HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Master’s thesis in Geography

Geoinformatics

Evaluating the use of crowdsourced data, trail camera images, and ranger-based knowledge in conservation geography: a case study in Taita Hills Wildlife Sanctuary, Kenya

Tuomas Kirjavainen

2025

Supervisors:
Enrico Di Minin
Miquel Torrents-Ticó

Master’s Programme in Geography

Faculty of Science

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Faculty	Department
Faculty of Science	Department of Geosciences and Geography

Author

Tuomas Kirjavainen

Title of thesis

Evaluating the use of crowdsourced data, trail camera images, and ranger-based knowledge in conservation geography: a case study in Taita Hills Wildlife Sanctuary, Kenya

Programme and study track

Master's programme in geography, Geoinformatics

Level of the thesis	Date	Number of pages
Master's thesis, 30 credits	January 2025	85 pages

Abstract

East African savannas are vital ecosystems harbouring biodiversity and high animal densities. The drastic decline in wildlife populations, particularly in Kenya, has raised concerns about the effectiveness of current conservation strategies. Protected areas play a crucial role in biodiversity conservation, but their success depends on effective conservation planning, supported by robust biodiversity data. However, the availability of high-quality biodiversity data remains a significant challenge. Crowdsourced data, with its broad spatial and temporal coverage, provides a potential tool for studying species distributions and seasonal patterns in poorly monitored regions. This study investigates the potential of crowdsourced biodiversity data from Flickr and the Global Biodiversity Information Facility (GBIF) to predict species occurrence and richness in the Taita Hills Wildlife Sanctuary in southeastern Kenya. The study aims to assess the degree of consistency between these data sources compared to ranger-based knowledge and trail cameras for mapping the distribution of the Big Four species, other mammals, and birds, while acknowledging that each data source offers distinct strengths and limitations.

Results showed a significant correlation between Flickr data and ranger-based knowledge for the Big Four species (Spearman's rho = 0.635, $p < 0.05$), highlighting the potential of social media as a tool for monitoring charismatic megafauna. GBIF data proved more reliable for mammals, showing strong correlations with both ranger-based knowledge (Spearman's rho = 0.674, $p < 0.05$) and trail camera data (Spearman's rho = 0.790, $p < 0.01$). The trail cameras, despite a limited deployment period of three months, provided valuable ground-truth data and broader spatial coverage in less accessible areas of the sanctuary. Additionally, they detected nocturnal and elusive mammal species that were underrepresented in the crowdsourced datasets. However, all three data sources struggled to accurately predict bird species richness, exhibiting weak correlations with ranger-based knowledge (Spearman's rho = 0.226 for Flickr and 0.209 for GBIF, both $p > 0.05$).

The study also highlights critical limitations, such as taxonomic bias and spatial clustering in crowdsourced data, which affect their reliability in fine-scale conservation planning. Despite these challenges, our findings demonstrate that integrating crowdsourced data, camera trap images, and ranger-based knowledge provides a cost-effective and complementary approach to wildlife monitoring. This research underscores the potential of combining innovative and traditional data collection strategies to enhance spatial conservation planning, ultimately supporting more informed and effective conservation actions in Kenya's rapidly changing landscapes.

Keywords

crowdsourced biodiversity data, wildlife conservation, spatial conservation planning, Digital Accessible Knowledge (DAK), GBIF, Flickr, Kenya, camera trapping, species occurrence, species richness, ecotourism

Where deposited

University of Helsinki electronic theses library E-thesis/HELDA

Additional information

Table of Contents

1. Introduction

Human activities have dramatically accelerated environmental changes leading to unprecedented declines in biodiversity and ecosystem services (Díaz et al., 2019). The growing demand for resources like food, energy, and timber has driven the widespread exploitation of natural ecosystems, contributing to their degradation (Díaz et al., 2019). As noted by Jaureguiberry et al. (2022), this overexploitation has caused most ecological health indicators to show concerning downward trends. Current research suggests that the rate of species loss is now between 100 and 10 000 times higher than natural background extinction rates (De Vos et al., 2015). If current trends persist, the loss of biodiversity and the degradation of ecosystems are anticipated to worsen in the coming decades, reinforcing the view that the ongoing Holocene extinction represents the planet's sixth mass extinction event (Díaz et al., 2019).

These alarming global patterns are also evident at regional scales (Ogutu et al., 2016). Among the regions profoundly affected are the East African savannas, renowned for their rich biodiversity and abundant wildlife (Lehmann et al., 2011). In recent decades, wildlife populations have decreased in size in East African protected areas, highlighting the need for more efficient conservation methods (Caro, T., 2008; Western et al., 2009; Craigie et al., 2010). Large-bodied mammal populations have been particularly affected and losses as high as 72 - 88% were recorded in Kenia between 1977 and 2016 (Ogutu et al., 2016). Main drivers of the decline include land cover conversion resulting in the loss of suitable habitats, increase in human population, poaching, climate change as well as general poor governance of the wilderness areas (Ogutu and Owen-Smith, 2003; Owen-Smith and Mills, 2006; Msoffe et al., 2011; Scholte, 2011; Ogutu et al., 2012).

Protected areas are the cornerstone of biodiversity conservation (Craigie et al., 2010). As the amount of biodiversity and land area covering natural indigenous vegetation and ecosystems have decreased across sub-Saharan Africa, it is vitally important to recognize areas of conservation importance (Brink et al., 2009; Western et al, 2009). While establishing protected areas is a crucial step, their effectiveness depends on various factors, including financial resources, adequate staffing as well as effective conservation planning informed by comprehensive biodiversity data (Margules and Pressey, 2000; Knight, Cowling and Campbell, 2006; Chen et al., 2022). Acquiring data of biodiversity features involves gathering comprehensive information on species distributions, habitat types, ecosystem functions, and environmental variables across the study area (Kujala et al., 2018). Margules and Pressey (2000) indeed argued that collecting rigorous biodiversity data is the first, and often the most difficult, step of a systematic conservation planning framework. Furthermore, as emphasized by Kukkala and Moilanen (2013), the acquisition of biodiversity data serves an essential role in monitoring the effectiveness of conservation actions within established protected areas. This underscores the pivotal significance of comprehensive biodiversity data collection within the framework of systematic conservation planning.

However, acquiring high-quality biodiversity data can be challenging due to limitations in resources, expertise, and data availability (Margules and Pressey, 2000; Troudet et al., 2017; Di Cecco et al., 2021; Tulloch et al., 2015). Efforts to overcome these challenges include integrating local ranger-based knowledge (Brackowski et al., 2024) and trail camera data (O'Brien et al., 2020) as well as using crowdsourced data sources such as Digital Accessible Knowledge (DAK) and social media platforms (Tulloch et al., 2016; Di Cecco et al., 2021). Ranger-based knowledge holds great potential for enhancing wildlife monitoring and conservation efforts, offering an effective and cost-efficient data collection method, particularly in regions with limited resources (Brackowski et al., 2024). Braczkowski et al. (2024) argue that where wildlife tourism rangers are present, they could be a powerful addition to future wildlife census efforts across Africa. Similarly, DAK databases and social media offer researchers abundant biodiversity data to complement traditional research data with a wide spatial and temporal range (Toivonen et al, 2019; Di Cecco et al, 2021). Indeed, crowdsourced geotagged observation records can be used to source spatial biodiversity data such as information on species occurrences and richness (Farooq et al., 2021). However, no previous studies have examined the predictive accuracy of crowdsourced data for wildlife distributions or compared these data with ranger-based knowledge to evaluate the effectiveness of different monitoring methods in Southeast Kenya.

Aiming to close this research gap, this study evaluates how crowdsourced biodiversity data can be used in a conservation area in Southeast Kenya to predict wildlife occurrence and species richness at a regional scale. By comparing crowdsourced data with ranger-based knowledge and trail camera observations, we aim to identify the most effective data sources for modelling distributions of different taxonomic groups. Specifically, we assess the predictive accuracy of crowdsourced information within a wildlife sanctuary setting, along with the relationships and level of agreement between each dataset. The outcomes offered here aim to inform spatial conservation planning in Kenya by providing insights into the effectiveness of integrating crowdsourced biodiversity data with ranger expertise and traditional monitoring methods. Ultimately, our study contributes to the development of more efficient and cost-effective strategies for wildlife monitoring and conservation planning in regions with limited resources.

The following research questions were addressed:

1. How accurately do crowdsourced species occurrence data correlate and agree with ranger-based knowledge and trail camera data?
2. What are the strengths and limitations of using crowdsourced data for mapping species occurrence and richness?
3. How effective is short-term trail camera deployment for mammal distribution mapping following the rainy season?

2. Background

2.1. Kenyan wildlife populations on decline

Kenya is renowned for its iconic wildlife and picturesque savanna landscapes, attracting safari-based tourism from all over the world (Ministry of Tourism and Wildlife, Kenya, 2022). With 411 protected areas covering 72,545 km², Kenya has protected more than 12% of the nation's land area (World Bank, 2021; IUCN, 2020). These protected areas, which include national parks, game reserves, and sanctuaries, are predominantly situated in arid and semi-arid regions (Ogutu et al., 2016). Savanna ecosystems dominate these areas and are globally recognized for their rich biodiversity, supporting a range of iconic megafauna, including the Big Five: lion (*Panthera leo*), elephant (*Loxodonta africana*), buffalo (*Syncerus caffer*), leopard (*Panthera pardus*), and rhinoceros (*Diceros bicornis*) (Du Toit and Cumming, 1999; Bond and Loffel, 2001).

Wildlife populations in the savannas of Kenya have however experienced significant declines, raising alarm about biodiversity conservation and the effectiveness of current conservation strategies. According to Ogutu et al. (2016), between 1977 and 2016, wildlife numbers in Kenya's rangelands decreased by an average of 68%. The study highlights that this decline is pervasive across various species, with some experiencing losses as high as 72-88%. Large-bodied mammals such as lesser kudu (*Tragelaphus imberbis*), greater kudu (*Tragelaphus strepsiceros*), Thomson's gazelle (*Eudorcas thomsonii*), eland (*Taurotragus oryx*), East African oryx (*Oryx beisa*), hartebeest (*Alcelaphus buselaphus*), impala (*Aepyceros melampus*), Grevy's zebra (*Equus grevyi*), and waterbuck (*Kobus ellipsiprymnus*) have been particularly affected (Ogutu et al., 2016). However, these declines are not confined to unprotected areas; they are also occurring within protected regions, highlighting the widespread nature of the issue (Caro and Scholte, 2007; Western et al., 2009).

The largest national parks in Kenya, Tsavo (East and West), Meru and Maasai Mara, have experienced the most severe declines in wildlife numbers (Ottichilo et al., 2000; Western et al., 2009; Ogutu et al., 2011). The most drastic decline has been recorded in Meru national park, with a disturbing 78% decline in wildlife numbers between 1977 and 2007 (Western et al., 2009). Similarly, the Maasai Mara national park experienced significant wildlife losses. Ogutu et al. (2011) described a 67% decline of wildlife populations in the Maasai Mara between 1977 and 2009, with many large-bodied mammal species such as wildebeest and zebras facing sharp declines. Wildebeest, especially, were subject to huge losses, with Ogutu et al. (2016) noting a decline of 81%. Correspondingly, remote sensed data shows that wildlife populations in the Tsavo national park declined by 63% between 1977 and 1997, with large herbivores such as elephants and buffalo being particularly affected (De Leeuw et al., 1998; Western et al., 2009). Newer research in Tsavo national park (Ogutu et al. 2016) confirms that this decline has not only persisted but has accelerated, with some ungulate species experiencing losses as high as 63-89%. Additionally, elephant populations in Tsavo have been especially hard hit. In the early 1970s, Tsavo had one of the largest elephant populations in Africa, with over 35,000 individuals (De Leeuw et al., 1998). However, by the late 1980s, poaching had reduced the population to approximately 6,000 elephants (De Leeuw et al., 1998). Although anti-poaching measures have led to recovery, the current elephant population is still far below historical levels, with around 13,000 elephants recorded in a survey done by Ngene et al. (2013). Similarly, black rhinoceros populations in Tsavo were decimated by poaching, and the species is now classified as critically endangered (Western et al., 2009). Ongoing efforts to reintroduce and protect rhinos in Tsavo have been met with some success, but the population remains small and vulnerable (Western et al., 2009). The severity of Kenya's biodiversity crisis is underscored by the 2013 classification of 7 mammal species as critically endangered, 19 as endangered, and 37 as vulnerable (Ogutu et al., 2016).

The causes behind these wildlife declines are complex and multifaceted (Ogutu et al., 2016). While poaching is often cited as a leading factor, it has been incorrectly assumed to be the dominant driver across all taxa (Western et al., 2009). Western et al. (2009) note that, with the exception of wildlife losses in Meru National Park, poaching is currently unlikely to be a major contributor in other parks, which are heavily patrolled by the Kenyan Wildlife Service. This is evidenced by the recovering populations of elephants and rhinos, the species most at risk of poaching (Western et al., 2009). However, Ogutu et al. (2016) highlight that poaching is still a significant problem in less-patrolled regions and outside protected areas. Furthermore, the Kenyan government’s ban on trophy hunting in the late 1970s rules out trophy hunting as a cause of recent wildlife declines (Ogutu et al., 2016).

Instead, one of the primary reasons for these declines is the significant limitations imposed by the design of protected areas in Kenya (Ottichilo et al., 2000; Western et al., 2009). Many of Kenya's national parks only cover the dry season ranges of migratory species, such as wildebeest and zebras, while the wet season ranges lie outside protected boundaries (Ottichilo et al., 2000). Consequently, the loss of migratory ranges due to agricultural expansion has been a significant contributor to wildlife declines, particularly in areas near Maasai Mara and Tsavo national parks (Western et al., 2009). Indeed, human population growth and its associated pressures are another major cause of wildlife declines (Caro and Scholte, 2007; Ogutu et al., 2017; Riggio et al., 2018). Kenya's population has grown nearly fivefold since 1960, resulting in increased demand for land for agriculture, settlements and infrastructure (Brink et al., 2014; Ogutu et al., 2016). The unregulated expansion of settlements and infrastructure, as well as overgrazing and uncontrolled wood harvesting, has led to habitat fragmentation, degradation and loss in Kenya's rangelands, where over 70% of the country's wildlife lives (Muriuki et al., 2011; Ogutu et al., 2016). Additionally, as Ogutu et al. (2016) note, the increasing livestock numbers introduce competition between livestock and wildlife for resources such as water and pasture, particularly during droughts.

Climate change has contributed to the decline in Kenya's wildlife populations (Ogutu and Owen-Smith, 2003; Ogutu et al., 2014). Kenya has been experiencing a reduction in rainfall, coupled with an increase in temperatures, which has both increased the frequency and the severity of droughts (Williams and Funk, 2011; Funk, 2012; Lyon and DeWitt, 2012). The increased spatial and temporal variation of rainfall has led to mass mortality events among wildlife (Wangai et al., 2013; Ogutu et al., 2014). Furthermore, the rising frequency of droughts forces wildlife to migrate and thus increases human-wildlife conflicts (Smith and Kasiki, 2000).

Addressing the decline in Kenyan wildlife numbers is crucial not only for biodiversity conservation but also for the country's economy, as wildlife tourism is a major contributor to Kenya's GDP (Ministry of Tourism and Wildlife, Kenya, 2022). According to the Kenyan Ministry of Tourism and Wildlife (2022), wildlife-based tourism, particularly safaris, generates yearly between USD 1.0 – 1.4 billion in inbound expenditure and accounts for 75% of the nation's leisure travel, highlighting the economic significance of the country's biodiversity. Declining wildlife populations pose a direct threat to the travel industry, as reduced wildlife numbers may deter tourists from visiting national parks and reserves (Otianga-Owiti et al., 2021). Otianga-Owiti et al. (2021) highlight that potential loss of tourists can in turn further reduce the funds available for wildlife conservation. Furthermore, the loss of biodiversity affects essential ecosystem services, such as water regulation and soil fertility, which support the livelihoods of local communities (Gudka, 2020).

To reverse wildlife losses in Kenya, comprehensive and robust conservation efforts are essential. Initiatives such as anti-poaching measures, habitat restoration, species translocation and reintroduction projects, as well as community-based conservation programs, have shown encouraging results (Western et al., 2009; Ogutu et al., 2016). However, to truly halt the decline in wildlife populations, particularly in key ecosystems like Tsavo, there is an urgent need for more effective, data-driven conservation strategies. This study aims to contribute to these efforts by investigating the most reliable methods and data sources for mapping wildlife distributions, which are critical for informed conservation planning and sustainable management in Kenya's rapidly evolving landscape. By integrating innovative approaches, this research hopes to offer practical solutions to support the conservation of the country's rich biodiversity.

2.2. Spatial conservation planning needs biodiversity data

The dynamic nature of conservation areas requires continued efforts to generate updated biodiversity knowledge data, ensuring the underlying data used in quantitative conservation geography are relevant and up to date (Di Minin et al., 2021b). Spatial conservation planning (SCP), a method in quantitative conservation geography, is a structured and data-driven conservation approach that aims to identify priority areas for biodiversity conservation while balancing ecological, social and economic factors (Margules and Pressey, 2000; Naidoo et al., 2006; Di Minin et al., 2013; Tulloch et al., 2015; Di Minin et al., 2021b). Given the limited resources available for conservation, it is crucial to prioritize areas where the greatest biodiversity can be protected with the least amount of effort (Naidoo et al., 2006). Margules and Pressey highlight that at the core of SCP is the need for accurate and comprehensive biodiversity data. Collecting biodiversity data requires obtaining detailed information on species distributions, habitat types, ecosystem functions, and environmental variables (Kujala et al., 2018). Such data form the foundation for making informed decisions about which areas should be prioritized for protection, restoration, or managed for conservation goals (Margules and Pressey, 2000). Biodiversity factors especially affecting the direction of conservation actions include endemism, endangered species, species richness and habitat quality (Arponen, 2012; Hernández-Quiroz et al., 2018). Furthermore, the collection of biodiversity data plays a crucial role in assessing the effectiveness of conservation efforts within designated protected areas (Kukkala and Moilanen, 2013). Without reliable biodiversity data, conservation efforts risk being inefficient, misplaced, or even counterproductive (Margules and Pressey, 2000; Kremen and Merenlender, 2018).

In Kenya, where rapid wildlife declines in key ecosystems like Tsavo are evident, biodiversity data is crucial for monitoring the population sizes, understanding the drivers of wildlife losses, and ultimately informing more effective conservation strategies (Western et al., 2009; Ogutu et al., 2016). As Western et al. (2009) argue, the design of many Kenyan national parks is flawed by their coverage of only dry season ranges of migratory herbivores, resulting in ineffective conservation management. Ogutu et al. (2016) further emphasize that the poor general governance driven partly by low amounts of comprehensive data account for the drastic declines in Kenyan wildlife numbers. This raises critical questions about the effectiveness and management of Kenya's national parks, which are unable to fully support wildlife populations throughout the year. To address these issues, comprehensive biodiversity data is needed to understand the spatial and temporal dynamics of wildlife and habitat use (Onditi et al., 2021).

A significant challenge for SCP is the availability and accessibility of comprehensive biodiversity data (Troudet et al., 2017). Moilanen (2012) and Joppa et al. (2016) highlight that the lack of accessible and high-quality data often becomes a key barrier to effective SCP implementation. Indeed, acquiring high-quality data on species and habitat distribution poses several challenges, including limitations in resources, expertise, and data availability and quality (Moilanen, 2012; Margules and Pressey, 2000; Troudet et al., 2017). Traditional methods of species distribution data collection include direct monitoring, such as line transect censuses and track surveys (Silveira et al., 2003) as well as remote sensing (Reddy et al., 2024). However, these methods have limitations mainly due to their labour-intensive nature and logistical problems (Lyra-Jorge et al., 2008; Palmer et al., 2018; Bruce et al., 2018).

Trail cameras have emerged as an effective tool for collecting wildlife data, particularly in areas where traditional methods like line transects and track surveys are impractical (Stein et al., 2008; Chapman and Balme, 2010; Brassine and Parker, 2015), or where certain habitat types necessitate their use (Tobler et al., 2008; Moolman et al., 2019). Trail cameras offer several advantages that contribute to their widespread use in wildlife research and conservation. Trail cameras are non-invasive, allowing for the continuous monitoring of wildlife without human presence, which minimizes disturbance and reduces observer bias (Silveira et al., 2003; Bruce et al., 2018). They are capable of operating day and night under various environmental conditions, making them especially useful for detecting elusive or nocturnal species that are difficult to observe through direct methods as highlighted by Brassine and Parker (2015). Additionally, trail cameras can collect large volumes of data over extended periods with relatively low effort and cost compared to traditional survey techniques (Rovero et al., 2013; O'Brien et al., 2020). Their ease of deployment in remote or rugged terrains further enhances their utility, allowing researchers to gather data in areas that are otherwise challenging to study (Stein et al, 2008; Assou et al., 2021).

