016. Phenological sensitivity to climate across taxa and trophic levels. *Nature*, *535*(7611), 241-245.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F., De Siqueira, M.F., Grainger, A., Hannah, L. and Hughes, L., 2004. Extinction risk from climate change. *Nature*, *427*(6970), 145-148.

Tulloch, A.I. and Szabo, J.K., 2012. A behavioural ecology approach to understand volunteer surveying for citizen science datasets. *Emu-Austral Ornithology*, *112*(4), 313-325.

Van der Walt, J., 2016. Butterfly Migration in South Africa. Available from: https://www.wildlifesouthafrica.com/blog/wildlife-nature/butterfly-migration-in-south-africa, accessed 11 March 2020.

van Swaay, C.A., Nowicki, P., Settele, J. and Van Strien, A.J., 2008. Butterfly monitoring in Europe: methods, applications, and perspectives. *Biodiversity and Conservation*, *17*(14), 3455-3469.

Van Wyk, A., Van den Berg, J. and Van Hamburg, H., 2008. Diversity and comparative phenology of Lepidoptera on Bt and non-Bt maize in South Africa. *International Journal of Pest Management*, *54*(1), 77-87.

Venter, S.M. and Witkowski, E.T., 2019. Phenology, flowering, and fruit-set patterns of baobabs, Adansonia digitata, in southern Africa. *Forest Ecology and Management*, *453*, DOI: 10.1016/j.foreco.2019.117593.

WallisDeVries, M.F. and Van Swaay, C.A., 2006. Global warming and excess nitrogen may induce butterfly decline by microclimatic cooling. *Global Change Biology*, *12*(9), 1620-1626.

WallisDeVries, M.F., Baxter, W., and Van Vliet, A.J., 2011. Beyond climate envelopes: effects of weather on regional population trends in butterflies. *Oecologia*, *167*(2), DOI: 10.1007/s00442-011-2007-z.

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J., Fromentin, J.M., Hoegh-Guldberg, O. and Bairlein, F., 2002. Ecological responses to recent climate change. *Nature*, *416*(6879), 389-395.

Wang, H., Can, D., Kazemzadeh, A., Bar, F. and Narayanan, S., 2012, July. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations,* 115-120.

Wessels, K., Steenkamp, K., Von Maltitz, G. and Archibald, S., 2011. Remotely sensed vegetation phenology for describing and predicting the biomes of South Africa. *Applied Vegetation Science*, *14*(1), 49-66.

Whitecross, M.A., Witkowski, E.T.F. and Archibald, S., 2016. No two are the same: Assessing variability in broad-leaved savanna tree phenology, with watering, from 2012 to 2014 at Nylsvley, South Africa. *South African Journal of Botany*, *105*, 123-132.

Whitecross, M.A., Witkowski, E.T.F. and Archibald, S., 2017. Savanna tree-grass interactions: A phenological investigation of green-up in relation to water availability over three seasons. *South African Journal of Botany*, *108*, 29-40.

Wickham, H., 2011. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180-185.

Widener, M.J. and Li, W., 2014. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography*, *54*, 189-197.

Wilkins, E.J., Wood, S.A., and Smith, J.W., 2020. Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. *Environmental Management*, DOI: 10.1007/s00267-020-01373-7.

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M. and Fortson, L., 2019. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, *10*(1), 80-91.

Willis, C.K. and Woodhall, S.E., 2010. *Butterflies of South Africa's National Botanical Gardens*. South African National Biodiversity Institute. Pretoria.

Willis, C.G., Law, E., Williams, A.C., Franzone, B.F., Bernardos, R., Bruno, L., Hopkins, C., Schorn, C., Weber, E., Park, D.S. and Davis, C.C., 2017. CrowdCurio: an online crowdsourcing platform to facilitate climate change studies using herbarium specimens. *New Phytologist*, *215*(1), 479-488.

Wilson, J.J., Jisming-See, S.W., Brandon-Mong, G.J., Lim, A.H., Lim, V.C., Lee, P.S., and Sing, K.W., 2015. Citizen science: the first Peninsular Malaysia butterfly count. *Biodiversity Data Journal*, *3,* DOI: 10.3897/BDJ.3.e7159.

Wolter, K. and Timlin, M.S., 2011. El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI. ext). *International Journal of Climatology*, *31*(7), 1074-1087.

Woodhall, S., 2012. *Field Guide to Butterflies of South Africa*. Penguin Random House, South Africa.

Woodhall, S., 2020. *Field guide to butterflies of South Africa*. Bloomsbury Publishing, London.

Wright, D.R., Underhill, L.G., Keene, M., and Knight, A.T., 2015. Understanding the motivations and satisfactions of volunteers to improve the effectiveness of citizen science programs. *Society & Natural Resources*, *28*(9), 1013-1029.

Xerces Society for Invertebrate Conservation, 2017. Step-by-Step Western Monarch Thanksgiving Count Monitoring Guide 2017. Available from https://www.westernmonarchcount.org/wp-content/uploads/2014/10/Step-by-Step-Update-10.20.17.pdf, accessed 9 October 2020.

Zagheni, E., Garimella, V.R.K., Weber, I. and State, B., 2014, April. Inferring international and internal migration patterns from twitter data. In *Proceedings of the 23rd international conference on World Wide Web*, pp. 439-444.

Zipper, S.C., 2017. Agricultural research using social media data. *Agronomy Journal*, *110*(1), 349-358.

Zhao, M., Peng, C., Xiang, W., Deng, X., Tian, D., Zhou, X., Yu, G., He, H. and Zhao, Z., 2013. Plant phenological modeling and its application in global climate change research: overview and future challenges. *Environmental Reviews*, *21*(1), 1-14.

Zhao, X., Hu, H., Shen, H., Zhou, D., Zhou, L., Myneni, R.B. and Fang, J., 2015. Satellite-indicated long-term vegetation changes and their drivers on the Mongolian Plateau. *Landscape Ecology*, *30*(9), 1599-1611.

Appendices:

Appendix A: Ethics Clearance

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

SCHOOL OF SOCIAL SCIENCES ETHICS COMMITTEE
CONSTITUTED UNDER THE UNIVERSITY HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)

CLEARANCE CERTIFICATE	PROTOCOL NUMBER: INTR2020/07/01
PROJECT TITLE	Chatter About the Clouds of White: Using Social Media to Track the Annual Migration of Belenois aurota
INVESTIGATOR	Subhashini Pillay
SCHOOL/DEPARTMENT OF INVESTIGATOR	SOSS (e-Science)
DATE CONSIDERED	27 July 2020
DECISION OF THE COMMITTEE	Approved unconditionally
RISK LEVEL	MINIMAL RISK
EXPIRY DATE	Date of submission of the project Research Report

ISSUE DATE OF CERTIFICATE 6 August 2020 CHAIRPERSON (Prof. Rod Alence)

cc: Co-supervisors: Jennifer Fitchett, Helen Robertson

DECLARATION OF INVESTIGATOR

To be completed in duplicate and ONE COPY returned to the Chairperson of the School/Department ethics committee.

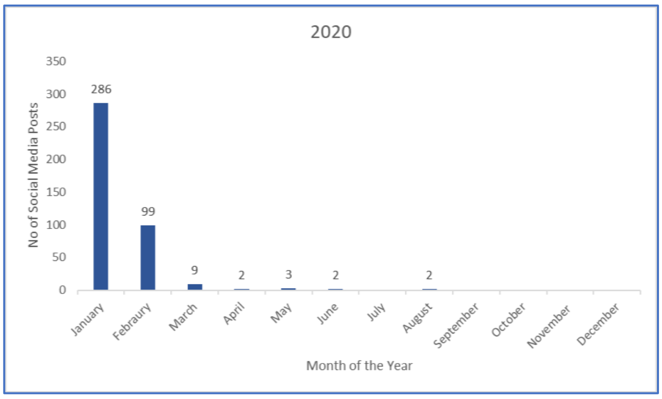
I fully understand the conditions under which I am are authorized to carry out the abovementioned research and I guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

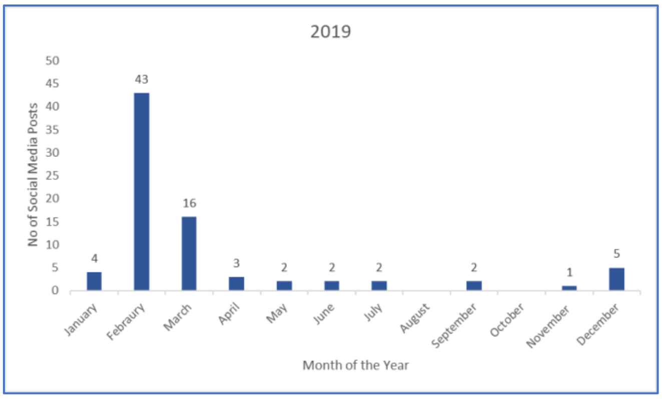
______________________ Signature

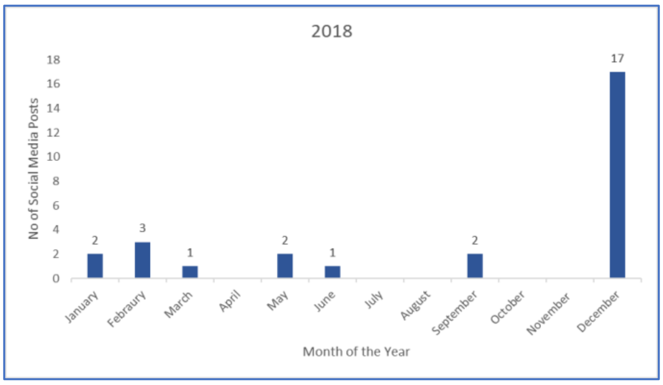
Date ______/______/______

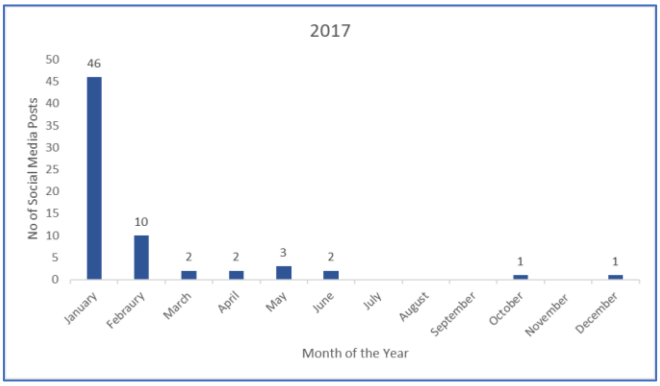
PLEASE QUOTE THE PROTOCOL NUMBER ON ALL ENQUIRIES

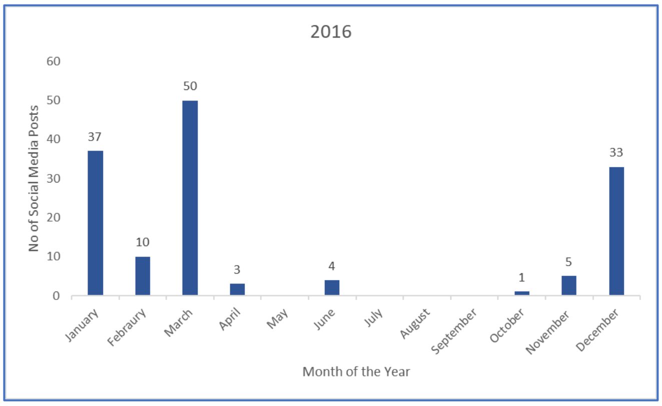
Appendix B: Bar graphs depicting the number of social media posts per month for the years 1914-2020