Despite their advantages, trail cameras too require direct fieldwork for deployment and ongoing maintenance (Rovero et al., 2013; Braczkowski et al., 2016; Cordier et al., 2022). The high initial cost of trail cameras can also be an obstacle in regions where resources for research are limited (Silveira et al., 2003). Additionally, Braczkowski et al. (2024) note that for nocturnal species such as lions, most commercially available trail cameras are insufficient for individual identification, making them unsuitable for analytical models that depend on precise identification of individual animals. To further improve the efficiency of biodiversity data collection and to complement traditional data collection methods, recent efforts have focused on integrating local ranger-based knowledge (Braczkowski et al., 2024) and utilizing alternative data sources, such as crowdsourced datasets (Tulloch et al., 2016; Di Cecco et al., 2021). Braczkowski et al. (2024) demonstrated the value of ranger-collected data in monitoring lions in Uganda, showing that search-encounter protocols led by rangers produced more reliable results for population estimates compared to camera traps. Furthermore, ranger-led patrols were found to be 50% less expensive than maintaining camera traps in 32 locations, underscoring both the economic and practical benefits of integrating ranger expertise into conservation efforts (Braczkowski et al., 2024). These alternative methods, including the use of ranger-based knowledge and crowdsourced data, can be particularly valuable for SCP in regions with limited scientific infrastructure or where large-scale biodiversity assessments are logistically challenging (Ball-Damerow et al., 2019; Braczkowski et al., 2024).

2.3. Crowdsourced biodiversity data

2.3.1. Flickr as a source of biodiversity data

The use of digital and crowdsourced data sources has emerged as a worthy method for addressing biodiversity conservation challenges (Di Minin et al., 2015; Sosef et al., 2017; Toivonen et al., 2019). Citizen science platforms such as iNaturalist and eBird have long been recognized for their importance in providing crowdsourced observation-based biodiversity data with a vast spatial and temporal scale (Coetzer, 2012; Callcutt et al., 2018; Di Cecco et al., 2021). On the other hand, data from social media platforms have been successfully used to study human-nature interactions such as tourist preferences and visitor numbers in national parks (Willemen et al., 2015; Tenkanen et al., 2017; Wu et al., 2017). However, the usage of social media platforms as biodiversity information sources remains a fairly understudied field. Among the many social media platforms, Flickr has been recognized as a leading source of potential biodiversity information (Di Minin et al., 2015; Toivonen et al., 2019). Flickr is a popular photo-sharing platform that allows users to upload and share images with metadata, including geolocation (Li et al., 2013).

One of the limiting factors in the use of social media data for research is the platforms' Application Programming Interface (API) policies (Fink et al., 2021). Many platforms, such as Instagram and Facebook, have restrictive API regulations, which limit their accessibility for conservation studies (Toivonen et al., 2019; Fink et al., 2021). In contrast, Flickr's API is notably open to scientific use, making it one of the few social media platforms that allows researchers to access and analyse its data (Toivonen et al., 2019). This makes Flickr a potential resource for conservation studies, as it seems to provide a large repository of user-generated image content with spatial metadata. Interestingly, Hirvonen et al. (2020) found that in Sub-Saharan Africa, Flickr posts are often concentrated in national parks and other popular tourist sites. Kenya, in particular, was identified as having one of the highest densities of Flickr posts within national parks across Africa, underscoring Flickr's potential for studying species distributions in wildlife viewing areas (Hirvonen et al., 2020). Additionally, studies by Hausmann et al. (2017) and Toivonen et al. (2019) have found that, compared to other social media platforms like Instagram, Twitter, and Facebook, Flickr was the primary source used to share biodiversity-related pictures. In contrast, Instagram and Twitter were more used to share content showcasing human-nature interactions.

Geotagged pictures make Flickr particularly interesting for ecological research, as they allow researchers to extract spatial biodiversity data directly from user-contributed photos (Angus et al., 2010; Li et al., 2013). This crowdsourced information can provide data for studies focusing on species distribution, human-wildlife interactions as well as other conservation initiatives (Fink et al., 2020, Edwards et al., 2021). However, a literature review by Toivonen et al. (2019) found that only 11 publications used data from Flickr for conservation purposes. Although the number of studies utilizing Flickr data has increased since 2019, it remains low. Despite this, the limited studies that have employed Flickr data in conservation research demonstrate its potential value. For instance, Allain (2019) used Flickr to map the distribution of invasive freshwater turtles in the United Kingdom. The study found that the number of sightings recorded on Flickr was significantly higher than those recorded in the National Biodiversity Network (NBN) database, highlighting its potential as a valuable proxy for species distribution data. Similarly, Edwards et al. (2021) found that in the UK, the spatial distribution of geotagged Flickr posts, particularly for common invasive species and garden birds, is closely aligned with data from the NBN database. This study emphasized that Flickr could provide a rich source of observational data for specific taxonomic groups. However, the authors noted that Flickr data best reflected the NBA dataset when considering it from a purely spatial perspective, as Flickr was shown to lack the seasonal variation captured by the NBN database (Edwards et al., 2021).

Users posting wildlife-related pictures on social media tend to focus on well-known, large-bodied mammals, particularly iconic species like the Big Five (Hausmann et al., 2017; Walden-Schreiner et al., 2018; Toivonen et al., 2019). Hausmann et al. (2017) analysed tourists' preferences for nature-based experiences in Kruger National Park in South Africa and found that Flickr users, compared to those on other social media platforms, however, showed a broader interest in biodiversity. Unlike users who primarily focused on large charismatic mammals, Flickr users also shared posts about less-charismatic species, highlighting a more diverse range of biodiversity content. Hausmann et al. (2017) also noted that geotagged images on Flickr could serve as an effective tool for studying species occurrences and population dynamics, based on their analysis of the platform's posts. The use of Flickr data thus has the potential to complement traditional data collection methods, such as field surveys, particularly in regions where systematic biodiversity assessments are logistically challenging or resource intensive.

Despite its potential, there are several considerations and challenges associated with using Flickr and other social media platforms for conservation research. One key issue is the potential inaccuracy in the location of geotagged pictures (Crampton et al., 2013; Toivonen et al., 2019). Indeed, Toivonen et al. (2019) pointed out that social media users may upload photos from places with Wi-Fi access, such as lodges, rather than the actual location of the wildlife sighting. Although modern smartphones and cameras have GPS functionality that includes precise geolocation data in the metadata of uploaded images, the potential for inaccuracies remains a concern, especially when this functionality is not used or available.

Additionally, studies have found that Flickr data often exhibits biases toward specific regions and taxa. Hirvonen et al. (2020) observed that Flickr posts tend to be clustered in easily accessible areas, such as well-maintained sections of national parks with good road networks and tourist facilities. On the other hand, more remote areas tend to be underrepresented. This geographical bias is further compounded by taxonomic bias. Social media users are more likely to photograph and post images of larger, more charismatic species, leading to an overrepresentation of these animals and an underrepresentation of smaller or less iconic species (Hausmann et al., 2017; Fink et al., 2020). Another challenge with social media data is the potential for misidentification. Kays et al. (2022) noted that users of crowdsourced platforms often display overconfidence in their ability to identify species, which can introduce inaccuracies into species distribution models if datasets are based solely on user-provided tags. While expert review of images can help mitigate this issue, it can be impractical for large datasets, meaning misidentifications may remain uncorrected.

Finally, the ethical implications of using data from Flickr and other social media platforms in conservation studies must be carefully considered, particularly when it comes to user privacy. Geotagged posts can unintentionally reveal users' identities or personal information, potentially exposing users to risks, including unwanted attention or, in extreme cases, threats to personal safety (Di Minin et al., 2021a). In addition to privacy concerns, Di Minin et al. argue that users may be unaware that their publicly shared data is being used for research purposes, raising questions about consent. Social media users might not fully comprehend the extent to which their posts can be accessed and analysed by third parties, including researchers. This highlights the importance of ensuring transparency and ethical responsibility when leveraging publicly available data (Di Minin et al., 2021a). To mitigate these risks, researchers should employ strategies such as data minimization, anonymization, and adhering to strict data management protocols (Bolognini and Bistolfi, 2017; Hintze and El Emam, 2018; Di Minin et al., 2021a).

2.3.2. Digital accessible knowledge data and its relevance to biodiversity studies

Digital Accessible Knowledge (DAK) is a term used for biodiversity data that is freely accessible and digitally available, often collected from various sources such as museum specimens, field observations and citizen science platforms (Escribano et al., 2019; Ivanova and Shashkov, 2021; Ganglo, 2023). Citizen science platforms, such as iNaturalist and eBrid, integrate with DAK systems, resulting in a constantly updating data source (Di Cecco et al., 2021; GBIF, 2021). A key component in all DAK initiatives is that uploaded observations include coordinates (Gaiji et al., 2013; Sousa-Baena et al., 2014; Hjarding et al., 2015). This has led to DAK becoming a critical tool in modern biodiversity research by offering researchers a data source with a vast spatial and temporal coverage (Boitani et al., 2011; Neves et al., 2018; Taylor et al., 2018). The Global Biodiversity Information Facility (GBIF) stands out as by far the largest DAK initiative with over two billion species occurrence records uploaded by early 2023 (Ganglo, 2023). Established to support scientific research, conservation and sustainable development, GBIF serves as a comprehensive repository that aggregates biodiversity data from multiple sources (GBIF, 2021). GBIF aims to provide open and accessible data to enhance our understanding of species distribution patterns and to inform and promote effective conservation strategies (GBIF, 2021; Ganglo, 2023).

The usage of GBIF data in wildlife distribution and species richness modelling has been globally increasingly recognized in conservation science (Beck et al., 2013; Robertson et al., 2014; Callcut et al., 2018). The yearly number of peer-reviewed articles utilizing GBIF data has increased nearly tenfold from 2009 to 2019 (Fig. 1) (GBIF, 2021), demonstrating the growing importance and recognition of GBIF as a vital resource in scientific research. A major advantage of GBIF is its large spatio-temporal scale that allows researchers to conduct large-scale biodiversity studies that would otherwise be difficult to achieve due to the logistical and financial limitations of extensive fieldwork (Boitani et al., 2011; Neves et al., 2018). Indeed, data from GBIF has been widely used to model species ranges, identify biodiversity hotspots and detect changes in species presence over time (Chamberlain and Boettiger, 2017; Moreno et al., 2023; Trofino-Falasco et al., 2023). These applications of

GBIF data are vital for conservation planning, particularly in regions where traditional field data collection is impractical and resources for conservation are limited (Beck et al., 2013).

In Africa, the application of GBIF data for wildlife research has been demonstrated in various studies, highlighting its role in understanding species distribution and guiding conservation efforts (Neves et al., 2019). While the majority of GBIF studies in Africa have focused on plant life (Idohou et al., 2015; Ganglo, 2023), studies on mammal distribution have also been conducted. For example, Neves et al. (2018) successfully utilized GBIF data to compile a comprehensive list of terrestrial mammal species in Mozambique, aiding in the understanding of the country's biodiversity amidst historical challenges such as political instability and limited scientific exploration. Similarly, research in Angola's upper Okavango catchment integrated GBIF data with recent field surveys, providing an updated mammal checklist that included the documentation of significant historical range changes for larger mammals (Taylor et al., 2018). These studies illustrate the potential of GBIF data to fill knowledge gaps, especially in regions where traditional data collection methods are constrained by logistical difficulties.

Despite its vast potential, the use of GBIF data is not without challenges. One significant issue is the inherent biases in data collection, which can lead to uneven spatial coverage and completeness (Troia et al., 2016; Huang et al., 2020; Mesaglio et al., 2023). Data availability in GBIF can vary significantly across regions, often reflecting historical patterns of exploration, collection efforts and funding (Yesson et al., 2007; Troia et al., 2016; Mesaglio et al., 2023). Supporting this, Ganglo (2023) found that although Africa harbors rich biodiversity with numerous hotspots, the continent’s contribution to the GBIF database remains remarkably low. As of early 2023, Africa accounted for only 2.69% of the global records on GBIF, with substantial disparities among countries (Ganglo, 2023). Indeed, countries with higher tourism income or better-funded research infrastructure tend to have more comprehensive biodiversity records in GBIF (Freeman and Peterson, 2019; Ganglo, 2023; Moreno et al., 2023). For example, South Africa alone contributed over half of the continent's data, indicating significant regional imbalances (Ganglo, 2023).

Spatial bias occurs not only between countries but also within individual countries. In Benin, for example, regions with well-maintained infrastructure, such as good road networks, showed higher completeness of GBIF data, whereas remote and less accessible areas were significantly underrepresented (Idohou et al., 2015). Additionally, the availability of DAK data often depends on proximity to research institutions and access to local research funding, further influencing data distribution (Meyer et al., 2015). Similar patterns were observed by Escribano et al. (2019), who found that only a small fraction of grid cells in the Iberian Peninsula were well-sampled for mammal data, highlighting notable gaps in species distribution knowledge, particularly in less developed areas. Beck et al. (2013) also reported that GBIF data for European hawkmoths did not fully capture their range and climatic niches, suggesting that independently compiled datasets could offer more comprehensive insights. Furthermore, a case study on East African chameleons revealed that the lack of spatial completeness of DAK data can affect the accuracy of species distribution models, emphasizing the need to integrate GBIF data with local expertise and targeted field surveys to improve data reliability and conservation outcomes (Hjarding et al., 2015). Similar findings were done by De Araujo et al. (2022) and Štípková et al. (2024) who noted that due to spatial bias, data from GBIF was not accurate enough when used alone.

While the number of GBIF-related articles has significantly increased in recent years, this growth has, again, not been uniformly distributed worldwide (Ganglo, 2023). The low amounts of GBIF data available in Africa are strongly reflected in the amount of publications using GBIF data in the area (Fig. 1). In 2019, Africa was the lowest ranking continent in the usage of GBIF data in peer-reviewed articles with only 58, accounting for just 3.9% of the global total (GBIF, 2021). In stark contrast, Europe, the leading continent in GBIF-related research, had just under six hundred publications the same year (GBIF, 2021). In Kenya, fewer than nine peer-reviewed articles utilized GBIF data in 2019, highlighting both the scarcity of available data and the limited research activity using GBIF in this region (GBIF, 2021).

Taxonomic bias presents another significant challenge when using crowdsourced biodiversity data from DAK platforms like GBIF (Stahlschmidt, 2011; Sousa-Baena et al., 2014; Troudet et al., 2017). This sampling bias results in certain taxa, such as birds and mammals, being overrepresented, while others, particularly invertebrates and less charismatic species, are often underrepresented (Bonnet et al., 2002; Clark et al., 2002; Wilson et al., 2007; Ford et al., 2017). Troudet et al. (2017) highlighted that birds are by far the most recorded taxa in GBIF, mostly due to their popularity among both professional scientists and citizen scientists. The widespread enthusiasm for birdwatching has led to a substantial amount of bird-related data in DAK platforms, making these datasets far more robust than those for other taxa groups (Troudet et al., 2017; Huang et al., 2020). In stark contrast, taxa like insects and amphibians are much less represented, which poses challenges for comprehensive biodiversity assessments (McKinney, 1999; Beck et al., 2014). Additionally, Stephenson et al. (2021) found that in Africa less data on small-bodied mammals were available on GBIF compared to large-bodied mammals, highlighting that in the realm of DAK platforms, size indeed matters.

The persistence of taxonomic bias over time indicates that societal preferences heavily influence which species are studied and recorded (Wilson et al., 2007; Di Marco et al., 2017; Troudet et al., 2017). Public interest, driven by cultural salience, plays a significant role in data collection efforts, often more so than scientific interest alone (Phaka et al., 2022). Ford et al. (2017) agree that charismatic species, such as birds and mammals, and especially megafauna, tend to be more culturally and visually appealing, leading to greater public engagement and, subsequently, more extensive data on these groups. This preference impacts not only data collection but also funding allocation and research focus, perpetuating a cycle where well-known species continue to be even more overrepresented in scientific studies and databases (Leather, 2009; Ford et al., 2017; Troudet et al., 2017). The implications of taxonomic bias are far-reaching. It limits the effectiveness of biodiversity conservation efforts by providing an incomplete picture of ecosystems, skewing our understanding of species distributions and potentially leading to uninformed conservation decisions (Di Marco et al., 2017). Underrepresented taxa, which often play crucial ecological roles, may be overlooked in conservation planning, risking their decline without detection (McKinney, 1999; Troudet et al., 2017). This gap in knowledge is particularly concerning for organisms that contribute significantly to ecosystem functioning, such as insects and fungi.

With these identified limitations in mind and given the limited number of similar studies in East Africa, this study is eager to evaluate the effectiveness of crowdsourced data using DAK data from GBIF as a proxy for species occurrence and richness in the Taita Hills Wildlife Sanctuary.

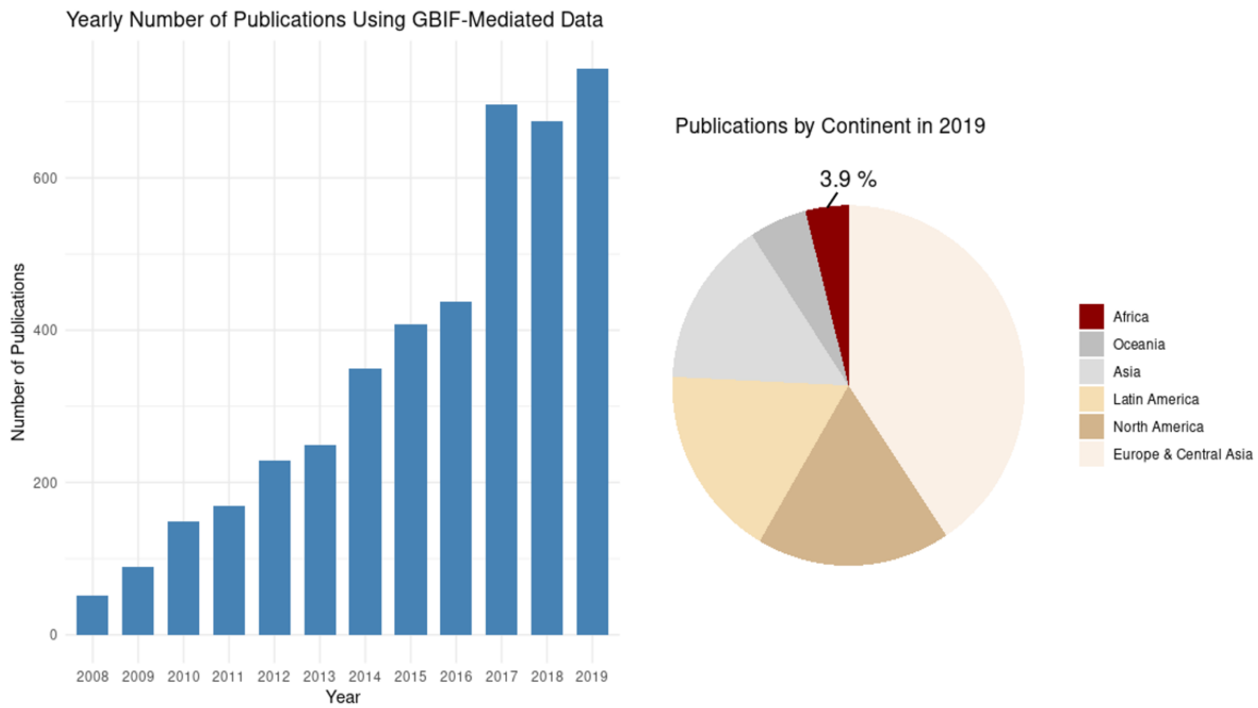

Figure 1. Yearly number of peer-reviewed articles utilizing Global Biodiversity Information Facility-mediated data alongside the percentage of publications originating from Africa. Data sourced from the GBIF Annual Report (2021).

3. Methodology

3.1. Study Area

The study was conducted in the wildlife watching area of Taita Hills Wildlife Sanctuary (THWS) (3° 30' 43.524'' S, 38° 15' 4.752'' E, Taita Hills Safari Resort & Spa), which is located in Taita-Taveta county in southeastern Kenya (Fig. 2). The study area spans approximately 110 km^2, with an altitude ranging from 900 to 1100 meters above sea level (Wachiye et al., 2022). THWS was selected as the study site due to its biodiversity and its importance for conservation (Taita Hills Wildlife Sanctuary, 2024). Additionally, THWS is a popular destination for wildlife tourism (Taita Hills Wildlife Sanctuary, 2024), which increases the availability of crowdsourced data from platforms like Flickr and GBIF. The presence of the Taita Research Station, established by the University of Helsinki, further enhances the suitability of the site by providing logistical support and facilitating the deployment of trail cameras. Furthermore, the sanctuary is staffed by wildlife rangers who contribute local knowledge of species distributions patterns. These factors make THWS a suitable location for studying the value of integrating crowdsourced data, trail camera images, and ranger-based knowledge for monitoring species occurrence and richness patterns. For the purposes of this study, the wildlife viewing area of THWS was divided into grid cells using a 3 km x 3 km grid (Fig. 2). This grid size was chosen to provide an optimal spatial resolution for data aggregation, as recommended by Wolters et al. (2006) in a study of a similarly sized area.