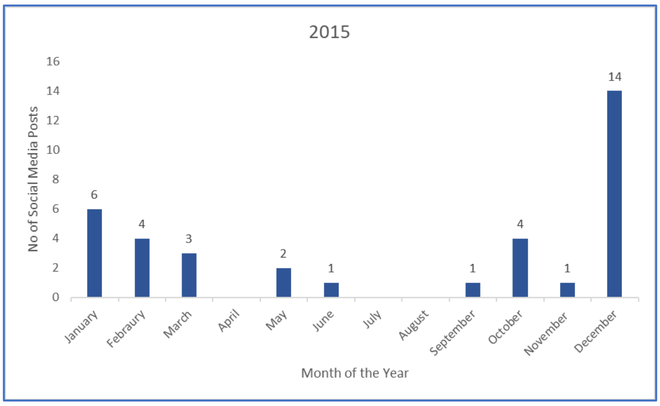


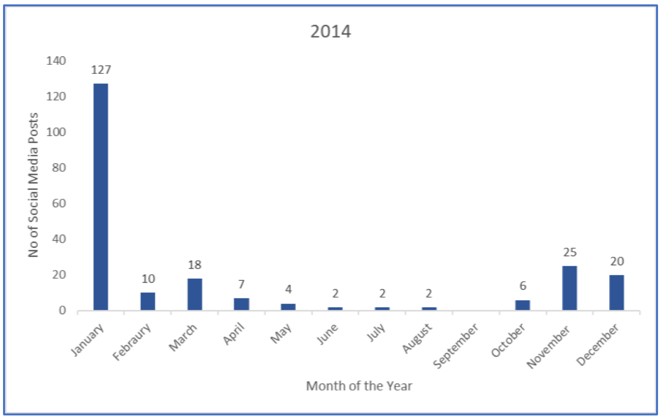


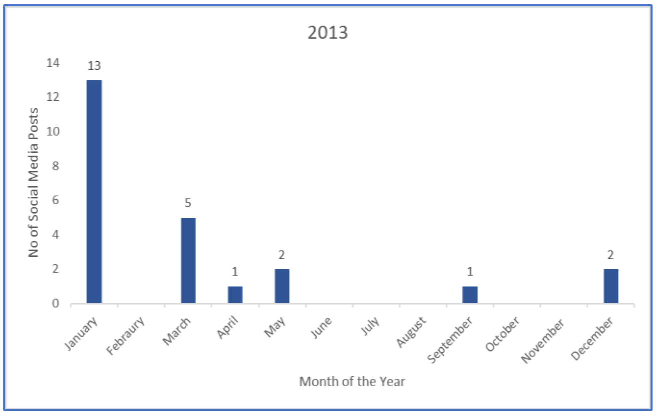


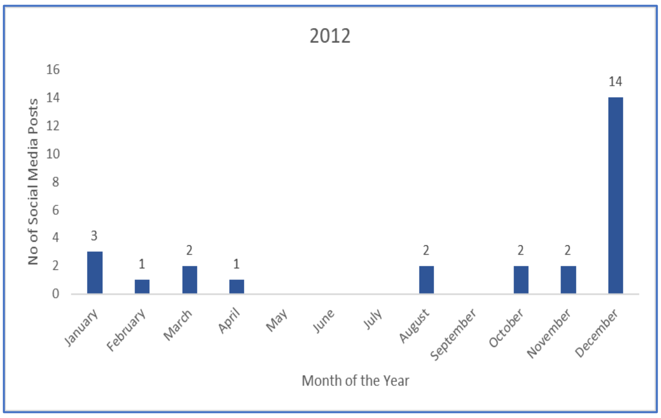


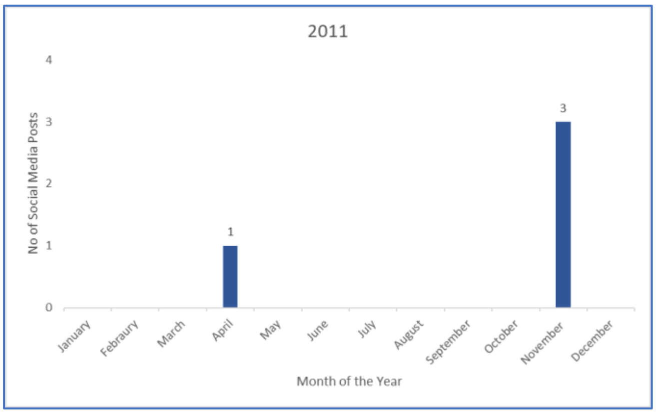


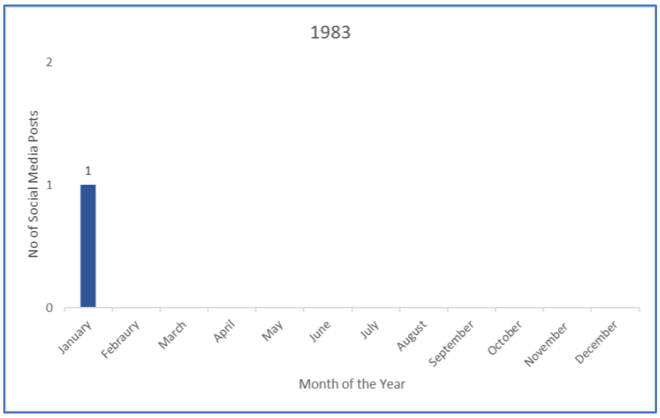


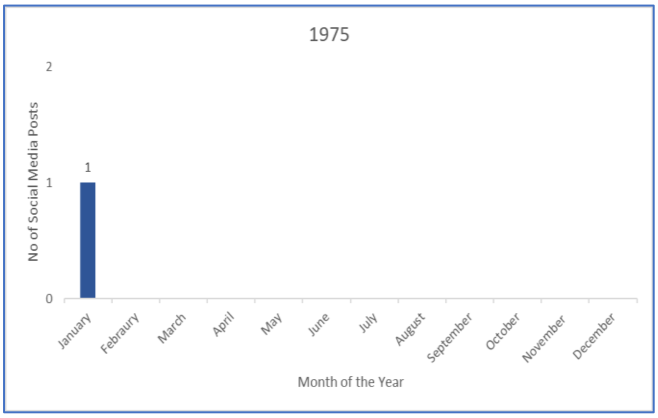

2014
No of Social Media Posts
140
120
100
80
60
40
20
0
127
10
18
7
4
2
2
2
6
25
20
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

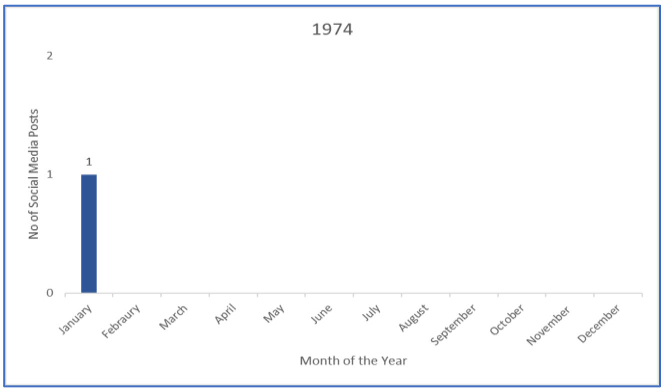

2013
No of Social Media Posts
14
12
10
8
6
4
2
0
13
5
1
2
1
2
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

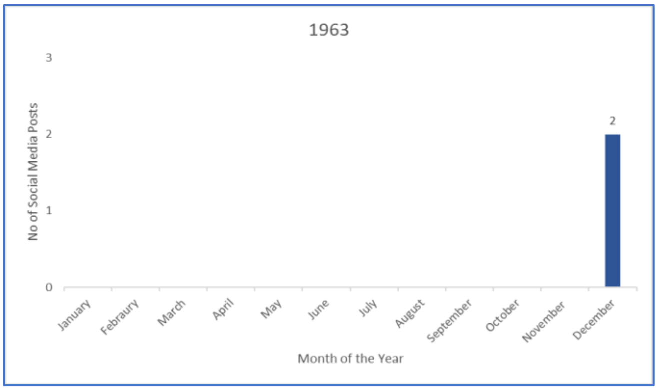

2012
No of Social Media Posts
16
14
12
10
8
6
4
2
0
3
1
2
1
2
2
2
14
January
February
March
April
May
June
July
August
September
October
November
December
Month of the Year

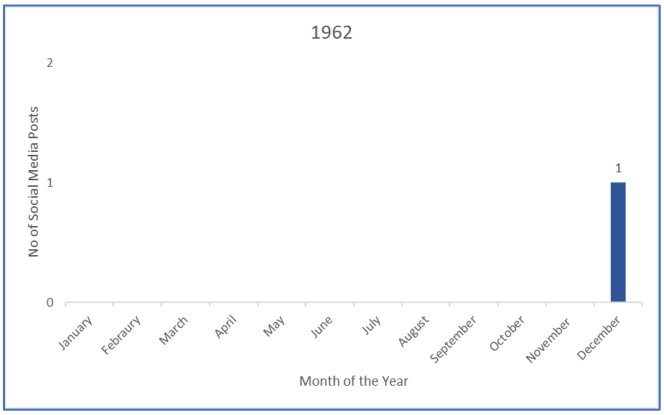

2011
No of Social Media Posts
4
3
2
1
0
1
3
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

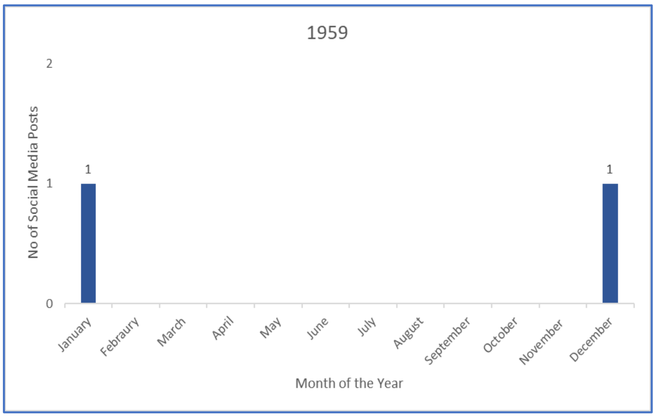

1983
No of Social Media Posts
2
1
0
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

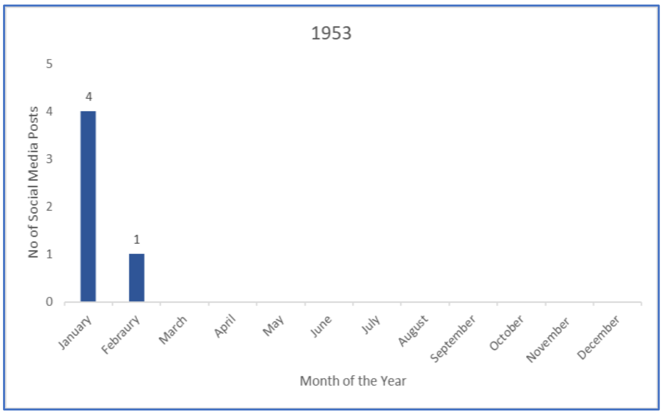

1975
No of Social Media Posts
2
1
0
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

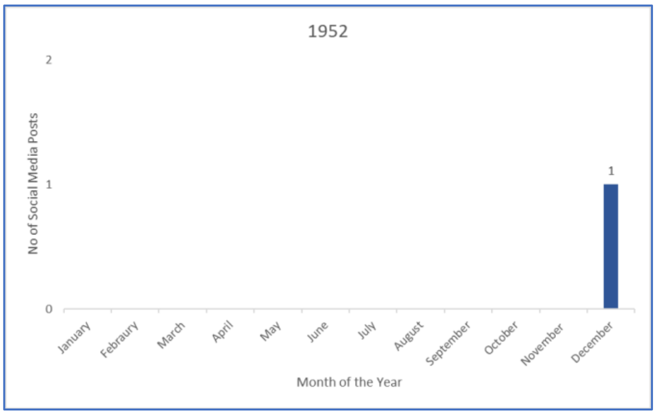

1974
No of Social Media Posts
2
1
0
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

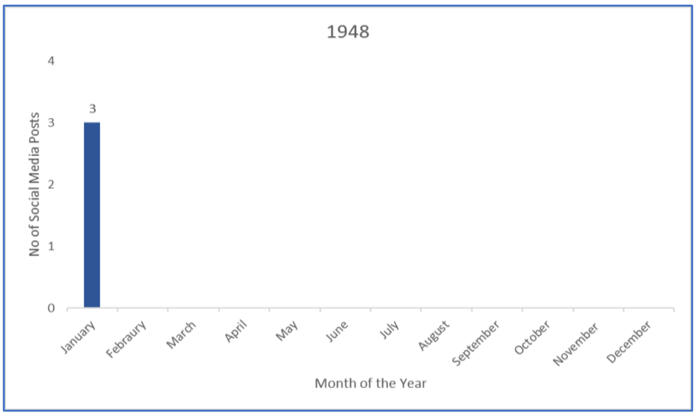

1963
No of Social Media Posts
3
2
1
0
2
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

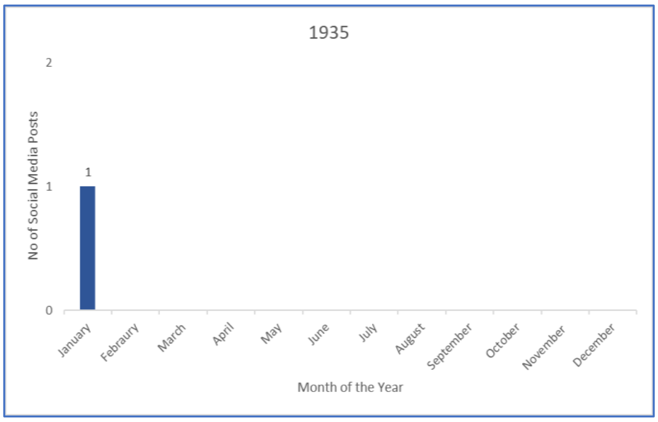

1962
No of Social Media Posts
2
1
0
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

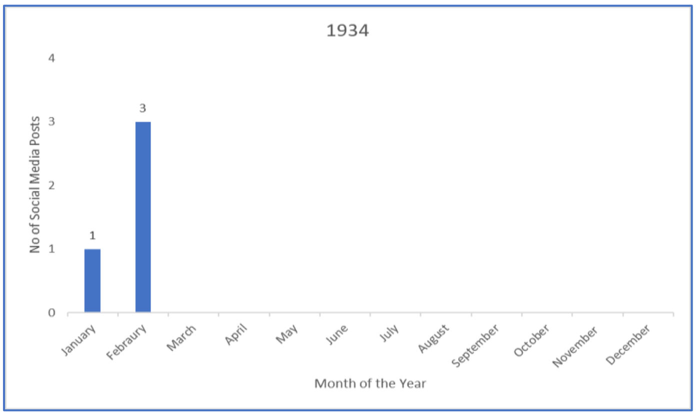

1959
No of Social Media Posts
2
1
0
1
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

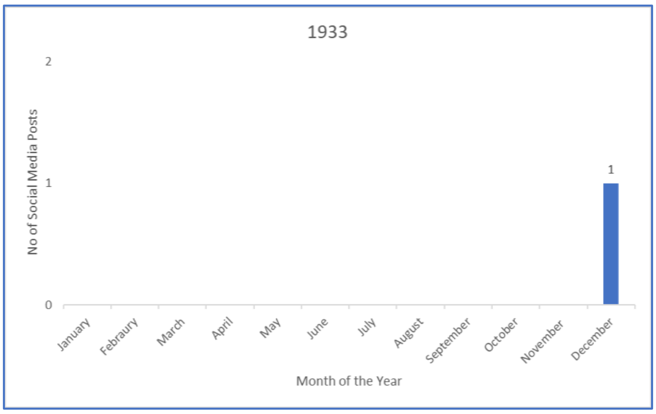

1953
No of Social Media Posts
5
4
3
2
1
0
4
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

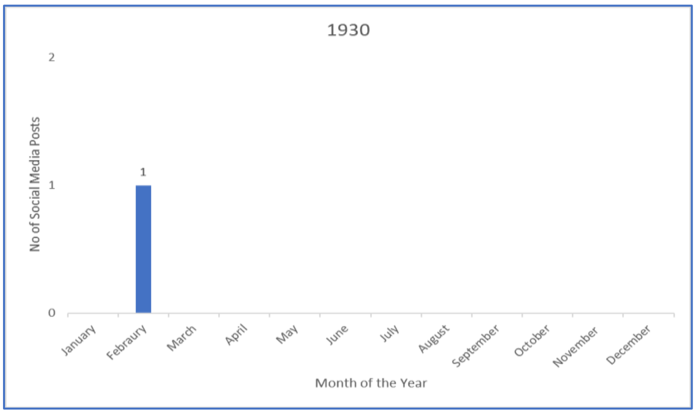

1952
No of Social Media Posts
2
1
0
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

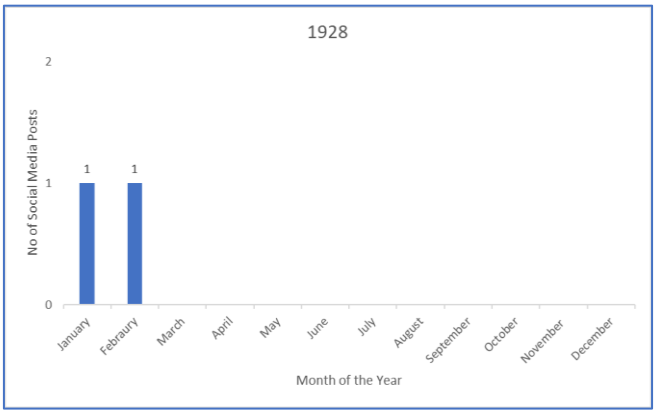

1948
No of Social Media Posts
Month of the Year
3
January
Febraury
March
April
May
June
July
August
September
October
November
December

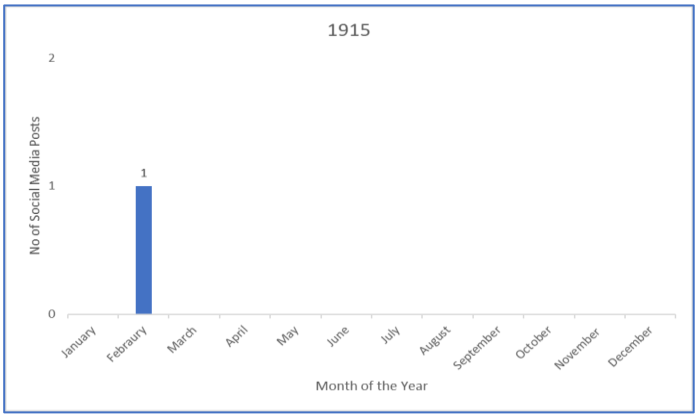

1935
No of Social Media Posts
Month of the Year
1
January
Febraury
March
April
May
June
July
August
September
October
November
December