As part of the Greater Tsavo ecosystem THWS is characterized by tropical savanna grassland with scattered tree cover; *Acacia sp., Newtonia sp.* and *Vachellia* sp. being some of the most important tree taxa (Amara et al., 2023; Dickens et al., 2018; Omoro, Starr and Pellikka, 2013; Mukeka et al., 2020). Taita-Taveta county experiences a tropical semi-arid climate with two distinct rainy seasons (Pellikka et al., 2018). Rains usually fall from March to May and October to December, while temperatures reach their lowest between June and August and peak in February and March (Nyambariga et al., 2023). The field work was carried out in 2024 from January to March, months that are generally characterized by hot and dry weather. However, El Niño phenomena had brought extensive rains during the previous months resulting in tall vegetation and unusually wet landscape (Fig. 3) (Kenya meteorological department, 2023).

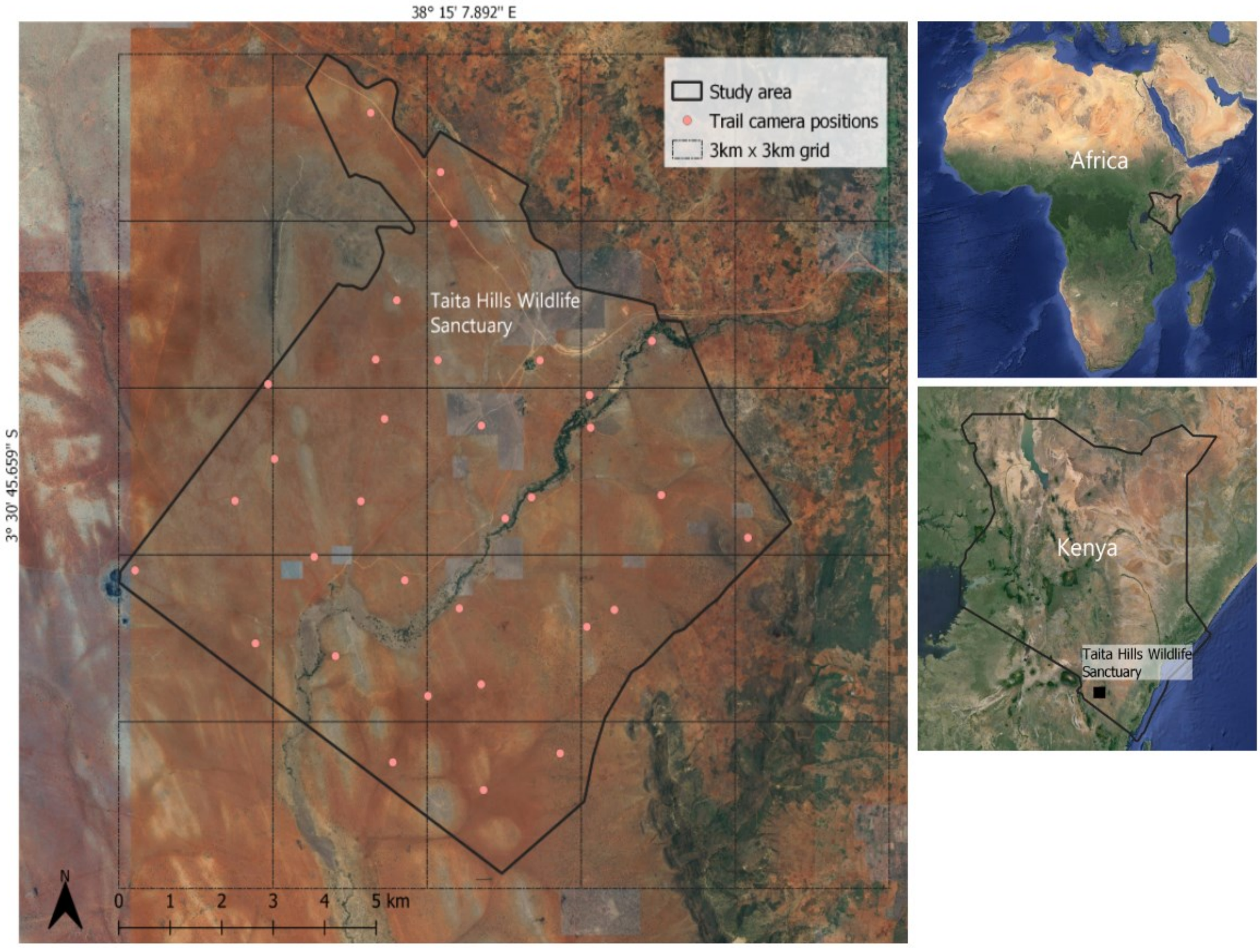

Figure 2. Location of the study area and trail camera positions.

THWS is a privately owned wildlife reserve that was established in 1970 to promote ecotourism and wildlife conservation (Njogu and Dietz, 2006). The sanctuary is regionally an important wildlife conservation area with over 50 mammal species and 390 bird species documented within its borders (Taita Hills Wildlife Sanctuary, 2024). THWS is also home to four of the Big Five species — lion, leopard, elephant, and buffalo — while black rhinos have gone extinct in the sanctuary. Therefore, for the remainder of this study, I will refer to these species collectively as the Big Four. The presence of the remaining Big Four species is an important contributor to local safari tourism, generating significant revenue (Ministry of Tourism and Wildlife, Kenya, 2022). However, the abundance of megafauna, especially large-bodied herbivores, in the sanctuary is greatly influenced by seasonal rainfall patterns, which directly affect the quality and availability of water sources and grazing (Muteti and Maloba, 2013; Mukeka et al., 2013). During the dry season, wildlife gathers around manmade water holes in THWS, the Bura River, and the surrounding riverine forests, attracted by the availability of water and fresh vegetation (Amara et al., 2020). In contrast, during the rainy season, wildlife disperses beyond the sanctuary's borders as water and grazing resources become more widely available (Schmitz, 2008). As observed by Muteti and Maloba (2013), during a dry season ground census in November 2013, 462 elephants were recorded within the sanctuary, whereas only 17 were sighted during the wet season census in June 2013. As a result of this seasonal movement, THWS also serves as an important migratory corridor for megafauna between Tsavo West and Tsavo East national parks (Smith and Kasiki, 2000; Mukeka et al., 2013).

From 1962 to 2019, the human population of Taita-Taveta County has more than tripled with an increase from 90,000 to 340,671 (Platts et al., 2011; Kenya National Bureau of Statistics, 2019). Agriculture is the main source of income for over 80% of the county’s residents, most of whom are small-scale farmers (Autio et al., 2021). This population growth and agricultural dependence have significantly increased the demand for land resources, water supplies and energy production (Pellikka et al., 2013). Consequently, there has been heightened pressure on local communities and wildlife, intensifying human-wildlife conflicts, such as crop raiding, in the region (Long et al., 2019; Mukeka et al., 2020; Mwakio Mwadime and Mbataru, 2022). According to Long et al. (2019), Taita Taveta County experienced the highest number of incidences of human-wildlife conflicts in Kenya between 2005 and 2016, further highlighting the need for effective and sustainable conservation planning.

Contrary to neighbouring LUMO conservancy, livestock grazing is minimal within the sanctuary borders (Wachiye et al., 2022). To reduce human-wildlife conflicts outside the sanctuary borders, THWS is fenced along its northern boundary (Amara et al., 2020). However, Li et al. (2014) and Packer et al. (2013) argue that fencing wilderness areas can disrupt ecological processes, such as natural wildlife migrations, often leading to issues like overgrazing. Supporting this, Amara et al. (2020) observed that large-bodied herbivores, particularly elephants, play a crucial role in maintaining the open grasslands of THWS by limiting tree growth, especially during the dry seasons. Consequently, this has created a landscape where tree coverage and aboveground biomass is significantly greater outside the sanctuary compared to inside (Massey et al., 2014; Vesala et al., 2017; Amara et al., 2020).

Figure 3. Taita Hills Wildlife Sanctuary had received atypical amounts of rain in previous months resulting in tall grass and weed growth. Picture taken in Taita Hills Wildlife Sanctuary in January 2024.

3.2. Data collection

3.2.1. Collection of crowdsourced biodiversity data

To assess whether crowdsourced biodiversity datasets can serve as reliable proxies for species richness and occurrence in THWS, we utilized two data sources: social media photos from Flickr (https://www.flickr.com) and DAK data from GBIF (https://www.gbif.org). Flickr was selected as the social media platform for this study due to its open API, which allows researchers to access and analyse spatial metadata associated with user-generated content. As noted in previous studies (Toivonen et al., 2019; Hirvonen et al., 2020), Flickr posts are often concentrated in national parks and tourist sites, making it a potential resource for studying species distributions in wildlife viewing areas. Additionally, Flickr has been identified as a primary platform for sharing biodiversity-related images, compared to other social media platforms like Instagram and Twitter (Hausmann et al., 2017; Toivonen et al., 2019). The study focused on three taxa groups: birds, mammals (excluding Big Four species), and the Big Four species. Table 1 provides a summary of the year of record submission for each data source and taxa group, offering the temporal context for the records used in this research.

Geotagged images from Flickr were extracted by filtering for location, initially covering the entire Greater Tsavo area to provide a comprehensive overview of the types of photos shared and how different taxa were represented. To separate data for different taxa groups, an initial keyword search was conducted. We used the terms "lion," "elephant," "buffalo," and "leopard," to find relevant posts for the Big Four taxa. These search terms were translated into Kiswahili, French, German, Italian and Spanish to account for the most commonly used languages by tourists and local residents in the region. Kiswahili and English are official languages in Kenya widely spoken by the local population (Kenya National Bureau of Statistics, 2020). French, German, Spanish and Italian were included because tourists from these nations constitute a significant proportion of international visitors to Kenya (Kenya Tourism Board, 2019). However, this search resulted in a very low number of detected posts. Additionally, we encountered several issues. Many posts contained the search terms but were unrelated to the target species, such as captions like “Buffalo weaver,” which refer to bird species rather than the intended mammal. This, along with concerns about potential species misidentification for more difficult taxa, led us to adopt a more comprehensive approach. We decided to download all available posts from the Greater Tsavo area, resulting in a total of 3,921 photos. Each image was manually reviewed to accurately identify the taxa group and, when possible, the specific species. The data were organized into an Excel table, with additional columns added for taxa group and species identification. Following this, the data was divided into three separate datasets: one for birds, one for mammals (excluding the Big Four), and one specifically for the Big Four species. To address potential privacy concerns, privacy measures such as data minimization and anonymization were applied in the subsequent pre-processing steps (Di Minin et al., 2021a). These data tables were subsequently imported into QGIS (QGIS Desktop 3.22.16 Long Term Release) for spatial analysis, utilizing the geotagged coordinates from each image to assess spatial distribution.

In parallel, species occurrence data was sourced from GBIF, using the Greater Tsavo area as the selection boundary. GBIF’s extensive database resulted in total of 243 297 observations for the Greater Tsavo ecosystem. After downloading, the dataset was processed by splitting it into three separate tables corresponding to the study's taxonomic groups. This data organization was performed in the Posit Cloud environment using Python language. Once organized, the datasets were imported into QGIS for further spatial analysis.

Table 1. Data availability years for each taxa group from the Global Biodiversity Information Facility and Flickr, providing temporal context for the species records used in this study.

Platform	Taxa group	Record year
GBIF	Mammal	2003–2023
GBIF	Big Four	2016–2020
GBIF	Birds	1981–2023
Flickr	Mammal	2013–2023
Flickr	Big Four	2009–2023
Flickr	Birds	2013–2023

3.2.2. Trail camera data collection

To monitor *on-the-ground* species occurrence and richness, we placed motion-activated trail cameras within each grid cell across THWS. The fieldwork was conducted from January 14th to March 23rd, 2024. To optimize our survey efforts, we installed cameras only in cells with a sufficient road network (n = 17), resulting in two cells without camera coverage (Fig. 2). We utilized two types of trail cameras: the APEMAN (https://apemans.com) Trail Cam H45 (n = 4) and the Wildgame (https://www.wildgameinnovations.com) Terra Extreme (n = 10). These models were chosen for their affordability and reliability. Both models feature a 90° detection angle and a 73° field of view (FOV). All units were fitted with a fast 32 GB memory card. The cameras were mounted on trees at heights ranging from 1.5 to 2.6 meters above the ground (Fig. 4), following recommendations from previous studies (Glover-Kapfer et al., 2019; Wachiye et al., 2022) to minimize disturbances from large mammals, such as elephants, which may rub against the trees. Each camera was set to capture three images whenever motion was detected, ensuring that we had multiple frames to reduce the risk of unclear or blurry images (Swanson et al., 2015; Egna et al., 2020). Additionally, the cameras were programmed with a 1-minute recovery time between triggers to avoid capturing multiple sets of images of the same individual, as advised by Van Berkel (2014) and Palencia et al. (2022).

Due to budget limitations, we had access to only 14 trail cameras during the fieldwork, which was insufficient to simultaneously cover all 17 grid cells in our study area. To address this limitation and ensure comprehensive spatial coverage, we implemented a rotation strategy for the cameras. Each 3 km × 3 km grid cell was designated to have a total temporal coverage of one month. Since we couldn't deploy cameras in all grid cells at the same time, we rotated the cameras between grid cells. While most cameras were placed in a grid cell for one month before being moved to a new location, some cameras were rotated every two weeks to ensure that all grid cells received coverage within the study timeframe. Within each grid cell, trail cameras were strategically positioned based on the local expertise provided by sanctuary rangers. Given the limited timeframe for fieldwork, we aimed to identify and target areas with the highest animal densities in each cell, rather than relying on random camera placement. When possible, cameras were placed along known wildlife trails or near permanent water sources that attract wildlife. For camera positioning, we prioritized shaded trees that provided shade to the cameras, minimizing the likelihood of false triggers, following the advice of Swanson et al. (2015). Additionally, tall grass was trimmed around the FOV of trail cameras and low-hanging branches were pruned to ensure an unobstructed view from the camera and minimize false triggers (Swanson et al., 2015). Each time a trail camera was installed, the tree species, surrounding un-cut grass length, nearest post number and GPS coordinates were recorded.

To achieve consistent spatial coverage, each full grid cell (9 square kilometres) was equipped with three cameras, resulting in an average density of one camera per 3 square kilometres. For grid cells that did not reach full capacity, calculations were made to determine the appropriate number of cameras. When calculations resulted in fractional camera allocations above or at .5, such as 2.6 cameras, the number was rounded up to ensure comprehensive monitoring. Conversely, fractional results below .5, such as 2.4, were rounded down to optimize camera use efficiency. For example, a grid cell covering 7 square kilometres (7km2/9km2 * 3 = 2.333) required two cameras.

Images from the trail cameras were downloaded and the batteries and memory card were checked every two weeks. At the conclusion of the field work, after retrieving all the cameras, each photo was manually reviewed, and all mammal species and the Big Four species were identified to species level. While trail cameras were an essential tool in our study, they were not used to count bird species, as previous research (Agha et al., 2018; Glover-Kapfer et al., 2019; Trofino-Falasco et al., 2023) have shown that this method does not yield meaningful results for avian species. The number of observations and species recorded in each grid cell was finally counted and then converted into rank values on a scale from 0 to 5, using natural breaks classification (Jenks) to determine the appropriate boundaries. The natural breaks classification method was selected as it is widely used in ecological and spatial studies due to its ability to identify natural groupings in the data, minimizing variance within classes while maximizing variance between them (Chen et al., 2013).

Finally, the use of trail cameras presents ethical considerations, particularly concerning the potential bycatch of humans in images (Rovero et al., 2014; Dupuis-Desormeaux et al., 2016; Palmer et al., 2018; Sandbrook et al., 2018). This is especially true in areas near the eastern and northern borders of the sanctuary where human activity is more frequent. Unintentionally capturing images of local people, tourists or potentially poachers may result in ethical dilemmas regarding consent and the use of such data (Sollmann, 2018; Sandbrook et al., 2018; O'Brien et al., 2020). To mitigate this issue, careful consideration was given to camera placement, aiming to reduce the likelihood of capturing human activity. Whenever a trail camera was placed, we made sure that the FOV of each camera was not covering roads used by tourists and park officials. However, in border regions where human-wildlife interaction is higher, the risk of bycatch remained. Following the ethical guidelines set by previous research (Sandbrook et al., 2018; Sandbrook et al., 2021), all images were reviewed upon trail camera retrieval. Any images containing identifiable human subjects were immediately deleted to ensure that no personal data or sensitive information was stored or used in the analysis.

Fig. 4. Placing a trail camera on the field. Picture taken in Taita Hills Wildlife Sanctuary in January 2024.

3.2.3. Ranger workshop

In order to complement trail camera data and gather accurate ground-truth information on wildlife distribution and species richness in THWS, we conducted a participatory ranger workshop with six wildlife rangers from THWS. The workshop was divided into two main sections, with the first section further split into three parts corresponding to the three taxa groups under study. During the first three parts, the rangers, working as a group, were presented with blank maps of the sanctuary divided into 3 km x 3 km grids. They were asked to rank each grid cell based on its suitability for wildlife watching for each of the three taxa groups. The ranking criteria varied slightly depending on the taxa. For the Big Four, rangers provided a ranking based on the likelihood of spotting any of the species present in the sanctuary. This approach aimed to identify the most important areas for the Big Four, emphasizing grid cells where individual abundance was the highest. For mammals and birds, the rankings were based on the expected richness of species in each cell. The rangers assigned ratings from 0 to 5, with 5 indicating the most suitable cells for observing each specific taxa group.

The second section was dedicated to reviewing the species distribution and richness maps based on data derived from both Flickr and GBIF. Again, maps for all three taxa groups were reviewed separately. Rangers were tasked with critiquing each map by marking a plus or minus sign in cells they believed were under- or overrepresented. If they felt a map accurately reflected the spatial distribution of a taxa group, no markings were made. This feedback was instrumental in validating the external datasets as well as prompting rangers to re-evaluate their initial rankings considering this additional information.

3.3. Pre-processing and mapping of external biodiversity data

Before mapping and analysing, the raw data collected from Flickr and GBIF underwent pre-processing. Firstly, data privacy concerns were addressed by minimizing the data to ensure user privacy as suggested by Di Minin et al. (2021a). Accordingly, data minimization involved only retaining the essential metadata, such as geolocation and species identification, necessary for spatial analysis, while any personally identifiable information (PII) was removed from the datasets. All datasets were processed to protect sensitive information while maintaining the integrity of the spatial and ecological data. These steps are crucial to adhering to ethical standards in research using crowdsourced data.

Secondly, in QGIS, we removed preserved specimens from each of the three GBIF datasets, ensuring that our study focused exclusively on observations of live individuals as advised by previous research (Kent et al., 2014; Phaka et al., 2022). As noted by Freeman and Peterson (2019), picture sharing platforms often feature multiple images stemming from a single observation. This leads to spatial bias in datasets with redundant entries. To address this issue, we removed duplicate records from both the Flickr and GBIF datasets, ensuring that each observation or photograph was only represented once. This process was carried out using the Delete Duplicate Geometries tool in QGIS. Next, the data from both sources were clipped to the extent of THWS, ensuring that only observations within the sanctuary's borders were considered for the final analysis. The spatial aggregation of the data was then carried out, using a 3 km x 3 km grid across the sanctuary. This was achieved in QGIS through the Points-in-Polygon tool, resulting in 19 grid cells, each containing the number of species for each taxa group as well as the total number of Big Four observations.

Finally, to facilitate comparative analysis across datasets and taxa groups, the datasets were standardized to a ranking scale from 0 to 5. This was accomplished using the natural breaks (Jenks) classification method to define appropriate boundaries, allowing for more meaningful comparisons across datasets and taxa groups. Finally, the processed data were mapped to visualize the spatial distribution of the rank values for each taxa group, highlighting patterns and trends within THWS.

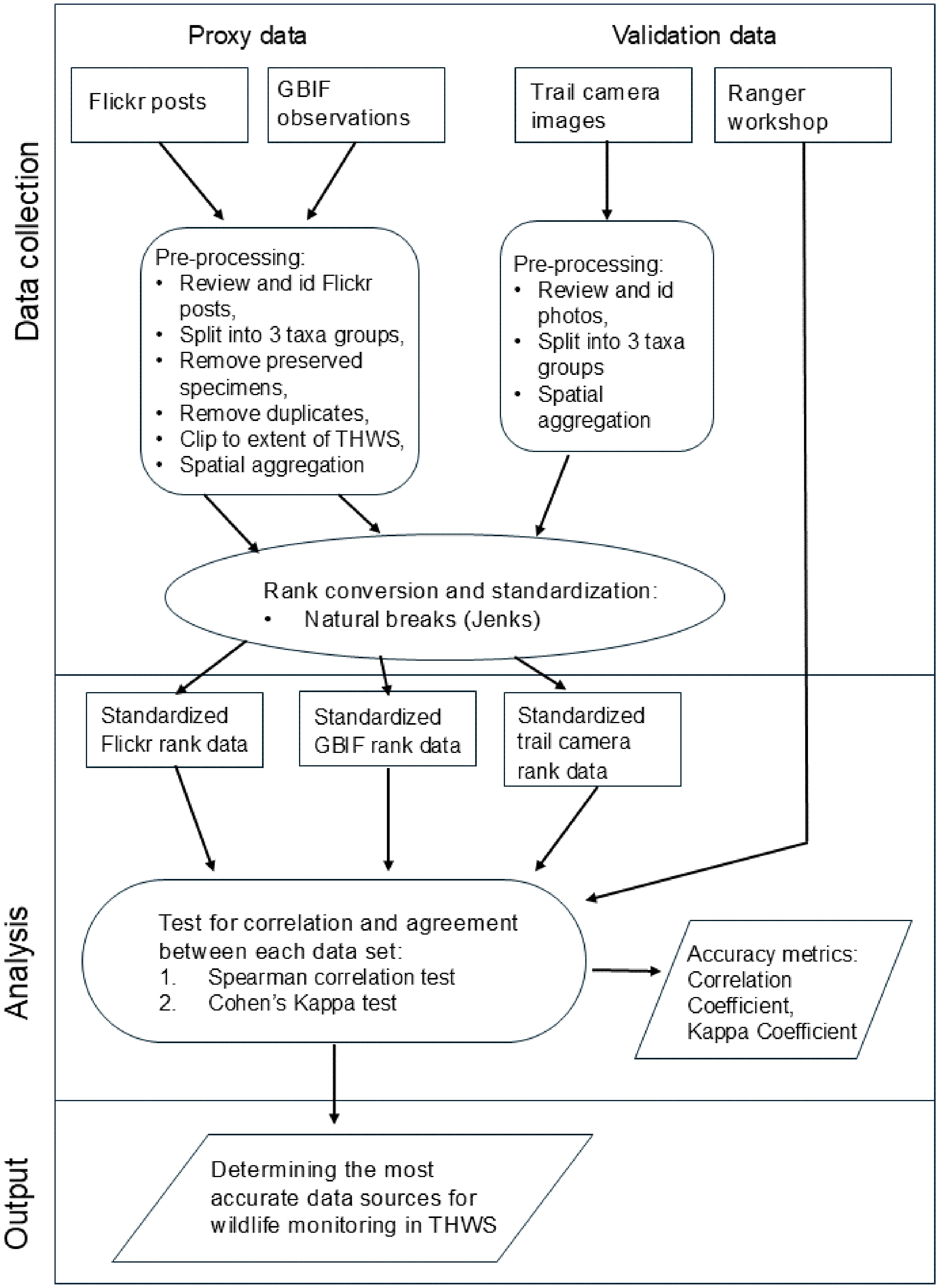

Fig. 5. Flowchart visualizing the workflow of data collection, pre-processing and analysis.

3.4. Data analysis

This study aimed to determine the most accurate data sources for predicting the spatial wildlife distribution and species richness in the study area. By comparing traditional field data with crowdsourced biodiversity data from social media photos from Flickr and DAK data from GBIF, the goal was to assess the reliability of these external sources as proxies for direct field observations (trail cameras) and ranger-based knowledge. We conducted statistical analyses to evaluate the agreement and correlation between these datasets, providing insights into their utility for wildlife monitoring and conservation efforts in THWS. Statistical analyses were calculated in RStudio version 2024.04.1 (RStudio, 2024).

In the first step of the data analysis, we began with a visual comparison of pre-processed and standardized rank maps for each taxa group across the four data collection methods. This allowed us to assess the spatial patterns of wildlife distribution across the study area and gain an early impression of how well the crowdsourced datasets matched the data derived from trail cameras and ranger workshops. This step was important for forming a preliminary understanding before moving into more formal statistical evaluations.

To quantitatively measure the relationships between the different datasets, we used Spearman's correlation test and Cohen's Kappa test. Initially, the Shapiro-Wilk test for normality was conducted, which indicated that none of the datasets followed a normal distribution. As a result, the non-parametric Spearman's correlation test was chosen to measure the strength and direction of the relationships between datasets for each taxonomic group as advised by previous research (Barve et al., 2005; Msoffe et al., 2007; Jori et al., 2011). This step of analysis was particularly focused on exploring how well the crowdsourced data sources correlated with ground-truth data from ranger-based knowledge and trail camera data. Additionally, we were interested in testing whether ground-truth data were internally consistent to assess the effectiveness of short-term trail camera deployment following the rainy season.

Complementing this, Cohen's Kappa test was conducted to measure the level of agreement for rank values between datasets. While Spearman's correlation coefficient assesses the strength of relationships, Cohen's Kappa coefficient measures the consistency of ranked categorical data across different data sources (Flather and King, 1992; Yang et al., 2017). The Kappa test allowed us to determine whether the datasets consistently agreed on the exact rank values assigned to each grid cell. This was especially important for discovering whether the standardized ranks, ranging from 0 to 5, were consistently applied across datasets. As a result of using both Spearman's correlation and Cohen's Kappa, we gained complementary insights into the data relationships.

We conducted pairwise comparisons between the four datasets to assess their alignment and effectiveness for each of the three taxa groups. Each pairwise comparison was done with standardized rank values. The analysis focused on how well each dataset aligns especially with rankings derived from the ranger workshop, considered the ground truth due to the rangers' extensive experience in the study area. Specifically, we compared ranger-based knowledge with crowdsourced datasets from Flickr and GBIF to determine how well our crowdsourced datasets align with expert knowledge. Additionally, we compared crowdsourced datasets with trail camera data to explore the relationships between crowdsourced data and direct field observations. Finally, we compared Flickr with GBIF to assess the consistency between the two crowdsourced datasets for each taxonomic group.

Furthermore, to evaluate the effectiveness of short-term trail camera deployment following the rainy season, we compared the standardized and ranked trail camera data with ranger-based knowledge. Focusing on the Big Four species and other mammals in THWS, this part of analysis assessed how well the trail cameras captured species richness and distribution over the limited deployment period. By examining the correlation and agreement between the two data sources, we aimed to determine whether short-term deployments, in the seasonal context of our study, provided reliable data for wildlife monitoring.

To further validate our findings, qualitative validation was incorporated into the analysis using reference data from the ranger workshop. During the second part of the workshop, rangers reviewed and critiqued the maps derived from external data sources. This hands-on review allowed for visual validation of the datasets, as rangers provided input on the accuracy of the spatial patterns presented. Their feedback highlighted areas where crowdsourced datasets either aligned with or diverged from their local knowledge, helping us to identify any over- or under-represented grid cells. This expert validation was important for reinforcing or questioning the findings from the quantitative analyses.

4. Results

4.1. Descriptive statistics of crowdsourced wildlife datasets

The numbers of Flickr posts and GBIF records for each taxonomic group within the study area are presented in Table 2. Observation counts varied widely, ranging from as few as 16 to over 1600, highlighting significant differences in sample sizes among the taxa groups. These disparities between taxa representativeness strongly suggest that these crowdsourced datasets exhibit taxonomic bias. Furthermore, both datasets exhibit clustering around the two lodges in the sanctuary as seen in the observation heatmap presented in Figure 6. Figure 6 highlights that while most of the data is concentrated in the central and western parts of the study area, the eastern and southern areas are under-represented.

Mammal and Big Four species recorded on both platforms are listed in Table 3, which shows that Flickr primarily documents more common mammal species, while GBIF provides a more comprehensive list of species. Additionally, GBIF presents a broader list of bird species as seen from the Table 3. In total, the crowdsourced datasets provided records for 273 bird species within the study area. Given the length of this list, the complete bird species inventory is provided in Appendix Table A1.

Table 2. Number of observations and species per taxa group in Taita Hills Wildlife Sanctuary.

Taxa group	GBIF n. obs	GBIF n. sp	Flickr n. obs	Flickr n. sp
Big Four	*16*	*4*	*108*	*4*
Mammals (excluding Big Four)	*68*	*25*	*43*	*7*
Birds	*1630*	*272*	*39*	*18*
Total	*1714*	*303*	*184*	*29*

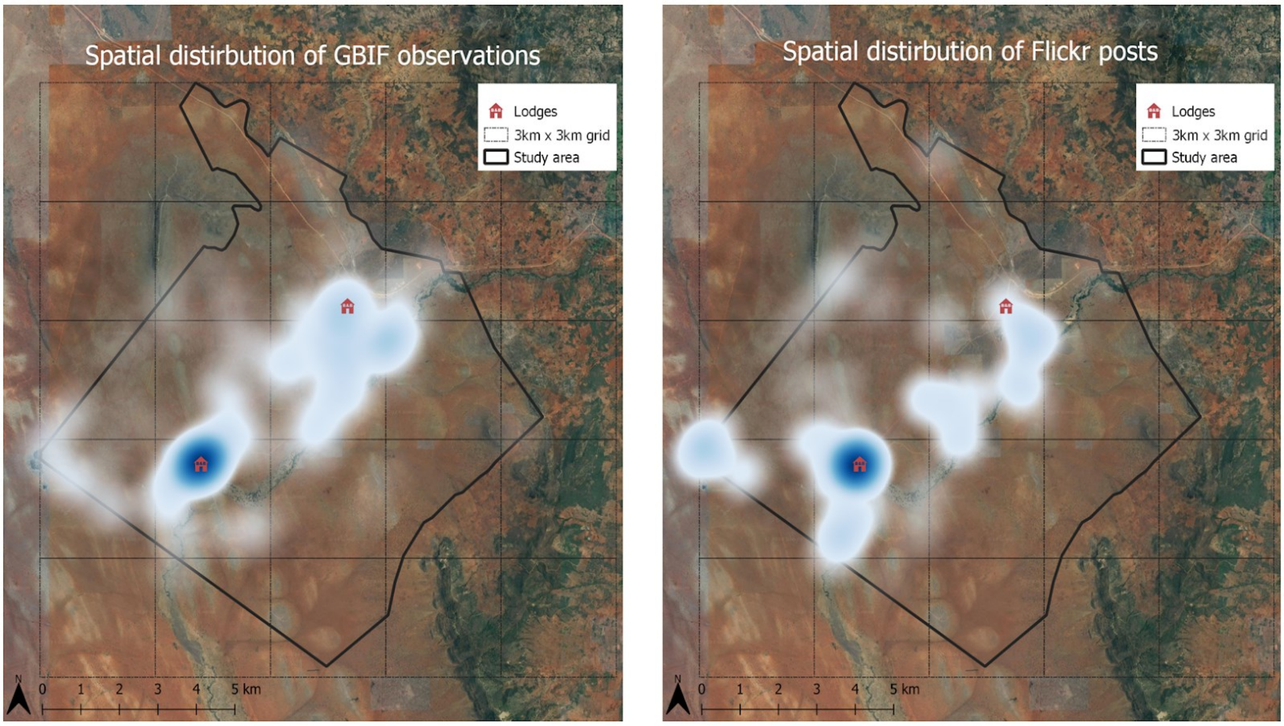

Fig. 6. Heatmaps representing the spatial distribution of Global Biodiversity Information Facility observations and Flickr posts.

Table 3. Mammal and Big Four Species recorded on Global Biodiversity Information Facility and Flickr in Taita Hills Wildlife Sanctuary.

Taxonomic group	Common name	Scientific name	Platform recorded
Big four	African elephant	*Loxodonta africana*	GBIF & Flickr
Big four	African Buffalo	*Syncerus caffer*	GBIF & Flickr
Big four	Lion	*Panthera leo*	GBIF & Flickr
Big four	African leopard	*Panthera pardus pardus*	GBIF & Flickr
Mammal	Bat-eared Fox	*Otocyon megalotis*	GBIF
Mammal	Black-backed Jackal	*Lupulella mesomelas*	GBIF
Mammal	Aardwolf	*Proteles cristatus*	GBIF
Mammal	Serval	*Leptailurus serval*	GBIF
Mammal	African civet	*Civettictis civetta*	GBIF
Mammal	Common genet	*Genetta genetta*	GBIF
Mammal	White-tailed mongoose	*Ichneumia albicauda*	GBIF
Mammal	Ethiopian dwarf mongoose	*Helogale hirtula*	GBIF
Mammal	Common dwarf mongoose	*Helogale parvula*	GBIF
Mammal	Banded mongoose	*Mungos mungo*	GBIF
Mammal	Red-legged sun squirrel	*Heliosciurus rufobrachium*	GBIF
Mammal	Brown greater galago	*Otolemur crassicaudatus*	GBIF
Mammal	Cape hare	*Lepus capensis*	GBIF
Mammal	Impala	*Aepyceros melampus*	GBIF & Flickr
Mammal	Grant's gazelle	*Nanger granti*	GBIF
Mammal	Thomson's gazelle	*Eudorcas thomsonii*	GBIF
Mammal	Hartebeest	*Alcelaphus buselaphus cokii*	GBIF & Flickr
Mammal	Plains zebra	*Equus quagga*	GBIF & Flickr
Mammal	Giraffe	*Giraffa tippelskirchi*	GBIF & Flickr
Mammal	Waterbuck	*Kobus ellipsiprymnus*	GBIF & Flickr
Mammal	Bushbuck	*Tragelaphus scriptus*	GBIF
Mammal	Greater kudu	*Tragelaphus strepsiceros*	GBIF
Mammal	Common warthog	*Phacochoerus africanus*	Flickr
Mammal	Vervet Monkey	*Chlorocebus pygerythrus*	GBIF
Mammal	Blue monkey	*Cercopithecus mitis*	GBIF
Mammal	Yellow baboon	*Papio cynocephalus*	GBIF & Flickr

4.2. Correlation and agreement analysis between data sources

The analysis revealed significant and high correlations between external biodiversity datasets and ranger opinions in the Taita Hills Wildlife Sanctuary. For the Big Four species, ranger rankings showed the highest level of correlation and agreement with rankings from Flickr data (Spearman's rho = 0.635, $p < 0.05$, $n = 19$; Cohen's Kappa = 0.539, $p < 0.05$, $n = 19$). Additionally, a significant but moderate correlation and agreement was found between Flickr data and trail camera rankings (Spearman's rho: 0.513, $p < 0.05$, $n = 17$; Cohen's Kappa: 0.463, $p < 0.05$, $n = 17$). However, GBIF did not show a significant correlation with ranger rankings for the Big Four (Spearman's rho: -0.3, $p > 0.1$, $n = 19$; Cohen's Kappa: -0.196, $p > 0.1$, $n = 19$). Similarly, there was no significant correlation or agreement ($p > 0.1$) with trail camera rankings and rankings from GBIF. Additional analysis between GBIF and Flickr data for Big Four species revealed a weak and non-significant negative correlation (Spearman's rho = -0.34, $p > 0.1$) and poor agreement (Cohen's Kappa = -0.2, $p > 0.1$). The limited number of GBIF observations for Big Four taxa likely contributes to the discrepancies observed between datasets.

Figure 7 illustrates the spatial distribution of mammal rank values for each dataset. Statistical analysis of mammal species richness rankings revealed significant correlations between both Flickr and GBIF with ranger-based knowledge, as depicted in Figure 8. While mammal rankings from Flickr correlated quite strongly with ranger rankings, the agreement between these datasets, while significant, was rather low (Spearman's rho: 0.657, $p < 0.05$, $n = 19$; Cohen's Kappa: 0.174, $p < 0.05$, $n = 19$). Mammal rankings from GBIF followed a very similar pattern (Spearman's rho: 0.674, $p < 0.05$, $n = 19$; Cohen's Kappa: 0.234, $p < 0.05$), indicating a high correlation but moderate agreement with ranger rankings. Interestingly, GBIF rankings correlated more strongly with trail camera data (Spearman's rho: 0.79, $p < 0.05$, $n = 19$; Cohen's Kappa: 0.232, $p < 0.05$, $n = 19$), indicating a high level of consistency with observed data from trail cameras. Flickr, on the other hand, did not show significant correlation with trail camera rankings ($p > 0.05$). Additionally, while mammal rankings between Flickr and GBIF showed moderate correlation (Spearman's rho = 0.414, $p = 0.077$, $n = 19$), the agreement between these two datasets was poor (Cohen's Kappa = 0.1, $p > 0.1$).

The analysis of bird species richness rankings showed weak and non-significant correlations between ranger rankings and both Flickr (Spearman's rho = 0.226, $p > 0.1$, $n = 19$) and GBIF data (Spearman's rho = 0.209, p > 0.1, n = 19), with low agreement values (Cohen's Kappa: 0.116 and 0.14, respectively). However, a strong correlation and moderate agreement were found between Flickr and GBIF rankings (Spearman's rho = 0.670, p < 0.05, n = 19; Cohen's Kappa = 0.342, p < 0.05, n = 19). These results suggest that while Flickr and GBIF data on birds are consistent with each other, they may not align well with local expert knowledge.

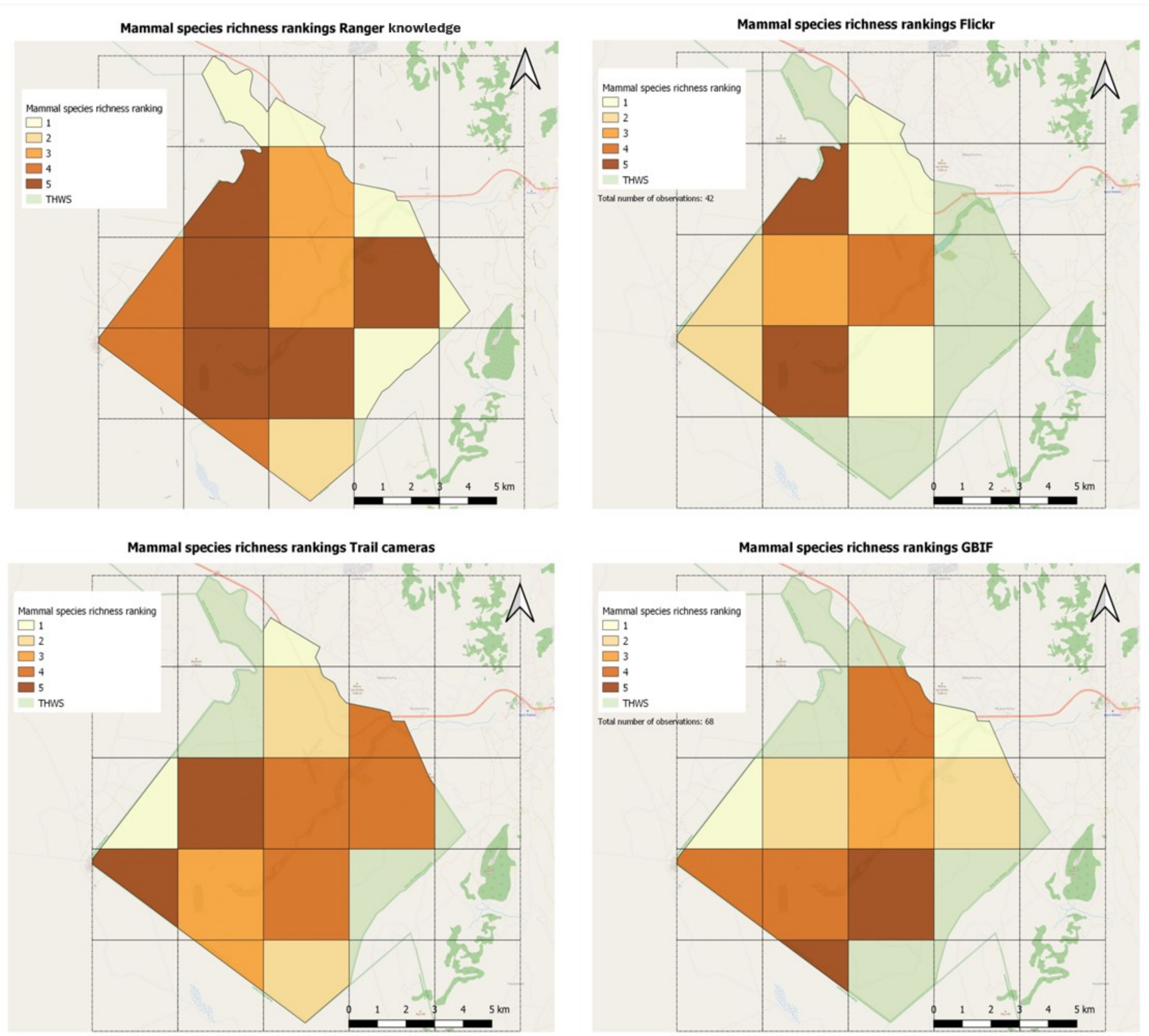

Fig. 7. Rank maps illustrating the spatial distribution of ranks for each data source, using mammals (excluding Big Four) as the representative taxa.