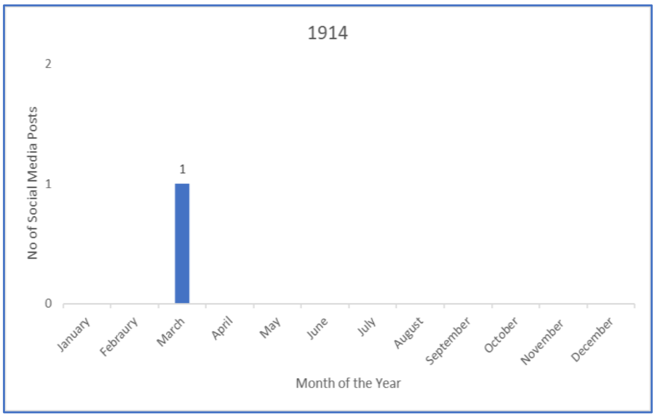

1934
No of Social Media Posts
Month of the Year
1
3
January
Febraury
March
April
May
June
July
August
September
October
November
December


1933
No of Social Media Posts
Month of the Year
1
January
Febraury
March
April
May
June
July
August
September
October
November
December

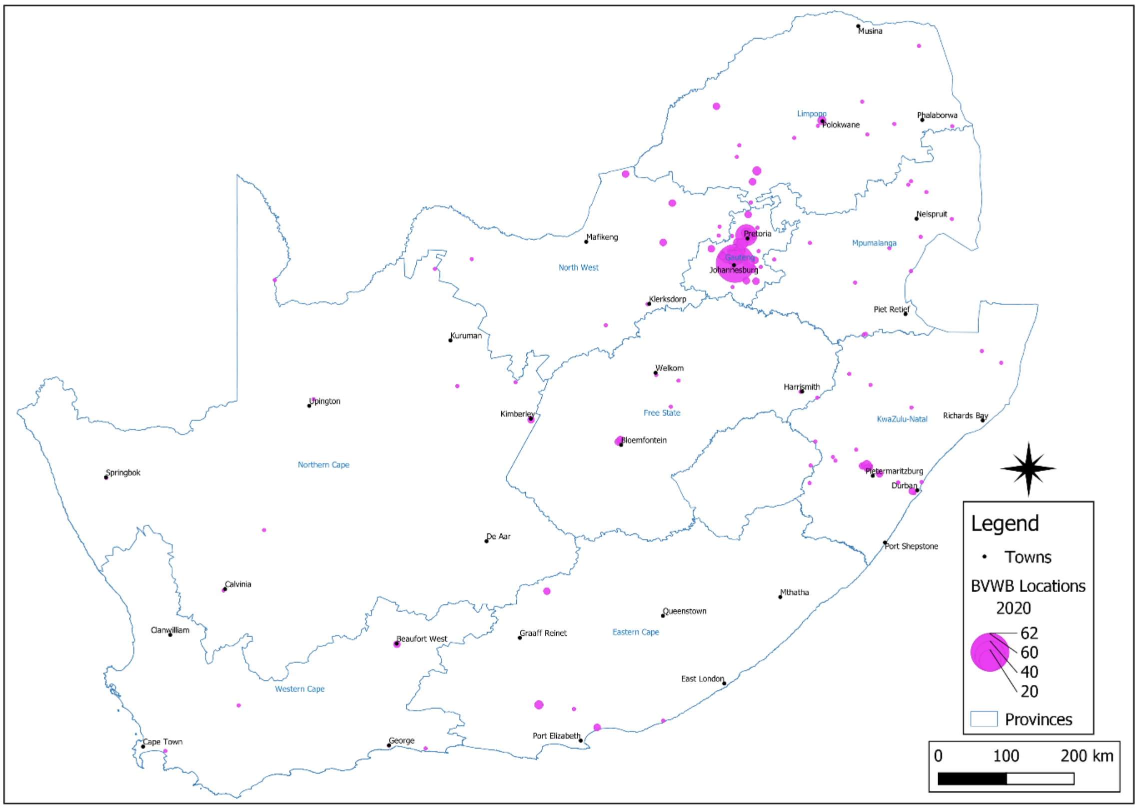

1930
No of Social Media Posts
Month of the Year
1
January
Febraury
March
April
May
June
July
August
September
October
November
December


1928
No of Social Media Posts
Month of the Year
1
1
January
Febraury
March
April
May
June
July
August
September
October
November
December

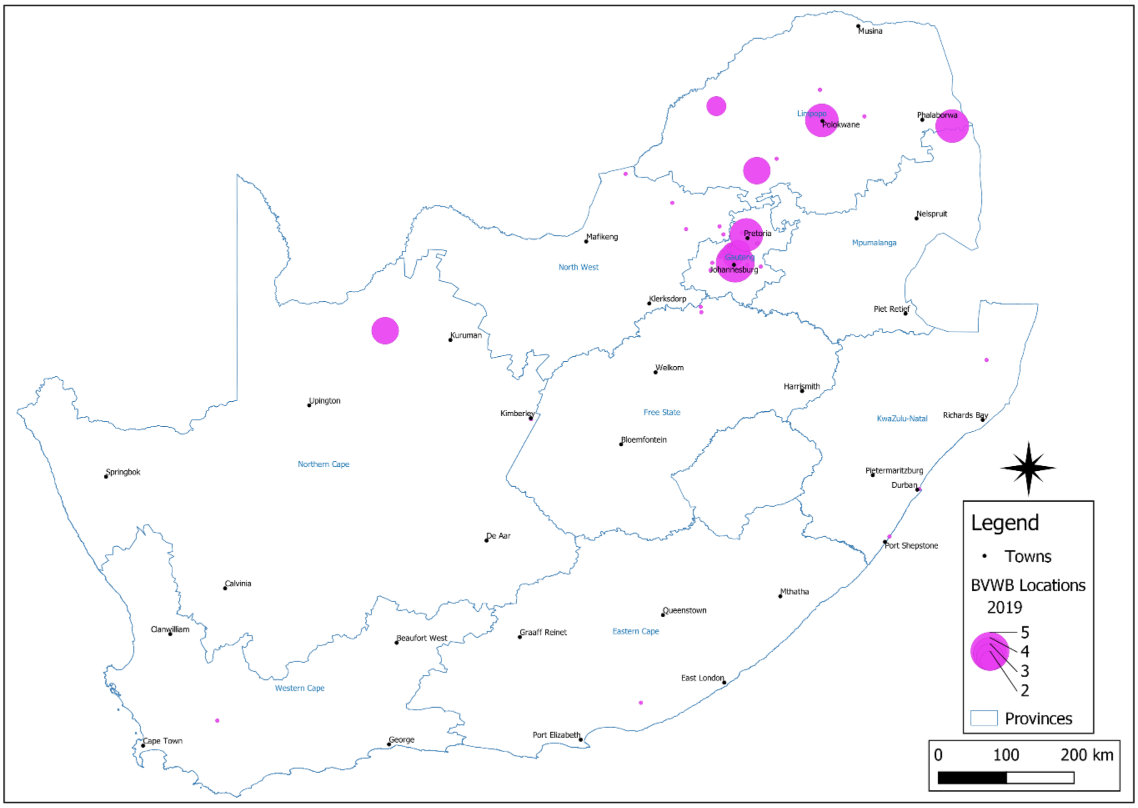

1915
No of Social Media Posts
2
1
0
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year


1914
No of Social Media Posts
2
1
0
1
January
Febraury
March
April
May
June
July
August
September
October
November
December
Month of the Year