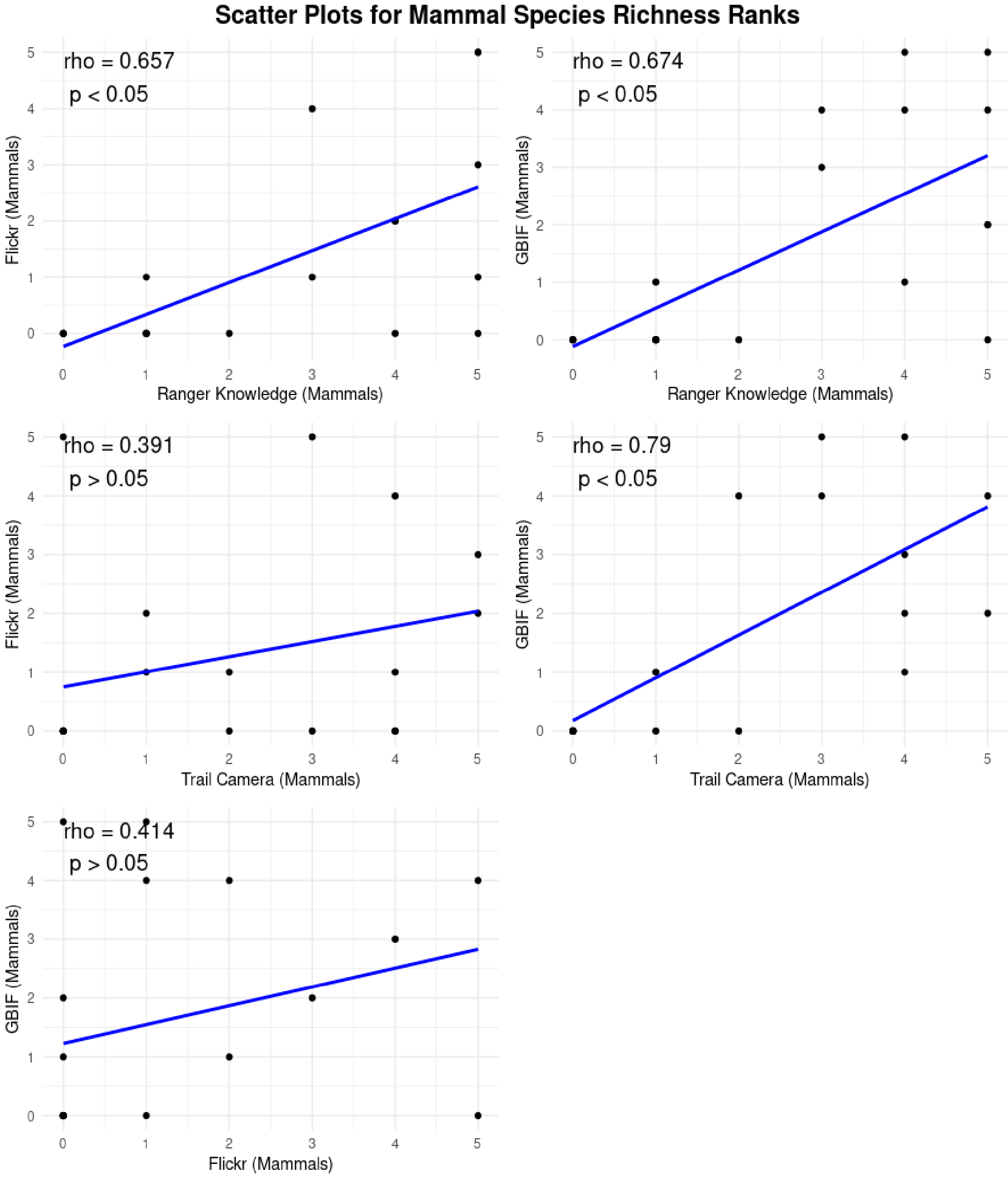

Fig. 8. Scatter plots visualizing the relationships between datasets, using mammals (excluding Big Four) as the representative taxa.

4.3. Effectiveness of short-term trail camera deployment

The 14 trail cameras deployed in this study accumulated a total of 420 field days. During this period our trail cameras managed to capture images of 15 species of mammals and two species of the Big Four, as shown in Table 3. The number of mammal species recorded per grid cell ranged from 0 to 6. Notably, the trail cameras recorded the bat-eared fox (*Otocyon megalotis*), black-backed jackal (*Lupulella mesomelas*), dik-dik (*Madoqua kirkii*), blue wildebeest (*Connochaetes taurinus*), lesser kudu (*Tragelaphus imberbis*), East African oryx (*Oryx beisa callotis*), common eland (*Taurotragus oryx*), and bushbuck (*Tragelaphus sylvaticus*)—species that were all absent from the Flickr dataset, with the dik-dik, lesser kudu, East African oryx, common eland and blue wildebeest also absent from GBIF. This underscores the complementary value of trail camera data. Furthermore, compared to the crowdsources datasets, the trail cameras provided more comprehensive coverage of mammal species across the study area, capturing data from 12 out of the 17 grid cells as demonstrated by Figure 7. For the Big Four, a total of 54 sightings were recorded, including 52 elephant sightings and 2 buffalo sightings. This demonstrates a limitation of the trail camera data compared to the crowdsourced datasets, as both Flickr and GBIF had observations of all four Big Four species present in the sanctuary. Though trail camera data fell short of the number of Flickr observations (108), it provided more than three times the number of Big Four sightings compared to the GBIF dataset (16), emphasizing its contribution to the study.

The relationships between ranger-based knowledge and trail camera data for the Big Four and other mammals are presented in Figure 9. Our statistical analysis revealed that for the Big Four species the correlation between trail camera data and ranger rankings was weak and non-significant (Spearman's rho = 0.219, $p > 0.1$, $n = 17$). Additionally, the agreement between the two datasets was low (Cohen's Kappa = 0.090, $p > 0.1$, $n = 17$). These results indicate that our trail camera deployment period was indeed too short and it did not comprehensively capture the distribution and presence of Big Four species as reflected in the ranger assessments. For other mammal species, on the other hand, trail camera data showed a moderate to strong correlation with ranger rankings (Spearman's rho = 0.635, $p < 0.01$, $n = 17$) as seen in Figure 9. The level of agreement, however, was relatively low, with Cohen's Kappa approaching significance (Cohen's Kappa = 0.182, $p = 0.059$, $n = 17$).

Table 3. Species detected by the trail cameras.

Taxonomic group	Common name	Scientific name
Big four	African elephant	*Loxodonta africana*
Big four	African buffalo	*Syncerus caffer*
Mammal	Bat-eared fox	*Otocyon megalotis*
Mammal	Black-backed jackal	*Lupulella mesomelas*
Mammal	Kirk's dik-dik	*Madoqua kirkii*
Mammal	Impala	*Aepyceros melampus*
Mammal	Hartebeest	*Alcelaphus buselaphus cokii*
Mammal	Plains zebra	*Equus quagga*
Mammal	Blue wildebeest	*Connochaetes taurinus*
Mammal	Giraffe	*Giraffa tippelskirchi*
Mammal	Waterbuck	*Kobus ellipsiprymnus*
Mammal	Bushbuck	*Tragelaphus scriptus*
Mammal	Lesser kudu	*Tragelaphus imberbis*
Mammal	East African oryx	*Oryx beisa callotis*
Mammal	Common eland	*Taurotragus oryx*
Mammal	Vervet monkey	*Chlorocebus pygerythrus*
Mammal	Yellow baboon	*Papio cynocephalus*

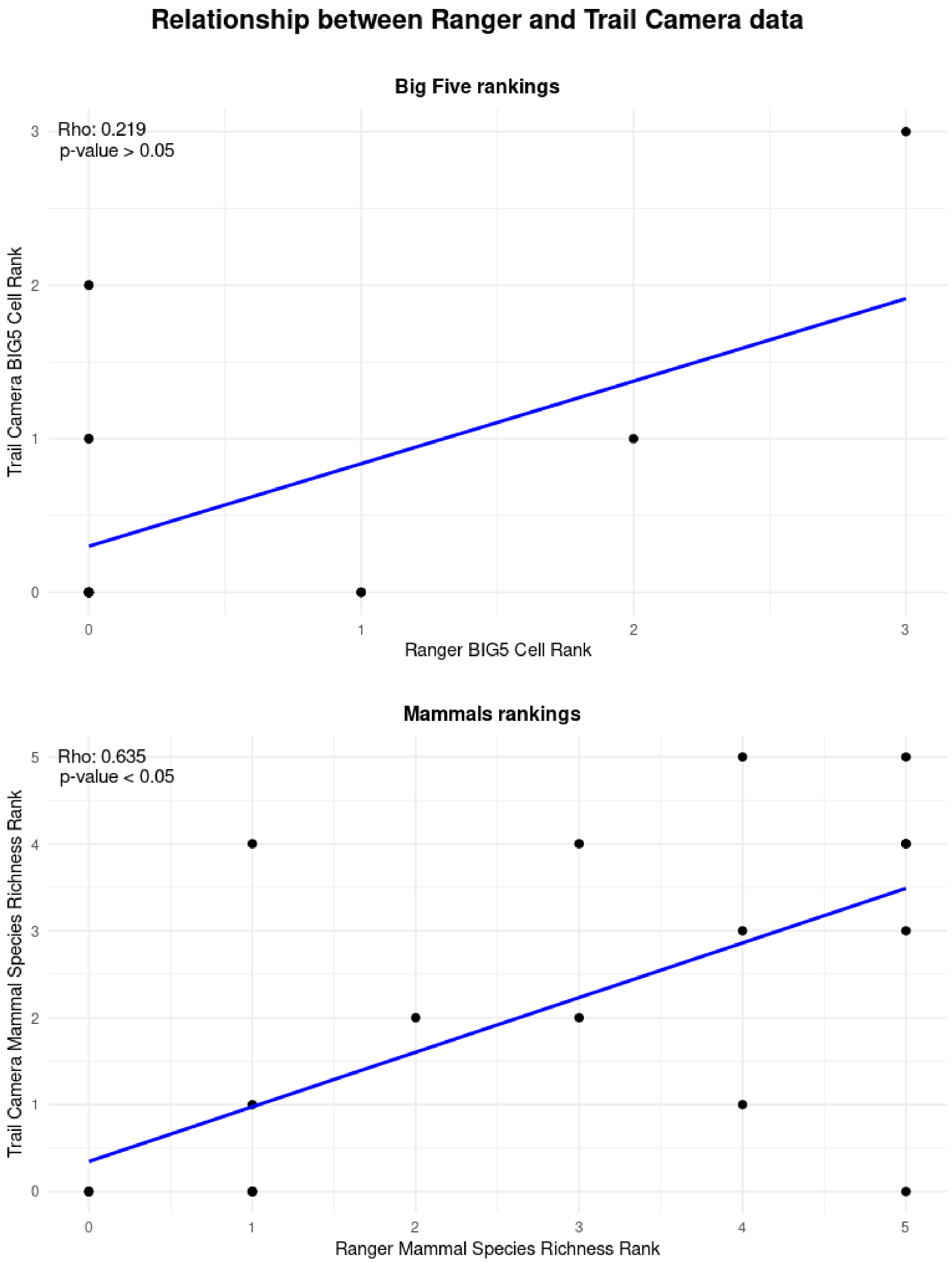

Fig. 9. Scatter plots visualizing the relationship between ranger-based knowledge and trail camera data for the Big Four and mammal taxa.

5. Discussion

5.1. Evaluation of crowdsourced data and trail cameras for species occurrence and richness mapping

The main objective of this study was to examine how well crowdsourced datasets from Flickr and GBIF align with ranger-based knowledge and trail camera data for species occurrence and richness mapping in Taita Hills Wildlife Sanctuary (THWS). This was accomplished by comparing rank values derived from each dataset through visual comparisons, statistical correlation and agreement testing, as well as expert validation, providing a comprehensive evaluation of the reliability of each dataset. The study focused on three taxonomic groups: birds, mammals, and the Big Four species. The principal findings of this study demonstrate the potential of crowdsourced biodiversity data and trail cameras as complementary tools for mapping species occurrence and richness in conservation areas like THWS. The study provides a nuanced understanding of the reliability and limitations of each data source. The results highlight significant strengths, particularly the ability of crowdsourced data to capture the occurrences of large, charismatic species such as the Big Four and other mammals. However, the study also emphasizes critical limitations, including taxonomic biases, spatial clustering of data, and the limited temporal coverage of trail camera deployment.

As noted by Toivonen et al. (2019), social media users are often more eager to document charismatic species. This tendency is evident in our study, where the number of Big Four sightings recorded on Flickr was notably high ($n = 108$). One of the key outcomes of our research is the strong relationship found between Flickr data and ranger-based knowledge for Big Four species, demonstrated by significant correlation and agreement (Spearman's rho = 0.635, Cohen's Kappa = 0.539, $p < 0.05$, $n = 19$). These findings suggest that Flickr, which relies heavily on tourist-generated content, can serve as a reliable proxy for ranger-based knowledge when monitoring the presence of charismatic species in areas where wildlife tourism is prominent. Our results align with previous studies that have highlighted the potential of Flickr data for monitoring large-bodied mammals in protected areas (Hausmann et al., 2017).

Furthermore, while the number of mammal observations ($n = 43$) and species observed ($n = 7$) in THWS in Flickr was relatively low, the mammalian species richness trend captured by Flickr correlated well with ranger-based knowledge (Spearman's rho: 0.657, $p < 0.05$).

This finding is supported by Allain et al. (2019) and Edwards et al. (2021), who discovered that Flickr can be a powerful data source for studying species occurrence also in less charismatic species, such as invasive species and garden birds. However, the results of our study also highlight the quite low agreement levels between the mammal richness ranks from Flickr and ranger-based knowledge (Cohen's Kappa: 0.174, $p < 0.05$). This indicates that while Flickr is capable of predicting the best general areas for mammalian richness, it does not necessarily rank the grid cells in the exact same order as rangers would. It thus lacks the fine scale capabilities of expert knowledge. Ultimately, in THWS, where tourism plays a significant role in contributing wildlife data, Flickr's open API and geotagged data provide an efficient means of gathering data on mammal species occurrence, including the Big Four, in accessible areas. While the number of studies using social media for these purposes is low, the consistency of these findings with previous literature supports the idea that social media data, especially from Flickr, can complement traditional field methods in providing species occurrence data for conservation science.

Another interesting finding of this study was the good performance of GBIF for mammals in the study area. GBIF provided over 50% more mammal sightings (n = 68) compared to Flickr and had over three times more species recorded (n = 25). Mammal ranks derived from GBIF showed a significant and high level of correlation with ranger-based knowledge (Spearman's rho: 0.674) and a moderate agreement (Cohen's Kappa: 0.234). Additionally, GBIF showed a particularly strong relationship with trail camera data for mammals (Spearman's rho: 0.79). The high correlation between GBIF and trail camera data highlights GBIF's potential as a practical proxy for direct field observations in mammal distribution studies. These findings are supported by previous research that has identified GBIF as an effective data source for studying mammal species occurrences independently (Neves et al., 2018) and for complementing traditional field surveys (Taylor et al., 2018). It is however interesting that GBIF exhibited a stronger alignment with trail camera data rather than direct ranger-based knowledge. This could be due to multiple reasons, including the small sample size of GBIF or the short temporal coverage of trail camera data. Nevertheless, these findings indicate that GBIF is a more reliable dataset than Flickr for representing mammal species richness patterns in our study area, given its strong alignment with both ranger-based knowledge and trail camera observations.

On the other hand, GBIF did not effectively capture trends for the presence of Big Four species. The weak and statistically insignificant correlation between GBIF data and ranger-based knowledge for the Big Four ranks (Spearman's rho = -0.3) and the low number of observations (n = 16) indicate that the platforms contributing to GBIF are not used to extensively record data for iconic megafauna in THWS. Rather, our findings indicate that in the study area, data contributors interested in sharing observations of the Big Four use social media for this purpose as suggested by Di Minin et al. (2015) and Toivonen et al. (2019). While we had acknowledged that taxonomic biases exist among less charismatic taxa in GBIF, as demonstrated by Troudet et al. (2017), it was surprising to find Big Four species such misrepresented. Indeed, previous research by Troudet et al. (2017) and Huang et al. (2020) have highlighted that birds and mammals, including the Big Four, are generally outliers in GBIF with disproportionate overrepresentation in most areas. Our findings hence underscore the importance of using multiple data sources for different taxa groups. While GBIF was shown to be a reliable proxy for mammal species richness, Flickr far outperformed GBIF when it came to Big Four species occurrence in our study area.

Flikcr's and GBIF's ability to predict bird species richness patterns in THWS, however, was poor. Correlations between ranger rankings and crowdsourced datasets were weak and statistically insignificant. Similarly, the agreements between these datasets were low and insignificant. These results suggest that Flickr and GBIF data do not align well with ranger-based knowledge when it comes to bird species distribution in the study area. The low number of observations (n = 39) and species (n = 18) recorded in Flickr reflect lower interest in birds from the users of Flickr. The proportion of bird pictures in Flickr within our study area follows a very similar trend observed in other African regions. Hausmann et al. (2017) found that in Kruger national park Flickr had approximately three times more posts on large-bodied mammals compared to birds. Contrastingly, Edwards et al. (2021) found that Flickr data in the UK aligns well with national databases for garden birds, highlighting the geographical bias within Flickr. Conversely, GBIF provided a substantial number of bird sightings (n = 1630) and species observed (n = 272) in the study area. This high volume of bird data aligns with previous studies that have noted the disproportionate over-representation of birds in GBIF datasets. The abundance of bird records emphasizes the taxonomic bias toward bird species within the platform, which has been documented in earlier research (Troudet et al., 2017; Huang et al., 2020).

Interestingly, a strong and significant correlation was found between GBIF and Flickr bird species richness rankings (Spearman's rho = 0.670), along with significant and moderate agreement (Cohen's Kappa = 0.342). This suggests that while the number of observations and species between the two platforms is vastly different, they tend to agree on the locations of high bird species diversity in THWS. This could be due to spatial clustering of GBIF and Flickr data. While both datasets align well with each other, they do not necessarily reflect the ground truth distribution as represented by ranger rankings. The high correlation between GBIF and Flickr indicates that bird sightings might be exceedingly reported from the same locations rather than comprehensively throughout the sanctuary. This limits the crowdsourced data's ability to be used for fine-scale bird research as shown here. However, the high number of bird occurrence data available from GBIF highlights its importance when conducting studies with lower spatial resolution.

The use of trail cameras in this study provided important ground-truth data to complement ranger-based knowledge. The 14 trail cameras used here accumulated 420 field days across the 17 grid cells where a sufficient road network existed. Mammal species richness rankings derived from trail camera data showed a significant strong correlation with ranger rankings (Spearman's rho = 0.635, $p < 0.01$). However, the level of agreement was relatively low (Cohen's Kappa = 0.182). While the data from trail cameras align with ranger-based knowledge on the general hotspots for mammalian diversity, the exact ranks for each grid cell differs. A main advantage of even short-term camera deployment was the spatial coverage of mammal data. While crowdsourced data sources provide data with larger temporal coverage, the datasets exhibit clustering. During our fieldwork period, trail cameras managed to capture pictures from more grid cells than either Flickr or GBIF. This has been previously observed too. Silveira et al. (2003) and Trofino-Falasco et al. (2023) documented that even when deployed for a short term, trail cameras offered a superior spatial coverage of mammal data compared to other methods such as line transects or crowdsourced data. Additionally, during our study trail cameras managed to record more mammal species (n = 15) than Flickr. However, the short deployment period and the season of deployment influenced the effectiveness of trail cameras in capturing the full range of species present in the sanctuary. The low number of recorded Big Four species and the insignificant relationship observed with ranger rankings for Big Four imply that wildlife dispersal during and after the wet season negatively affected the capture rate. Extending the temporal coverage of trail camera deployment to a minimum of one year, as done by Wachiye et al. (2022), would provide a more comprehensive view of species presence and richness patterns over time. Ultimately, our findings reinforce the complementary role that even short-term trail camera deployment plays in species monitoring, especially in areas that are less frequently visited by tourists. If no time constraints exist, trail cameras offer the best means to comprehensively map mammal distributions in most habitats (Silveira et al., 2003; Lyra-Jorge et al., 2008).

The decline in Kenya's wildlife populations is driven by a combination of habitat loss, rangeland degradation, human population growth, climate change, and policy failures (Ogutu and Owen-Smith, 2003; Scholte, 2011; Ogutu et al., 2012). While efforts have been made to curb poaching and reintroduce wildlife, the long-term survival of many species depends on addressing the root causes of these declines and promoting more efficient data-driven conservation methods (Western et al., 2009; Ogutu et al., 2016). However, wildlife monitoring is often constrained by limited resources and therefore selecting an appropriate and efficient data collection method is essential to maximize cost-effectiveness (Lyra-Jorge et al., 2008). Our results suggest that crowdsourced biodiversity datasets, particularly those derived from Flickr and GBIF, can serve as reliable indicators of species occurrence trends in THWS. The implications of these findings are important for conservation planning, particularly in regions where financial and logistical constraints limit the capacity for extensive field monitoring. Crowdsourced data offers a cost-effective way to gather large amounts of biodiversity information, particularly for charismatic well-known species. However, the analysis also underscores the importance of using a diverse array of datasets for comprehensive species occurrence assessments in conservation planning as different taxa were shown to correlate with different datasets. Additionally, the limitations of GBIF and Flickr, both in terms of taxonomic and spatial bias, mean that they should not be used in isolation. Integrating crowdsourced data with more systematic field methods, such as trail cameras and ranger-based knowledge, provides a more holistic approach to species monitoring. This study contributes to the growing body of research exploring alternative data sources in conservation science and provides a comprehensive overview of the reliability and utility of crowdsourced biodiversity datasets in predicting wildlife distribution and species richness in THWS. The results highlight the potential for integrating social media and DAK data into ongoing conservation efforts and wildlife monitoring strategies, helping to identify the most accurate and cost-effective data sources for wildlife conservation.

5.2. Challenges and future prospects

It is important to address the limitations faced during this study. Utilizing crowdsourced biodiversity datasets as proxies for species occurrence and richness raises several critical considerations, particularly concerning data quality and methodologies. Notably, our crowdsourced taxa sample sizes varied significantly, ranging from just tens of observations to over a thousand, suggesting taxonomic bias within these datasets. This finding is in line with previous research that has repeatedly noted taxonomic bias in both DAK data sources (Stahlschmidt, 2011; Sousa-Baena et al., 2014; Troudet et al., 2017) as well as social media platforms (Hausmann et al., 2017; Fink et al., 2020). Specifically, we found that GBIF had a disproportionately large number of bird records compared to other taxa, consistent with findings by Troudet et al. (2017) and Huang et al. (2020). In contrast, the Flickr dataset displayed a skewed distribution, with iconic Big Four species being over-represented. These disparities in taxa representation impact the reliability and usability of Flickr and GBIF datasets, necessitating careful interpretation of results derived from them. However, it is important to acknowledge that since this study focused on only three taxa groups, the information on taxonomic bias in GBIF and Flickr presented here is not comprehensive.

Addressing taxonomic bias in the study area would require targeted strategies to ensure a more balanced representation of biodiversity (Troudet et al., 2017). One approach is to develop specific data collection initiatives focused on underrepresented taxa, utilizing both traditional scientific research and citizen science programs (Martín-López et al., 2009; Newman et al., 2012). As noted by Newman et al. (2012), engaging the general public in documenting less charismatic species through outreach and education not only expands biodiversity data but also actively involves local communities in conservation initiatives. This involvement enhances the general acceptance of conservation actions among locals, often fostering greater support and effectiveness for these efforts (Newman et al., 2012; Troudet et al., 2017). Additionally, integrating multiple data sources into research can mitigate individual dataset's biases (Troudet et al., 2017; Phaka et al., 2022).

Moreover, this study found that both Flickr and GBIF datasets were clustered around the two lodges in the sanctuary. Studies by Hirvonen et al. (2020) and Toivonen et al. (2019) support this finding as both of these studies highlight crowdsourced data's tendency to be clustered around easily accessed, and WiFi-enabled, areas. This clustering likely skews the spatial representation of species occurrences, leaving more remote or less accessible areas underrepresented in the data. This presents a significant limitation for assessing species occurrences across the entire sanctuary, as areas located in the southern and eastern part of THWS may be overlooked. Such gaps in spatial coverage further highlight the importance of integrating diverse data sources, including ground-truth observations, to create a more comprehensive understanding of species occurrences within the sanctuary.

One critical consideration is the potential bias introduced by ranger-tourist interactions. As rangers guide tourists to known wildlife hotspots to optimize sightings, frequent visits are made to these well-established areas. This pattern is reflected in crowdsourced data, as many of the records in platforms like GBIF and Flickr are contributed by tourists. Consequently, the data tends to cluster around accessible and well-known wildlife observation sites, potentially skewing species distribution analyses as noted by Usui et al. (2014). Furthermore, this bias is particularly significant as it implies that crowdsourced datasets are not fully independent from ranger-based knowledge. Rather, ranger-based knowledge indirectly shapes the distribution of crowdsourced records, necessitating careful consideration when interpreting the results. Acknowledging this bias is crucial for ensuring the validity of results derived from these datasets.

To improve the quality and coverage of crowdsourced biodiversity data, it would be important to engage tourists more effectively in data collection efforts. Since tourism plays a significant role in THWS, encouraging visitors to contribute their wildlife observations could greatly enhance the amount and diversity of available data for conservation purposes. One strategy would be to implement citizen science programs that facilitate and encourage tourist participation (Zheng et al., 2024). Developing user-friendly mobile applications or collaborating with existing platforms like iNaturalist could make it easier for tourists to document and share their sightings. Furthermore, providing educational materials at lodges about the importance of comprehensive biodiversity monitoring, including the need to document a wider variety of species, can raise awareness and motivation (Leung et al., 2018). Such materials should emphasize the importance of recording not only charismatic megafauna like the Big Four, but also less-known or less-visible species, helping to combat taxonomic bias in crowdsourced datasets. By encouraging tourists to document a broader range of species, the quality and representativeness of biodiversity data can be significantly improved (Zheng et al., 2024). Additionally, as suggested by Leal et al. (2021), offering incentives such as recognition for contributions, certificates or featuring exceptional photographs in sanctuary publications can encourage more tourists to participate. Training guides and rangers to promote these initiatives would further enhance engagement by making tourists aware of how their observations aid conservation efforts (Zheng et al., 2024). By actively involving tourists in data collection, not only would the quantity and quality of biodiversity data improve, but tourists would also develop a deeper connection to the sanctuary's conservation goals. This increased awareness and involvement can create a sense of stewardship, leading to more responsible tourism practices and greater support for conservation initiatives (Leung et al., 2018).

The aggregation of spatial data into 3 by 3-kilometer grids presents another significant consideration. While aggregation facilitates the manageable analysis of large datasets and enhances comparability, it risks oversimplifying habitat use and species richness by potentially masking fine-scale habitat variations. This issue became particularly evident during the ranger workshop. Rangers often reported the suitability for each taxa group based on specific habitats, such as areas around rivers or water holes. In contrast, grid-based aggregation tends to generalize these observations, potentially overlooking the diversity of habitats. This discrepancy highlights the challenge of reconciling fine-scale, ground-truth data with broader, coarser-scale data.

Ground-truth mapping done with rangers also highlighted temporal dynamics as a key consideration. Seasonal variations, particularly between wet and dry seasons, significantly influence wildlife occurrence patterns, including those of migratory birds and mammals (Schmitz, 2008; Western et al., 2009). A study by Muteti and Maloba (2013) highlights that besides purely migratory species, many large-bodied mammals, such as elephants, are more common in the sanctuary during the dry season and disperse during the wet season. Indeed, during our workshop, rangers emphasized the need to provide distinct ranks for both dry and wet seasons to account for this temporal aspect. However, this study chose to exclude any temporal aspects from the crowdsourced datasets, focusing instead on purely spatial data as advised by previous research. While Edwards et al. (2021) advocated for the use of Flickr data in spatial species occurrence studies, they acknowledged the limitations of such data for capturing temporal variations. As a result, the seasonal dynamics were not captured in our analysis. Incorporating temporal data in future research would provide a more comprehensive understanding of species occurrence, movements and habitat use in regions where seasonality plays a critical role in shaping wildlife patterns.

Seasonality is also likely a significant factor contributing to the low capture rate of trail cameras in this study. While trail cameras provided complementary data, the short deployment duration and the specific season during which they were deployed must be considered. Previous studies have highlighted the effective use of trail cameras for species occurrence research but often have had a higher density of trail cameras (Glen et al., 2013; Assou et al., 2021) as well as longer deployment period (McCallum, 2013; Aramsirirujiwet and Duengkae, 2022). With only a three-month deployment period following the rainy season, our trail cameras did not effectively record species that exhibit seasonal variations in abundance. Additionally, following the unusually intensive rains, long grasses and weeds likely presented another limitation - visibility. With such high vegetation many smaller-bodied mammals go unrecorded, introducing taxonomic bias into our trail camera data. Indeed, our trail camera data may not accurately reflect the typical presence and abundance of species in the study area throughout the year. To improve the reliability of trail camera data in future studies, extending the deployment period to include multiple seasons would be beneficial. This approach would allow for the detection of seasonal patterns in species occurrence and abundance. Additionally, increasing the number and density of cameras across the study area could enhance detection rates, providing a more comprehensive understanding of species distributions and habitat use.

To our knowledge, no prior studies have assessed the predictive accuracy of crowdsourced biodiversity datasets from Flickr and GBIF in mapping wildlife occurrences and richness in southeastern Kenya. While crowdsourced datasets are increasingly utilized in other regions (Neves et al., 2018; Allain et al., 2019; Edwards et al., 2021), their application in East African savanna ecosystems remains underexplored. This study demonstrated that integrating crowdsourced data with ranger-based knowledge and trail camera data offers an effective complementary approach on discovering and monitoring species occurrence and richness patterns. The promising results show the potential for using such data to complement traditional field methods, especially in regions where wildlife attracts tourism. Importantly, we found that tailoring data sources to specific taxonomic groups led to stronger correlations with ground-truth data. Specifically, Flickr data was most effective for the Big Four species, while GBIF data yielded the most robust results for mammals overall. Although bird datasets did not correlate with ground-truth data, the vast number of bird records available on GBIF suggests that this data source could be valuable for further studies of the region's avifauna. Additionally, this approach supports the development of novel, data-based wildlife monitoring techniques that are especially valuable regarding the challenges posed by the declining wildlife numbers in Kenyan wilderness areas. By utilizing crowdsourced data, this study addresses a significant research gap and demonstrates the viability of alternative data sources in conservation science. However, further research is required to consider the limitations addressed in this study, including the geographical and taxonomic bias of crowdsourced datasets, the lack of temporal aspect of our analysis, as well as the limited temporal coverage of trail camera deployment.

In Kenya, wildlife conservation is not only crucial for preserving biodiversity but also vital for sustaining the country's economy and cultural heritage. Wildlife numbers across the country have declined drastically over the last decades, underscoring the urgent need for more effective, data-driven conservation strategies (Western et al., 2009; Ogutu et al., 2016). Crowdsourced data offers a promising, cost-effective method to complement traditional research efforts, especially in regions where resources for extensive field data collection are limited and tourism is prominent. Obtaining reliable and robust species occurrence data is essential for effectively establishing new protected areas and monitoring wildlife populations within existing conservation areas. Therefore, it is critically important to explore how well crowdsourced data aligns with ground-truth data for species occurrence and richness mapping. Results presented here provide valuable insights that support informed decision-making in the region's wildlife conservation.

Acknowledgements:

I would like to thank everyone who has been part of my thesis and made this journey possible. First and foremost, I would like to express my gratitude to my supervisors, Enrico Di Minin and Miquel Torrents-Ticó, for their guidance, encouragement and feedback throughout the course of my research. Their support and expertise have been fundamental in shaping this thesis.

I am also deeply thankful to all the staff at the Taita Hills Research Station in Wundanyi, Kenya. Their assistance made my research possible, and their warm hospitality made my time there truly memorable. I owe special thanks to everyone who assisted me in the fieldwork, including Darius Kimuzi, Peter Mwasi and Enrico Di Minin.

A special thank you goes to the rangers at the Taita Hills Wildlife Sanctuary, whose knowledge of the region and their commitment to conservation were invaluable to my fieldwork. Their support and passion for protecting the local ecosystem inspired and motivated me throughout my research.

Finally, I extend my heartfelt thanks to my family, friends, and peers for their constant encouragement and belief in my work. Their support has been a source of strength during this journey.

Appendix:

Table A1. List of bird species recorded by GBIF and Flickr in the study area.

Species scientific name	Platform recorded
Accipiter badius	GBIF
Actitis hypoleucos	GBIF
Actophilornis africanus	GBIF
Alopochen aegyptiaca	GBIF
Amadina fasciata	GBIF
Anas erythrorhyncha	GBIF
Andropadus importunus	GBIF
Anthoscopus musculus	GBIF
Anthreptes orientalis	GBIF
Anthus caffer	GBIF
Anthus trivialis	GBIF
Apalis flavida	GBIF
Apus affinis	GBIF
Apus apus	GBIF
Aquila nipalensis	GBIF
Aquila rapax	GBIF
Aquila spilogaster	Flickr
Ardea alba	GBIF
Ardea cinerea	GBIF
Ardea melanocephala	GBIF
Ardeola ralloides	GBIF
Ardeotis kori	GBIF & Flickr
Asio capensis	GBIF
Batis molitor	GBIF
Batis perkeo	GBIF
Bostrychia hagedash	GBIF
Bradornis microrhynchus	GBIF
Bradornis pallidus	GBIF
Bubalornis niger	GBIF
Bubo africanus	GBIF
Bubo capensis	GBIF
Bubo lacteus	GBIF
Bubulcus ibis	GBIF
Bucorvus abyssinicus	GBIF
Bucorvus leadbeateri	GBIF & Flickr
Buphagus africanus	GBIF
Buphagus erythrorhynchus	GBIF

Burhinus capensis	GBIF
Buteo augur	GBIF
Buteo buteo	GBIF
Calamonastes simplex	GBIF
Calendulauda africanoides	GBIF
Calendulauda poecilosterna	GBIF
Calidris minuta	GBIF
Camaroptera brachyura	GBIF
Campephaga flava	GBIF
Campethera cailliautii	GBIF
Campethera nubica	GBIF
Caprimulgus clarus	GBIF
Caprimulgus fraenatus	GBIF
Cecropis abyssinica	GBIF
Centropus superciliosus	GBIF
Ceryle rudis	GBIF
Chalcomitra amethystina	GBIF
Chalcomitra hunteri	GBIF
Charadrius tricollaris	GBIF
Chlorophoneus sulfureopectus	GBIF
Chloropicus fuscescens	GBIF
Chloropicus namaquus	GBIF
Ciconia abdimii	GBIF
Ciconia ciconia	GBIF
Ciconia episcopus	GBIF
Ciconia microscelis	GBIF
Ciconia nigra	GBIF
Cinnyris tsavoensis	GBIF
Cinnyris venustus	GBIF
Circaetus cinereus	GBIF
Circaetus pectoralis	GBIF
Circus macrourus	GBIF
Circus pygargus	GBIF
Cisticola aridulus	GBIF
Cisticola chiniana	GBIF
Cisticola cinereolus	GBIF
Cisticola marginatus	GBIF
Cisticola nana	GBIF
Clamator jacobinus	GBIF
Colius striatus	GBIF
Columba guinea	GBIF
Coracias caudatus	GBIF
Coracias garrulus	GBIF & Flickr

Coracias naevius	GBIF
Corvus albicollis	GBIF
Corvus albus	GBIF
Corvus splendens	GBIF
Corythaixoides concolor	GBIF
Corythaixoides leucogaster	GBIF
Corythornis cristatus	GBIF
Cossypha heuglini	GBIF
Creatophora cinerea	GBIF
Crithagra dorsostriata	GBIF
Crithagra reichenowi	GBIF
Cuculus canorus	GBIF
Cuculus solitarius	GBIF
Cursorius temminckii	GBIF
Cypsiurus parvus	GBIF
Dendrocygna bicolor	GBIF
Dendrocygna viduata	GBIF
Dendropicos fuscescens	GBIF
Dicrurus adsimilis	GBIF
Dinemellia dinemelli	GBIF
Dryoscopus cubla	GBIF
Egretta garzetta	GBIF
Egretta intermedia	GBIF
Elanus caeruleus	GBIF
Ephippiorhynchus senegalensis	GBIF
Eremopterix leucopareia	GBIF
Erythropygia galactotes	GBIF
Euodice cantans	GBIF
Euplectes albonotatus	GBIF
Eupodotis senegalensis	GBIF
Eurocephalus anguitimens	GBIF
Eurocephalus ruppelli	GBIF
Falco biarmicus	GBIF
Falco naumanni	GBIF
Falco peregrinus	GBIF
Falco tinnunculus	GBIF
Gallinula chloropus	GBIF
Glaucidium perlatum	GBIF
Guttera pucherani	GBIF
Gymnoris pyrgita	GBIF
Gypohierax angolensis	GBIF
Gyps africanus	GBIF & Flickr
Gyps fulvus	GBIF

Gyps rueppellii	GBIF
Halcyon albiventris	GBIF
Halcyon chelicuti	GBIF
Halcyon leucocephala	GBIF
Haliaeetus vocifer	GBIF
Hieraaetus wahlbergi	GBIF
Himantopus himantopus	GBIF
Hirundo linnaeus	GBIF
Hirundo rustica	GBIF
Indicator indicator	GBIF
Indicator minor	GBIF
Kaupifalco monogrammicus	GBIF
Lagonosticta senegala	GBIF
Lamprotornis chalybaeus	GBIF
Lamprotornis fischeri	GBIF
Lamprotornis hildebrandti	GBIF
Lamprotornis purpuroptera	GBIF
Lamprotornis regius	GBIF
Lamprotornis shelleyi	GBIF
Lamprotornis superbus	GBIF & Flickr
Laniarius funebris	GBIF
Lanius cabanisi	GBIF & Flickr
Lanius collurio	GBIF
Lanius dorsalis	GBIF
Lanius excubitoroides	GBIF
Lanius humeralis	GBIF
Lanius isabellinus	GBIF & Flickr
Lanius minor	GBIF
Leptoptilos crumenifer	GBIF & Flickr
Lissotis hartlaubii	GBIF
Lissotis melanogaster	GBIF
Lophaetus occipitalis	GBIF
Lophoceros alboterminatus	GBIF
Lophoceros nasutus	GBIF
Lophotis gindiana	GBIF
Macronyx ameliae	GBIF
Macronyx aurantiigula	GBIF & Flickr
Megaceryle maxima	GBIF
Melaenornis pammelaina	GBIF
Melierax canorus	GBIF
Melierax poliopterus	GBIF
Merops apiaster	GBIF
Merops persicus	GBIF

Merops pusillus	GBIF & Flickr
Microcarbo africanus	GBIF
Micronisus gabar	GBIF
Mirafra hypermetra	GBIF
Mirafra javanica	GBIF
Mirafra rufocinnamomea	GBIF
Monticola saxatilis	GBIF
Motacilla aguimp	GBIF
Motacilla flava	GBIF
Muscicapa adusta	GBIF
Muscicapa striata	GBIF
Mycteria ibis	GBIF
Necrosyrtes monachus	GBIF
Neotis heuglinii	GBIF
Nilaus afer	GBIF
Notopholia corrusca	GBIF
Numida meleagris	GBIF & Flickr
Oena capensis	GBIF
Oenanthe isabellina	GBIF
Oenanthe oenanthe	GBIF
Onychognathus morio	GBIF
Oriolus larvatus	GBIF
Oriolus oriolus	GBIF
Ortygornis sephaena	GBIF
Parus albiventris	GBIF
Parus thruppi	GBIF
Passer diffusus	GBIF
Passer gongonensis	GBIF
Phoeniculus damarensis	GBIF
Phoeniculus purpureus	GBIF
Platalea alba	GBIF
Platalea leucorodia	GBIF
Plectropterus gambensis	GBIF
Plegadis falcinellus	GBIF
Plocepasser mahali	GBIF
Ploceus cucullatus	GBIF
Ploceus rubiginosus	GBIF
Ploceus subaureus	GBIF
Ploceus vitellinus	GBIF
Pogoniulus pusillus	GBIF
Poicephalus cryptoxanthus	GBIF
Poicephalus rufiventris	GBIF
Polemaetus bellicosus	GBIF & Flickr

Polihierax semitorquatus	GBIF
Porphyrio porphyrio	GBIF
Prinia subflava	GBIF
Prionops plumatus	GBIF
Prionops retzii	GBIF
Pseudonigrita cabanisi	GBIF
Pternistis afer	GBIF
Pternistis leucoscepus	GBIF & Flickr
Pterocles decoratus	GBIF
Pterocles exustus	GBIF
Pycnonotus barbatus	GBIF
Pycnonotus tricolor	GBIF
Quelea cardinalis	GBIF
Quelea quelea	GBIF
Rhinopomastus cyanomelas	GBIF
Riparia cincta	GBIF
Sagittarius serpentarius	GBIF & Flickr
Sarkidiornis melanotos	GBIF
Scleroptila shelleyi	GBIF
Scopus umbretta	GBIF
Spilopelia senegalensis	GBIF
Streptopelia capicola	GBIF
Streptopelia decipiens	GBIF
Streptopelia semitorquata	GBIF
Struthio camelus	GBIF & Flickr
Sylvia atricapilla	GBIF
Sylvia boehmi	GBIF
Sylvia communis	GBIF
Sylvietta brachyura	GBIF
Sylvietta whytii	GBIF
Tachybaptus ruficollis	GBIF
Tchagra jamesi	GBIF
Telophorus cruentus	GBIF
Terathopius ecaudatus	GBIF
Terpsiphone viridis	GBIF
Threskiornis aethiopicus	GBIF
Tmetothylacus tenellus	GBIF
Tockus deckeni	GBIF
Tockus erythrorhynchus	GBIF
Tockus flavirostris	GBIF
Tockus rufirostris	GBIF
Torgos tracheliotos	GBIF & Flickr
Trachyphonus darnaudii	GBIF

Trachyphonus erythrocephalus	GBIF & Flickr
Tricholaema lacrymosa	GBIF
Tricholaema melanocephala	GBIF
Trigonoceps occipitalis	GBIF
Tringa glareola	GBIF
Tringa nebularia	GBIF
Turdoides aylmeri	GBIF
Turdus tephronotus	GBIF
Turtur chalcospilos	GBIF
Upupa africana	GBIF
Upupa epops	GBIF
Uraeginthus bengalus	GBIF
Uraeginthus ianthinogaster	GBIF
Urocolius macrourus	GBIF
Vanellus armatus	GBIF
Vanellus coronatus	GBIF
Vanellus spinosus	GBIF
Vidua fischeri	GBIF
Vidua macroura	GBIF
Vidua paradisaea	GBIF

References:

Agha, M., Batter, T., Bolas, E. C., Collins, A. C., Gomes da Rocha, D., Monteza-Moreno, C. M. & Sollmann, R. (2018). A review of wildlife camera trapping trends across Africa. African Journal of Ecology, 56(4), 694-701. https://doi.org/10.1111/aje.12565

Allain, S. J. (2019). Mining Flickr: a method for expanding the known distribution of invasive species. Herpetological Bulletin, 148(Summer 2019), 11-4. https://doi.org/10.33256/hb148.1114

Amara, E., Adhikari, H., Heiskanen, J., Siljander, M., Munyao, M., Omondi, P., Pellikka, P. (2020). Aboveground Biomass Distribution in a Multi-Use Savannah Landscape in Southeastern Kenya: Impact of Land Use and Fences. Land. 2020; 9(10):381. https://doi.org/10.3390/land9100381

Amara, E., Adhikari, H., Mwamodenyi, J. M., Pellikka, P. K., & Heiskanen, J. (2023). Contribution of tree size and species on aboveground biomass across land cover types in the Taita Hills, Southern Kenya. Forests, 14(3), 642. https://doi.org/10.3390/f14030642

Angus, E., Stuart, D., & Thelwall, M. (2010). Flickr's potential as an academic image resource: an exploratory study. Journal of Librarianship and Information Science, 42(4), 268-278. https://doi.org/10.1177/0961000610384656

Aramsirirujiwet, P., & Duengkae, P. (2022). Species, Abundance, and Activity Pattern of Large Mammals using Camera Trap around Managed Grassland in Thap Lan National Park (Doctoral dissertation, Kasetsart University).