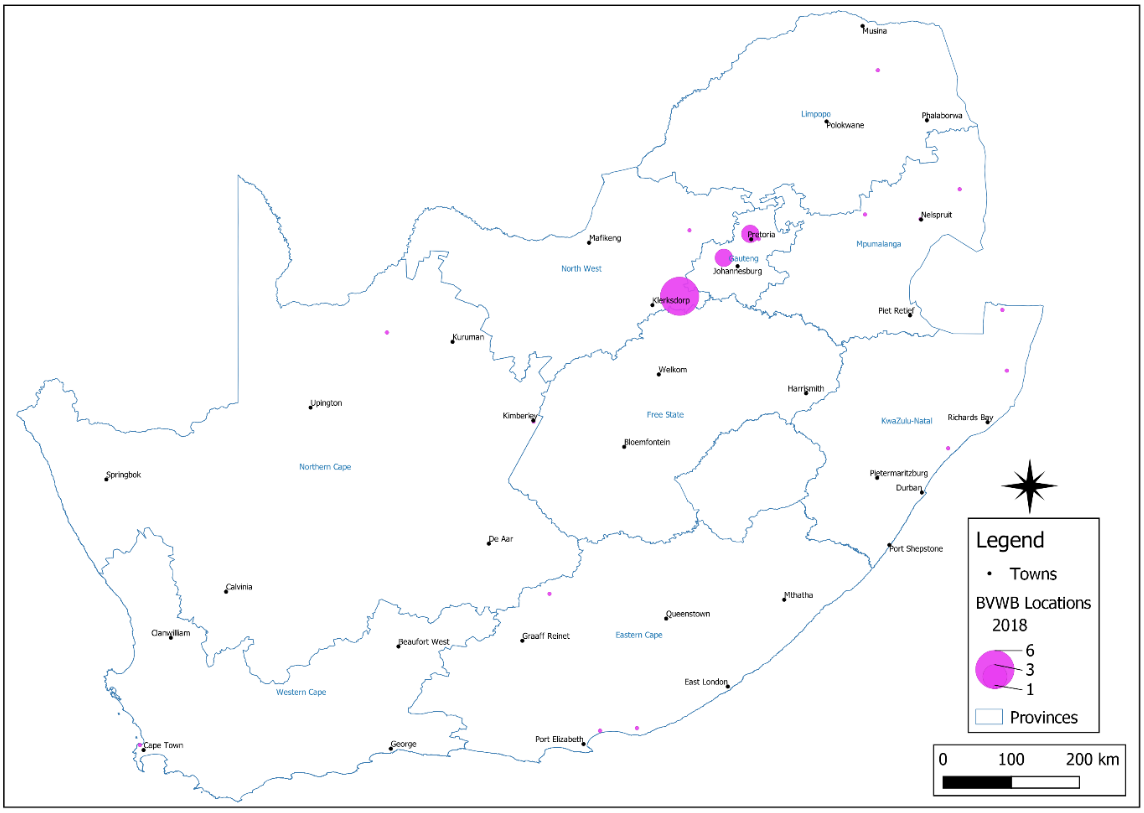
Appendix C: Spatial observations of *Belenois aurota* for period 2011-2020



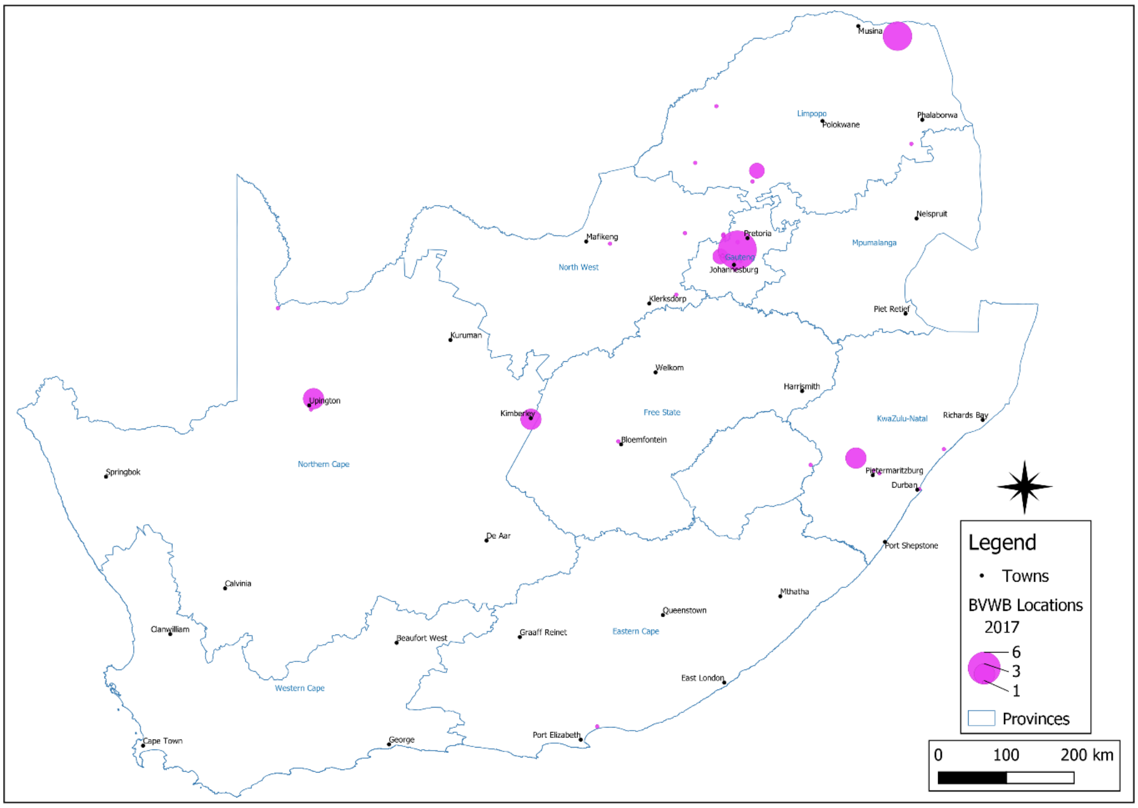
Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2020.



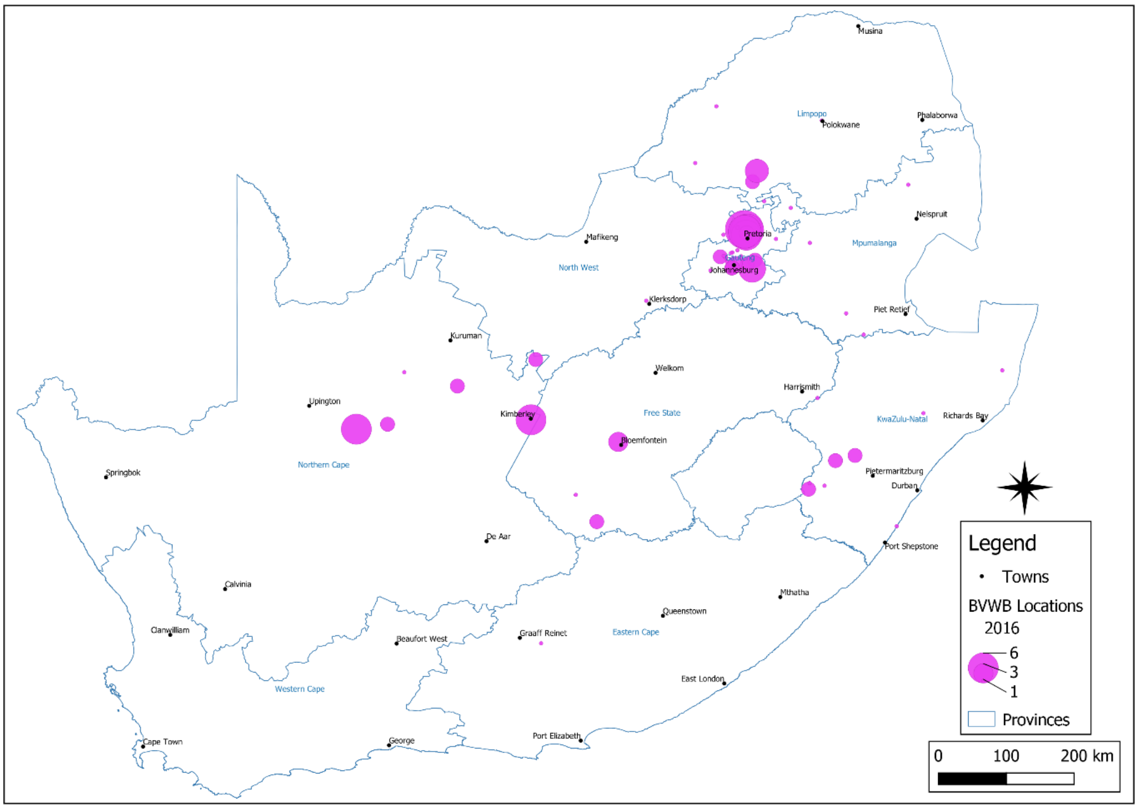
Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2019.



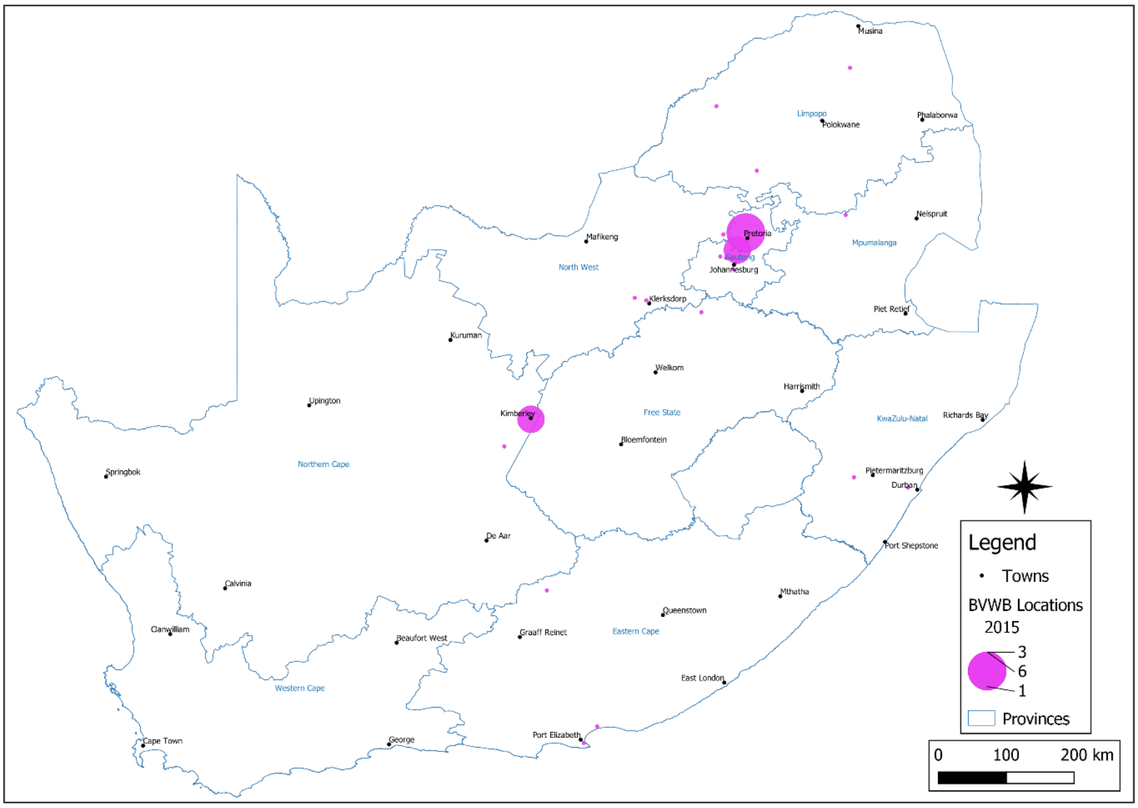
Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2018.



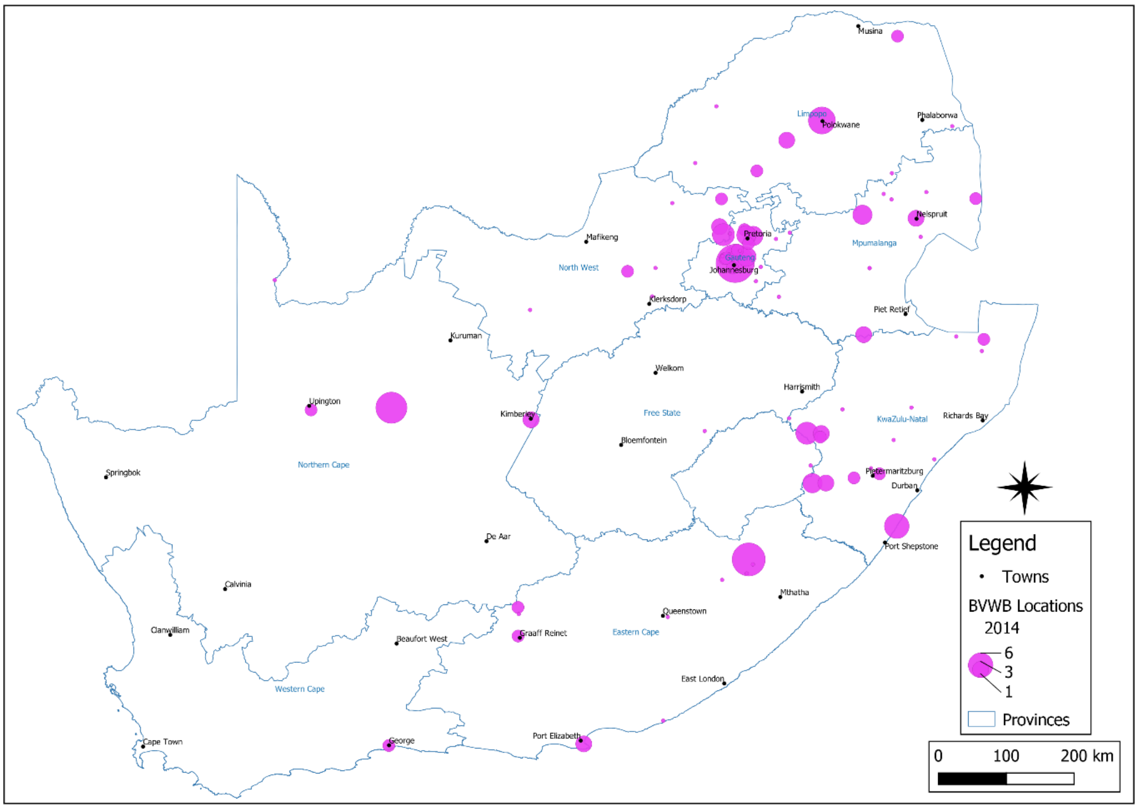
Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2017.