Arponen, A. (2012). Prioritizing species for conservation planning. Biodiversity and Conservation, 21, 875-893. https://doi.org/10.1007/s10531-012-0242-1

Assou, D., D'Cruze, N., Kirkland, H., Auliya, M., Macdonald, D. W., & Segniagbeto, G. H. (2021). Camera trap survey of mammals in the Fazao-Malfakassa National

Park, Togo, West Africa. African Journal of Ecology, 59(3), 583-596. https://doi.org/10.1111/aje.12856

Autio, A., Johansson, T., Motaroki, L., Minoia, P., & Pellikka, P. (2021). Constraints for adopting climate-smart agricultural practices among smallholder farmers in Southeast Kenya. Agricultural Systems, 194, 103284. https://doi.org/10.1016/j.agsy.2021.103284.

Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R. & Guralnick, R. P. (2019). Research applications of primary biodiversity databases in the digital age. PloS one, 14(9), e0215794. https://doi.org/10.1371/journal.pone.0215794 (2019)

Barve, N., Kiran, M. C., Vanaraj, G., Aravind, N. A., Rao, D., Shaanker, R. U. & Poulsen, J. G. (2005). Measuring and mapping threats to a wildlife sanctuary in southern India. Conservation Biology, 19(1), 122-130. https://doi.org/10.1111/j.1523-1739.2005.00532.x

Beck, J., Ballesteros-Mejia, L., Nagel, P., & Kitching, I. J. (2013). Online solutions and the 'Wallacean shortfall': what does GBIF contribute to our knowledge of species' ranges?. Diversity and Distributions, 19(8), 1043-1050. https://doi.org/10.1111/ddi.12083

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecological Informatics, 19, 10-15. https://doi.org/10.1016/j.ecoinf.2013.11.002

Braczkowski, A. R., Balme, G. A., Dickman, A., Fattebert, J., Johnson, P., Dickerson, T. & Hunter, L. (2016). Scent lure effect on camera-trap based leopard density estimates. PloS one, 11(4), e0151033. https://doi.org/10.1371/journal.pone.0151033

Braczkowski, A. R., Namukose, L., Musobozi, S., Cornille, O., Mudumba, T., Drileyo, G., ... & Elliot, N. (2024). Rangers on the frontline of wildlife monitoring: a case study on African lions in Uganda's Nile Delta. *Communications Biology*, *7*(1), 1308. https://doi.org/10.1038/s42003-024-06796-0

Boitani, L., Maiorano, L., Baisero, D., Falcucci, A., Visconti, P., & Rondinini, C. (2011). What spatial data do we need to develop global mammal conservation strategies?. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1578), 2623-2632. https://doi.org/10.1098/rstb.2011.0117

Bolognini, L., & Bistolfi, C. (2017). Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation. Computer law & security review, 33(2), 171-181. https://doi.org/10.1016/j.clsr.2016.11.002

Bond, W.J. and Loffell, D. (2001), Introduction of giraffe changes acacia distribution in a South African savanna. African Journal of Ecology, 39: 286-294. https://doi.org/10.1046/j.1365-2028.2001.00319.x

Bonnet, X., Shine, R., & Lourdais, O. (2002). Taxonomic chauvinism. Trends in Ecology & Evolution, 17(1), 1-3. https://doi.org/10.1016/S0169-5347(01)02381-3

Brassine, E., & Parker, D. (2015). Trapping elusive cats: using intensive camera trapping to estimate the density of a rare African felid. PloS one, 10(12), e0142508. https://doi.org/10.1371/journal.pone.0142508

Brink, A. B., & Eva, H. D. (2009). Monitoring 25 years of land cover change dynamics in Africa: A sample based remote sensing approach. Applied geography, 29(4), 501-512. https://doi.org/10.1016/j.apgeog.2008.10.004

Brink, A. B., Bodart, C., Brodsky, L., Defourney, P., Ernst, C., Donney, F. & Tuckova, K. (2014). Anthropogenic pressure in East Africa—Monitoring 20 years of land cover changes by means of medium resolution satellite data. International

Journal of Applied Earth Observation and Geoinformation, 28, 60-69. https://doi.org/10.1016/j.jag.2013.11.006.

Bruce, T., Amin, R., Wacher, T., Fankem, O., Ndjassi, C., Ngo Bata, M. & Olson, D. (2018). Using camera trap data to characterise terrestrial larger-bodied mammal communities in different management sectors of the Dja Faunal Reserve, Cameroon. African Journal of Ecology, 56(4), 759-776. https://doi.org/10.1111/aje.12574

Callcutt, K., Croft, S., & Smith, G. C. (2018). Predicting population trends using citizen science data: do subsampling methods produce reliable estimates for mammals?. European Journal of Wildlife Research, 64, 1-6. https://doi.org/10.1007/s10344-018-1189-7

Caro, T., & Scholte, P. (2007). When protection falters. African Journal of Ecology, 45(3), 233-235. https://doi.org/10.1111/j.1365-2028.2007.00814.x

Caro T. (2008). Decline of large mammals in the Katavi-Rukwa ecosystem of western Tanzania. African Zoology 43: 99–116. https://doi.org/10.1080/15627020.2008.11407412

Chamberlain, S. A., & Boettiger, C. (2017). R Python, and Ruby clients for GBIF species occurrence data (No. e3304v1). PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.3304v1

Chapman, S., & Balme, G. (2010). An estimate of leopard population density in a private reserve in KwaZulu-Natal, South Africa, using camera-traps and capture-recapture models. South African Journal of Wildlife Research-24-month delayed open access, 40(2), 114-120. https://hdl.handle.net/10520/EJC117351

Chen, J., Yang, S. T., Li, H. W., Zhang, B., & Lv, J. R. (2013). Research on geographical environment unit division based on the method of natural breaks (Jenks). *The International Archives of the Photogrammetry, Remote*

Sensing and Spatial Information Sciences, *40*, 47-50. https://doi.org/10.5194/isprsarchives-XL-4-W3-47-2013

Chen, C., Brodie, J. F., Kays, R., Davies, T. J., Liu, R., Fisher, J. T. & Burton, A. C. (2022). Global camera trap synthesis highlights the importance of protected areas in maintaining mammal diversity. Conservation Letters, 15(2), e12865. https://doi.org/10.1111/conl.12865

Clark, J. A., & May, R. M. (2002). How biased are we?: Even now, conservation research is still lopsided. Conservation in Practice, 3(3), 28-29.

Craigie, I. D., Baillie, J. E., Balmford, A., Carbone, C., Collen, B., Green, R. E., & Hutton, J. M. (2010). Large mammal population declines in Africa's protected areas. Biological conservation, 143(9), 2221-2228. https://doi.org/10.1016/j.biocon.2010.06.007

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. Cartography and geographic information science, 40(2), 130-139. https://doi.org/10.1080/15230406.2013.777137

Coetzer, W. (2012). A new era for specimen databases and biodiversity information management in South Africa. Biodiversity Informatics, 8. https://doi.org/10.17161/bi.v8i1.4263

Cordier, C. P., Smith, D. A. E., Smith, Y. E., & Downs, C. T. (2022). Camera trap research in Africa: A systematic review to show trends in wildlife monitoring and its value as a research tool. Global Ecology and Conservation, 40, e02326. https://doi.org/10.1016/j.gecco.2022.e02326

De Araujo, M. L., Quaresma, A. C., & Ramos, F. N. (2022). GBIF information is not enough: national database improves the inventory completeness of Amazonian epiphytes. Biodiversity and Conservation, 31(11), 2797-2815. https://doi.org/10.1007/s10531-022-02458-x

De Leeuw, J., Prins, H. H. T., Njuguna, E. C., Said, M. Y., & De By, R. (1998). Interpretation of DRSRS Animal Counts (1977-1997) in the Rangeland Districts of Kenya.

De Vos, J.M., Joppa, L.N., Gittleman, J.L., Stephens, P.R. and Pimm, S.L. (2015). Estimating the normal background rate of species extinction. Conservation Biology, 29: 452-462. https://doi.org/10.1111/cobi.12380

Díaz, S., Settele, J., Brondízio, E., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S., Chan, K., Garibaldi, L., Ichii, K., Liu, J., Subramanian, S., Midgley, G., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., & Zayas, C. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. Science (New York, N.Y.), 366. https://doi.org/10.1126/science.aax3100

Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., & Hurlbert, A. H. (2021). Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. Bioscience, 71(11), 1179–1188. https://doi.org/10.1093/biosci/biab093

Di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N. & Watson, J. E. (2017). Changing trends and persisting biases in three decades of conservation science. Global Ecology and Conservation, 10, 32-42. https://doi.org/10.1016/j.gecco.2017.01.008

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021a). How to address data privacy concerns when using social media data in conservation science. Conservation Biology, 35(2), 437–446. https://doi.org/10.1111/cobi.13708

Di Minin, E., Correia, R. A., & Toivonen, T. (2021b). Quantitative conservation geography. Trends in Ecology & Evolution, 37(1), 42-52. https://doi.org/10.1016/j.tree.2021.08.009

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. Frontiers in Environmental Science, 3, 1–6. https://doi.org/10.3389/fenvs.2015.00063

Di Minin, E., Macmillan, D. C., Goodman, P. S., Escott, B., Slotow, R., & Moilanen, A. (2013). Conservation Businesses and Conservation Planning in a Biological Diversity Hotspot: Conservation Businesses and Planning. Conservation Biology, 27(4), 808–820. https://doi.org/10.1111/cobi.12048

Dickens, O., Faith, K., Geoffrey, M., Petri, P., & Rob, M. (2018). Comparison and Prediction of the above Ground Carbon Storage in Croplands on the Inhabited Slopes on Mount Kilimanjaro (Tanzania) and the Taita Hills (Kenya). Journal of Geographic Information System, 10(4), 415-438. https://doi.org/10.4236/jgis.2018.104022

Du Toit, J. T., & Cumming, D. H. (1999). Functional significance of ungulate diversity in African savannas and the ecological implications of the spread of pastoralism. Biodiversity & Conservation, 8, 1643-1661. https://doi.org/10.1023/A:1008959721342

Dupuis-Desormeaux, M., Davidson, Z., Mwololo, M., Kisio, E., & MacDonald, S. E. (2016). Comparing motion capture cameras versus human observer monitoring of mammal movement through fence gaps: a case study from Kenya. African Journal of Ecology, 54(2), 154-161. https://doi.org/10.1111/aje.12277

Edwards, T., Jones, C. B., Perkins, S. E., & Corcoran, P. (2021). Passive citizen science: The role of social media in wildlife observations. PLoS One, 16(8), e0255416. https://doi.org/10.1371/journal.pone.0255416

Egna, N., O'Connor, D., Stacy-Dawes, J., Tobler, M. W., Pilfold, N., Neilson, K. & Owen, M. (2020). Camera settings and biome influence the accuracy of citizen science approaches to camera trap image classification. Ecology and evolution, 10(21), 11954-11965. https://doi.org/10.1002/ece3.6722

Escribano, N., Galicia, D., & Ariño, A. H. (2019). Completeness of Digital Accessible Knowledge (DAK) about terrestrial mammals in the Iberian Peninsula. PloS one, 14(3), e0213542. https://doi.org/10.1371/journal.pone.0213542

Farooq, H., Azevedo, J. A., Soares, A., Antonelli, A., & Faurby, S. (2021). Mapping Africa's biodiversity: More of the same is just not good enough. Systematic Biology, 70(3), 623-633. https://doi.org/10.1093/sysbio/syaa090

Fink, C., Hausmann, A., Di Minin, E. (2020). Online sentiment towards iconic species. https://doi.org/10.1016/j.biocon.2019.108289

Fink, C., Toivonen, T. , Correia, R., Di Minin, E. (2021). Mapping the online songbird trade in Indonesia. https://doi.org/10.1016/j.apgeog.2021.102505

Flather, C. H., & King, R. M. (1992). Evaluating performance of regional wildlife habitat models: implications to resource planning. Journal of Environmental Management, 34(1), 31-46. https://doi.org/10.1016/S0301-4797(05)80108-2

Ford, A. T., Cooke, S. J., Goheen, J. R., & Young, T. P. (2017). Conserving megafauna or sacrificing biodiversity?. BioScience, 67(3), 193-196. https://doi.org/10.1093/biosci/biw163

Freeman, B., & Peterson, A. T. (2019). Completeness of digital accessible knowledge of the birds of western Africa: Priorities for survey. The Condor, 121(3), duz035. https://doi.org/10.1093/condor/duz035

Funk, C. C. (2012). Exceptional warming in the western Pacific-Indian Ocean warm pool has contributed to more frequent droughts in eastern Africa. Bulletin of the American Meteorological Society, 93, 1049-1051. https://doi.org/10.1175/BAMS-D-12-00021.1

Gaiji, S., Chavan, V., Ariño, A. H., Otegui, J., Hobern, D., Sood, R., & Robles, E. (2013). Content assessment of the primary biodiversity data published through GBIF network: status, challenges and potentials. Biodiversity Informatics, 8(2). https://doi.org/10.17161/bi.v8i2.4124

Ganglo, J. C. (2023). Will Africa meet the challenge of biodiversity conservation in the digital age?. https://doi.org/10.21203/rs.3.rs-2182259/v5

Glen, A. S., Cockburn, S., Nichols, M., Ekanayake, J., & Warburton, B. (2013). Optimising camera traps for monitoring small mammals. PloS one, 8(6), e67940. https://doi.org/10.1371/journal.pone.0067940

Glover-Kapfer, P., Soto-Navarro, C. A., & Wearn, O. R. (2019). Camera-trapping version 3.0: current constraints and future priorities for development. Remote Sensing in Ecology and Conservation, 5(3), 209-223. https://doi.org/10.1002/rse2.106

Gudka, M. S. (2020). Kenya national biodiversity threat assessment: Direct human threats impacting Kenya's biodiversity. IUCN–BIODEV2030, Gland, Switzerland. https://www.biodev2030.org/wp-content/uploads/sites/2/2021/07/Annexe-33_Rapport-Final_National-Biodiversity-Threat-Assessment_Kenya.pdf

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2017). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. Conservation Letters, 11(1), e12343. https://doi.org/10.1111/conl.12343

Hernández-Quiroz, N.S., Badano, E.I., Barragán-Torres, F., Flores, J. (2018). Habitat suitability models to make conservation decisions based on areas of high species richness and endemism. Biodivers Conserv 27, 3185–3200. https://doi.org/10.1007/s10531-018-1596-9

Hintze, M., & El Emam, K. (2018). Comparing the benefits of pseudonymisation and anonymisation under the GDPR. Journal of Data Protection & Privacy, 2(2), 145-158.

Hirvonen, H., Leppämäki, T., Rinne, J. J., Muukkonen, P., & Fink, C. (2020). Modifying and analyzing Flickr data for wildlife conservation. In Examples and progress in geodata science (pp. 66-90). University of Helsinki, Faculty of Science, Department of Geosciences and Geography. http://hdl.handle.net/10138/326542

Hjarding, A., Tolley, K. A., & Burgess, N. D. (2015). Red List assessments of East African chameleons: a case study of why we need experts. Oryx, 49(4), 652-658. doi:10.1017/S0030605313001427

Huang, X., Lin, C., & Ji, L. (2020). The persistent multi-dimensional biases of biodiversity digital accessible knowledge of birds in China. Biodiversity and Conservation, 29(11), 3287-3311. https://doi.org/10.1007/s10531-020-02024-3

Idohou, R., Arino, A., Assogbadjo, A., Kakai, R. G., & Sinsin, B. (2015). Diversity of wild palms (Arecaceae) in the Republic of Benin: Finding the gaps in the national inventory combining field and digital accessible knowledge.

IUCN. (2020). ESA SoPACA Country Profile - Kenya. https://esahub-rris.rcmrd.org/sites/default/files/2020-12/IUCN2020%20ESA%20SoPACA%20Country%20Profile%20-%20Kenya.pdf

Ivanova, N. V., & Shashkov, M. P. (2021). The possibilities of GBIF data use in ecological research. Russian Journal of Ecology, 52(1), 1-8. https://doi.org/10.1134/S1067413621010069

Jaureguiberry, P., Titeux, N., Wiemers, M., Bowler, D. E., Coscieme, L., Golden, A. S. & Purvis, A. (2022). The direct drivers of recent global anthropogenic biodiversity loss. *Science advances*, *8*(45), eabm9982. https://doi.org/10.1126/sciadv.abm9982

Joppa, L. N., Visconti, P., Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2016). Filling in biodiversity threat gaps. Science, 352(6284), 416–418. https://doi.org/10.1126/science.aaf3565

Jori, F., Brahmbhatt, D., Fosgate, G. T., Thompson, P. N., Budke, C., Ward, M. P. & Gummow, B. (2011). A questionnaire-based evaluation of the veterinary cordon fence separating wildlife and livestock along the boundary of the Kruger National Park, South Africa. Preventive veterinary medicine, 100(3-4), 210-220. https://doi.org/10.1016/j.prevetmed.2011.03.015

Kays, R., Lasky, M., Allen, M. L., Dowler, R. C., Hawkins, M. T., Hope, A. G. & Cove, M. V. (2022). Which mammals can be identified from camera traps and crowdsourced photographs?. Journal of Mammalogy, 103(4), 767-775. https://doi.org/10.1093/jmammal/gyac021

Kent, R., Bar-Massada, A., & Carmel, Y. (2014). Bird and mammal species composition in distinct geographic regions and their relationships with environmental factors across multiple spatial scales. Ecology and evolution, 4(10), 1963-1971. https://doi.org/10.1002/ece3.1072

Kenya meteorological department. (2023). ANALYSIS OF EL NIÑO INFLUENCE ON WEATHER AND CLIMATE IN KENYA: A BRIEF UPDATE. https://meteo.go.ke/node/15860

Kenya Tourism Board. (2019). Tourism Performance Report. Retrieved from https://tri.go.ke/wp-content/uploads/2023/12/TOURISM-SECTOR-PERFORMANCE-REPORT-2019.pdf

Knight, A. T., Cowling, R. M., & Campbell, B. M. (2006). An operational model for implementing conservation action. Conservation biology, 20(2), 408-419. https://doi.org/10.1111/j.1523-1739.2006.00305.x

Kremen, C. and Merenlender, A. (2018). Landscapes that work for biodiversity and people. http://dx.doi. org/10.1126/ science.aau6020

Kujala, H., Lahoz-Monfort, J. J., Elith, J., & Moilanen, A. (2018). Not all data are equal: Influence of data type and amount in spatial conservation prioritisation. Methods in Ecology and Evolution, 9(11), 2249-2261. https://doi.org/10.1111/2041-210X.13084

Kukkala, A. and Moilanen, A. (2013). Core concepts of spatial prioritisation in systematic conservation planning. Biol Rev, 88: 443-464. https://doi.org/10.1111/brv.12008

Leal, F., Veloso, B., Malheiro, B., & Burguillo, J. C. (2021). Crowdsourced data stream mining for tourism recommendation. In *World Conference on Information Systems and Technologies* (pp. 260-269). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-72657-7_25

Leather, S. R. (2009). Taxonomic chauvinism threatens the future of entomology. Biologist, 56(1), 10-13. (PDF). https://cdn.harper-adams.ac.uk/document/profile/Leather_Biologist_2009.pdf

Lehmann, C. E., Archibald, S. A., Hoffmann, W. A., & Bond, W. J. (2011). Deciphering the distribution of the savanna biome. New Phytologist, 191(1), 197-209. https://doi.org/10.1111/j.1469-8137.2011.03689.x

Leung, Y. F., Spenceley, A., Hvenegaard, G., Buckley, R., & Groves, C. (2018). Tourism and visitor management in protected areas: Guidelines for sustainability (Vol. 27). Gland, Switzerland: IUCN.