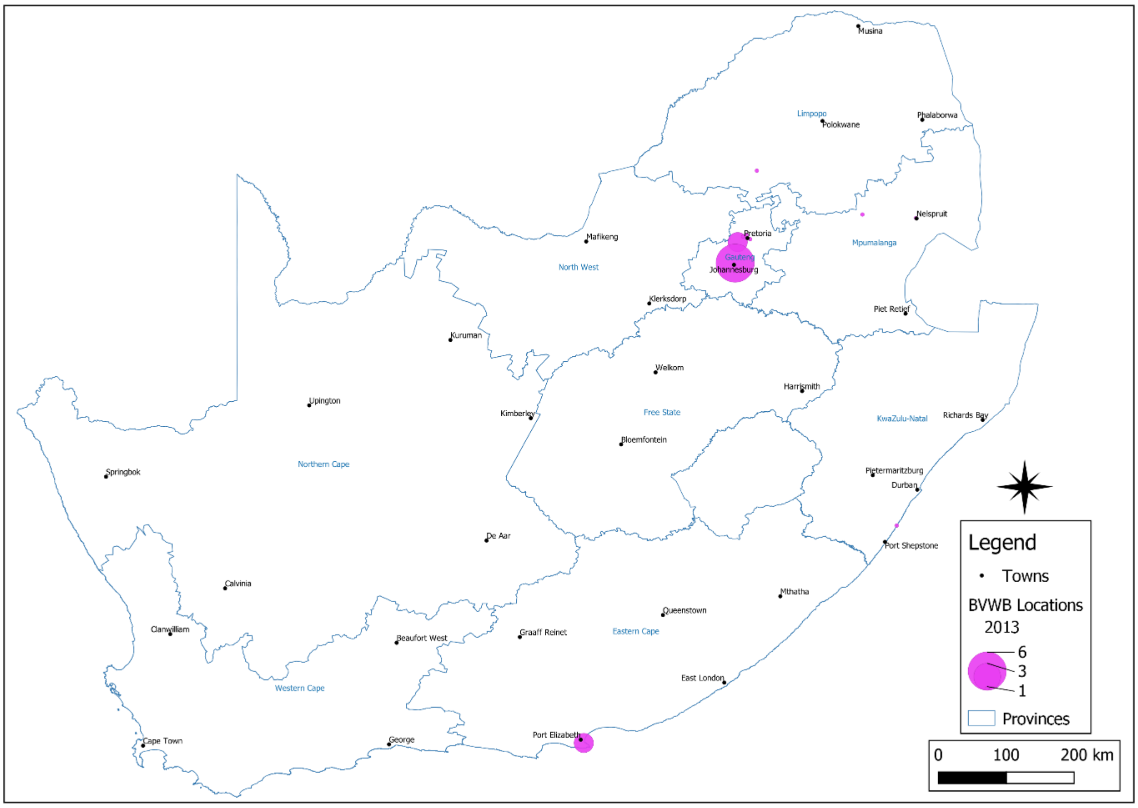
Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2016.



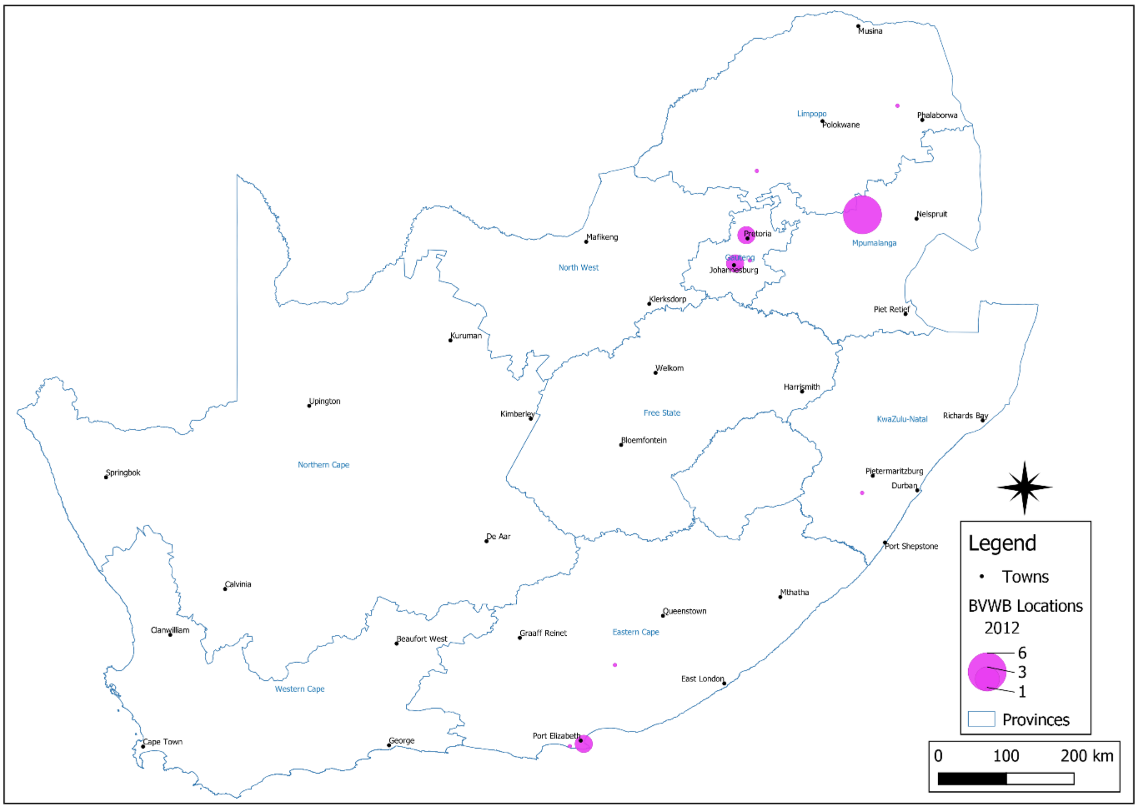
Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2015.



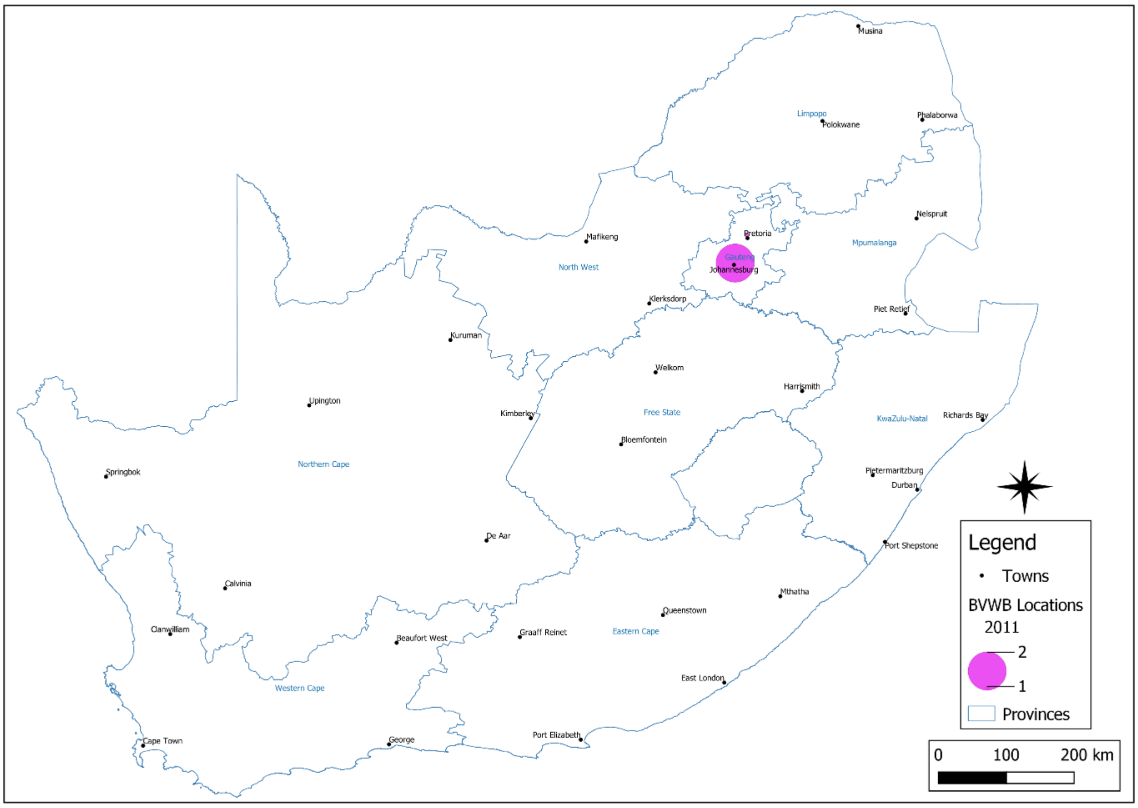
Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2014.



Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2013.



Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2012.



Map of locations from which recorded sightings of *Belenois aurota* were reported on social media for 2011.