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and geographic information science, 40(2), 61-77. https://doi.org/10.1080/15230406.2013.777139

Li, Q., Zhou, D., Jin, Y., Wang, M., Song, Y., & Li, G. (2014). Effects of fencing on vegetation and soil restoration in a degraded alkaline grassland in northeast China. Journal of Arid Land, 6, 478-487.

https://doi.org/10.1007/s40333-013-0207-6

Long, H., Mojo, D., Fu, C., Wang, G., Kanga, E., Oduor, A. M. O., & Zhang, L. (2019). Patterns of human-wildlife conflict and management implications in Kenya: a national perspective. Human Dimensions of Wildlife, 25(2), 121–135. https://doi.org/10.1080/10871209.2019.1695984

Lyon, B., & DeWitt, D. G. (2012). A recent and abrupt decline in the East African long rains. Geophysical Research Letters, 39(2). https://doi.org/10.1029/2011GL050337

Lyra-Jorge, M. C., Ciocheti, G., Pivello, V. R., & Meirelles, S. T. (2008). Comparing methods for sampling large-and medium-sized mammals: camera traps and track plots. European Journal of Wildlife Research, 54, 739-744. https://doi.org/10.1007/s10344-008-0205-8

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. Nature, 405(6783), 243-253. https://doi.org/10.1038/35012251

Martín-López, B., Montes, C., Ramírez, L., & Benayas, J. (2009). What drives policy decision-making related to species conservation?. Biological Conservation, 142(7), 1370-1380. https://doi.org/10.1016/j.biocon.2009.01.030

Massey, A. L., King, A. A., & Foufopoulos, J. (2014). Fencing protected areas: A long-term assessment of the effects of reserve establishment and fencing on African mammalian diversity. Biological Conservation, 176, 162-171. https://doi.org/10.1016/j.biocon.2014.05.023.

McCallum, J. (2013). Changing use of camera traps in mammalian field research: habitats, taxa and study types. Mammal Review, 43(3), 196-206. https://doi.org/10.1111/j.1365-2907.2012.00216.x

McKinney, M. L. (1999). High rates of extinction and threat in poorly studied taxa. Conservation biology, 13(6), 1273-1281. https://doi.org/10.1046/j.1523-1739.1999.97393.x

Mesaglio, T., Callaghan, C. T., Samonte, F., Gorta, S. B., & Cornwell, W. K. (2023). Recognition and completeness: two key metrics for judging the utility of citizen science data. Frontiers in Ecology and the Environment, 21(4), 167-174. https://doi.org/10.1002/fee.2604

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. Nature communications, 6(1), 1-8. https://doi.org/10.1038/ncomms9221

Ministry of Tourism and Wildlife, Kenya. (2022). New tourism strategy for Kenya 2021-2025. [PDF]. https://www.tourism.go.ke/wp-content/uploads/2022/10/New-Tourism-Strategy-for-Kenya-2021-2025.pdf

Moolman, L., De Morney, M. A., Ferreira, S. M., Ganswindt, A., Poole, J. H., & Kerley, G. I. (2019). And Then There Was One: a camera trap survey of the declining population of African Elephants in Knysna, South Africa. African Journal of Wildlife Research, 49(1), 16-26. https://hdl.handle.net/10520/EJC-130f909485

Moreno, I., Gippet, J. M. W., Fumagalli, L., & Stephenson, P. J. (2023). Factors affecting the availability of data on East African wildlife: the monitoring needs of conservationists are not being met. Biodiversity and Conservation, 32(1), 249-273. https://doi.org/10.1007/s10531-022-02497-4

Msoffe, F., Mturi, F. A., Galanti, V., Tosi, W., Wauters, L. A., & Tosi, G. (2007). Comparing data of different survey methods for sustainable wildlife management in hunting areas: the case of Tarangire–Manyara ecosystem, northern Tanzania. European Journal of Wildlife Research, 53, 112-124. https://doi.org/10.1007/s10344-006-0078-7

Msoffe, F. U., Kifugo, S. C., Said, M. Y., Neselle, M. O., Van Gardingen, P., Reid, R. S. & De Leeuw, J. (2011). Drivers and impacts of land-use change in the Maasai Steppe of northern Tanzania: an ecological, social and political analysis. Journal of Land Use Science, 6(4), 261-281. https://doi.org/10.1080/1747423X.2010.511682

Mukeka, J., Muya, S., Ihwagi, F., & Omondi, P. (2013). Status and trends of the elephant population in the Tsavo–Mkomazi ecosystem. Pachyderm, 53, 38–50. Retrieved from https://pachydermjournal.org/index.php/pachyderm/article/view/322

Mukeka, J. M., Ogutu, J. O., Kanga, E., & Røskaft, E. (2020). Spatial and temporal dynamics of human–wildlife conflicts in the Kenya Greater Tsavo Ecosystem. Human–Wildlife Interactions, 14(2), 14. https://doi.org/10.26077/bf21-497e

Muriuki, G., Seabrook, L., McAlpine, C., Jacobson, C., Price, B., & Baxter, G. (2011). Land cover change under unplanned human settlements: A study of the Chyulu Hills squatters, Kenya. Landscape and Urban Planning, 99(2), 154-165. https://doi.org/10.1016/j.landurbplan.2010.10.002

Muteti, D., & Maloba, M. (2013). Taita Hills Wildlife Sanctuary: Dry season total ground wildlife census. Unpublished data.

Mwadime, H. M., & Mbataru, P. (2022). Public Participation and Human-Wildlife Conflict Management in Taita-Taveta County, Kenya. Journal of Public Policy and Governance, 2(1), 17-25. https://edinburgjournals.org/journals/index.php/journal-of-public-policy/article/view/73

Naidoo, R., Balmford, A., Ferraro, P. J., Polasky, S., Ricketts, T. H., & Rouget, M. (2006). Integrating economic costs into conservation planning. Trends in Ecology & Evolution (Amsterdam), 21(12), 681–687. https://doi.org/10.1016/j.tree.2006.10.003

Neves, I. Q., da Luz Mathias, M., & Bastos-Silveira, C. (2018). The terrestrial mammals of Mozambique: Integrating dispersed biodiversity data. Bothalia-African Biodiversity & Conservation, 48(1), 1-23. https://hdl.handle.net/10520/EJC-1202af0872

Neves, I., da Luz Mathias, M., & Bastos-Silveira, C. (2019). Sharing the Decision Process Framework to Identify Well-supported Records of Mammal Species-occurrence in Mozambique. Biodiversity Information Science and Standards, 3, e35265.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., & Crowston, K. (2012). The future of citizen science: emerging technologies and shifting paradigms. Frontiers in Ecology and the Environment, 10(6), 298-304. https://doi.org/10.1890/110294

Ngene, S., Njumbi, S., Nzisa, M., Kimitei, K., Mukeka, J., Muya, S. & Omondi, P. (2013). Status and trends of the elephant population in the Tsavo–Mkomazi ecosystem. Pachyderm, 53, 38-50. https://pachydermjournal.org/index.php/pachyderm/article/view/322

Njogu, J. G., & Dietz, T. (2011). Land use and tenure: entitlement rights for community-based wildlife and forest conservation in Taita Taveta, Kenya. http://opendocs.ids.ac.uk/opendocs/handle/123456789/1053

Nyambariga, F. K., Opere, A. O., Kituyi, E., & Amwata, D. A. (2023). Climate change scenario projections and their implications on food systems in Taita Taveta County, Kenya. PLOS Climate, 2(6), e0000114. https://doi.org/10.1371/journal.pclm.0000114

O'Brien, T. G., Ahumada, J., Akampurila, E., Beaudrot, L., Boekee, K., Brncic, T. & Strindberg, S. (2020). Camera trapping reveals trends in forest duiker populations in African National Parks. Remote Sensing in Ecology and Conservation, 6(2), 168-180. https://doi.org/10.1002/rse2.132

Ogutu, J. O., & Owen-Smith, N. (2003). ENSO, rainfall and temperature influences on extreme population declines among African savanna ungulates. https://doi.org/10.1046/j.1461-0248.2003.00447.x

Ogutu, J. O., Owen-Smith, N., Piepho, H. P., & Said, M. Y. (2011). Continuing wildlife population declines and range contraction in the Mara region of Kenya during 1977–2009. Journal of Zoology, 285(2), 99-109. https://doi.org/10.1111/j.1469-7998.2011.00818.x

Ogutu, J. O., Owen-Smith, N., Piepho, H. P., Kuloba, B., & Edebe, J. (2012). Dynamics of ungulates in relation to climatic and land use changes in an insularized African savanna ecosystem. Biodiversity and Conservation, 21, 1033-1053. https://doi.org/10.1007/s10531-012-0239-9

Ogutu, J. O., Piepho, H. P., Said, M. Y., & Kifugo, S. C. (2014). Herbivore dynamics and range contraction in Kajiado County Kenya: climate and land use changes, population pressures, governance, policy and human-wildlife conflicts. The Open Ecology Journal, 7(1). https://benthamopenarchives.com/abstract.php?ArticleCode=TOECOLJ-7-1-9

Ogutu, J. O., Piepho, H. P., Said, M. Y., Ojwang, G. O., Njino, L. W., Kifugo, S. C., & Wargute, P. W. (2016). Extreme wildlife declines and concurrent increase in livestock numbers in Kenya: What are the causes?. PloS one, 11(9), e0163249. https://doi.org/10.1371/journal.pone.0163249

Ogutu, J. O., Kuloba, B., Piepho, H. P., & Kanga, E. (2017). Wildlife population dynamics in human-dominated landscapes under community-based conservation: the example of Nakuru Wildlife Conservancy, Kenya. PloS one, 12(1), e0169730. https://doi.org/10.1371/journal.pone.0169730

Omoro, L., Starr, M., & Pellikka, P. K. (2013). Tree biomass and soil carbon stocks in indigenous forests in comparison to plantations of exotic species in the Taita Hills of Kenya. Silva Fennica, 47(2). https://doi.org/10.14214/sf.935

Onditi, K. O., Li, X., Song, W., Li, Q., Musila, S., Mathenge, J., ... & Jiang, X. (2021). The management effectiveness of protected areas in Kenya. Biodiversity and Conservation, 30, 3813-3836. https://doi.org/10.1007/s10531-021-02276-7

Otianga-Owiti, G. E., Okori, J. J. L., Nyamasyo, S., & Amwata, D. A. (2021). Governance and challenges of wildlife conservation and management in kenya. Wildlife Biodiversity Conservation: Multidisciplinary and Forensic Approaches, 67-99. https://doi.org/10.1007/978-3-030-64682-0_4

Ottichilo, W. K., De Leeuw, J., Skidmore, A. K., Prins, H. H., & Said, M. Y. (2000). Population trends of large non-migratory wild herbivores and livestock in the Masai Mara ecosystem, Kenya, between 1977 and 1997. African journal of Ecology, 38(3), 202-216. https://doi.org/10.1046/j.1365-2028.2000.00242.x

Owen-Smith, N., & Mills, M. G. L. (2006). Manifold interactive influences on the population dynamics of a multispecies ungulate assemblage. Ecological monographs, 76(1), 73-92. https://doi.org/10.1890/04-1101

Packer, C., Loveridge, A., Canney, S., Caro, T., Garnett, S. T., Pfeifer, M., & Polasky, S. (2013). Conserving large carnivores: dollars and fence. Ecology letters, 16(5), 635-641. https://doi.org/10.1111/ele.12091

Palencia, P., Vicente, J., Soriguer, R. C., & Acevedo, P. (2022). Towards a best-practices guide for camera trapping: assessing differences among camera trap models and settings under field conditions. Journal of Zoology, 316(3), 197-208. https://doi.org/10.1111/jzo.12945

Palmer, M. S., Swanson, A., Kosmala, M., Arnold, T., & Packer, C. (2018). Evaluating relative abundance indices for terrestrial herbivores from large-scale camera trap surveys. African journal of ecology, 56(4), 791-803. https://doi.org/10.1111/aje.12566

Pellikka, P. K., Clark, B. J., Gosa, A. G., Himberg, N., Hurskainen, P., Maeda, E. & Siljander, M. (2013). Agricultural expansion and its consequences in the Taita Hills, Kenya. In Developments in earth surface processes (Vol. 16, pp. 165-179). Elsevier. https://doi.org/10.1016/B978-0-444-59559-1.00013-X

Pellikka, P. K. E., Heikinheimo, V., Hietanen, J., Schäfer, E., Siljander, M., & Heiskanen, J. (2018). Impact of land cover change on aboveground carbon stocks in Afromontane landscape in Kenya. Applied geography, 94, 178-189. https://doi.org/10.1016/j.apgeog.2018.03.017

Phaka, F. M., Vanhove, M. P., Du Preez, L. H., & Hugé, J. (2022). Reviewing taxonomic bias in a megadiverse country: primary biodiversity data, cultural salience, and scientific interest of South African animals. Environmental Reviews, 30(1), 39-49. https://doi.org/10.1139/er-2020-0092

Platts, P. J., Burgess, N. D., Gereau, R. E., Lovett, J. C., Marshall, A. R., McClean, C. J. & Marchant, R. O. B. (2011). Delimiting tropical mountain ecoregions for conservation. Environmental Conservation, 38(3), 312-324. https://doi.org/10.1017/S0376892911000191

Reddy, C. S., Satish, K. V., Saranya, K. R. L., Sri Surya, N. N., Neha, P. A., & Rajashekar, G. (2024). Harnessing essential biodiversity variables and remote sensing of earth observations-synthesizing biodiversity insights. Spatial Information Research, 32(3), 265-276. https://doi.org/10.1007/s41324-023-00558-6

Riggio, J., Kija, H., Masenga, E., Mbwilo, F., Van de Perre, F., & Caro, T. (2018). Sensitivity of Africa's larger mammals to humans. Journal for Nature Conservation, 43, 136-145. https://doi.org/10.1016/j.jnc.2018.04.001.

Robertson, T., Döring, M., Guralnick, R., Bloom, D., Wieczorek, J., Braak, K., ... & Desmet, P. (2014). The GBIF integrated publishing toolkit: facilitating the efficient publishing of biodiversity data on the internet. PloS one, 9(8), e102623. https://doi.org/10.1371/journal.pone.0102623

Rovero, F., Zimmermann, F., Berzi, D., & Meek, P. (2013). " Which camera trap type and how many do I need?" A review of camera features and study designs for a range of wildlife research applications. Hystrix. https://doi.org/10.1016/S0006-3207(03)00063-6

Rovero, F., Martin, E., Rosa, M., Ahumada, J. A., & Spitale, D. (2014). Estimating species richness and modelling habitat preferences of tropical forest mammals from camera trap data. PloS one, 9(7), e103300. https://doi.org/10.1371/journal.pone.0103300

Sandbrook, C., Luque-Lora, R., & Adams, W. M. (2018). Caught on camera: Human bycatch in camera trap images and its conservation implications. Conservation and Society, 16, 493–504.

Sandbrook, C., Clark, D., Toivonen, T., Simlai, T., O'Donnell, S., Cobbe, J., & Adams, W. (2021). Principles for the socially responsible use of conservation monitoring technology and data. Conservation Science and Practice, 3(5), Article 374. https://doi.org/10.1111/csp2.374

Scholte, P. (2011). Towards understanding large mammal population declines in Africa's protected areas: A west-central African perspective. Tropical Conservation Science 1: 1–11. https://doi.org/10.1177/194008291100400102

Schmitz, O. J. (2008). Herbivory from individuals to ecosystems. Annual Review of Ecology, Evolution, and Systematics, 39(1), 133-152. https://doi.org/10.1146/annurev.ecolsys.39.110707.173418

Secretariat, G. B. I. F. (2021). GBIF Science Review 2020. Copenhagen: GBIF Secretariat. https://gbif.jp/library/pdf/science-review/SR2020_interactive_paginated.pdf

Silveira, L., Jácomo, A. T., & Diniz-Filho, J. A. F. (2003). Camera trap, line transect census and track surveys: a comparative evaluation. Biological conservation, 114(3), 351-355. https://doi.org/10.1016/S0006-3207(03)00063-6

Smith, R. & Kasiki, S. (2000). A Spatial Analysis of Human-Elephant Conflict in the Tsavo Ecosystem. Kenya; African Elephant Specialist Group, Human-Elephant Conflict Task Force of IUCN: Gland, Switzerland.

Sollmann, R. (2018). A gentle introduction to camera-trap data analysis. African Journal of Ecology, 56(4), 740-749. https://doi.org/10.1111/aje.12557

Sosef, M. S., Dauby, G., Blach-Overgaard, A., van Der Burgt, X., Catarino, L., Damen, T. & Couvreur, T. L. (2017). Exploring the floristic diversity of tropical Africa. BMC biology, 15, 1-23. https://doi.org/10.1186/s12915-017-0356-8

Sousa-Baena, M. S., Garcia, L. C., & Peterson, A. T. (2014). Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Diversity and distributions, 20(4), 369-381. https://doi.org/10.1111/ddi.12136

Stahlschmidt, Z. R. (2011). Taxonomic chauvinism revisited: insight from parental care research. PLoS One, 6(8), e24192. https://doi.org/10.1371/journal.pone.0024192

Statistics, N. B. (2019). Kenya Population and Housing Census Volume IV: Distribution of Population by Socio-Economic Characteristics. Kenya National Bureau of Statistics.

Stein, A. B., Fuller, T. K., & Marker, L. L. (2008). Opportunistic use of camera traps to assess habitat-specific mammal and bird diversity in northcentral Namibia. Biodiversity and Conservation, 17, 3579-3587. https://doi.org/10.1007/s10531-008-9442-0

Stephenson, P. J., Bakarr, M., Bowles-Newark, N., Kleinschroth, F., Mapendembe, A., Ntiamoa-Baidu, Y. & Teferi, T. (2021). Conservation science in Africa: mainstreaming biodiversity information into policy and decision-making. Closing the Knowledge-Implementation Gap in Conservation Science:

Interdisciplinary Evidence Transfer Across Sectors and Spatiotemporal Scales, 287-321. https://doi.org/10.1007/978-3-030-81085-6_11

Štípková, Z., Tsiftsis, S., & Kindlmann, P. (2024). Is the GBIF appropriate for use as input in models of predicting species distributions? Study from the Czech Republic. Nature Conservation Research. Заповедная наука, 9(1), 84-95.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scientific data, 2(1), 1-14. https://doi.org/10.1038/sdata.2015.26

Taita Hills Wildlife Sanctuary. 2024. Conservancy page. URL: https://taitahillswildlifesanctuary.com/taita-hills-wildlife-conservancy/

Taylor, P. J., Neef, G., Keith, M., Weier, S., Monadjem, A., & Parker, D. M. (2018). Tapping into technology and the biodiversity informatics revolution: updated terrestrial mammal list of Angola, with new records from the Okavango Basin. ZooKeys, (779), 51. https://doi.org/10.3897/zookeys.779.25964

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Scientific reports, 7(1), 17615. https://doi.org/10.1038/s41598-017-18007-4

Tobler, M. W., Carrillo-Percastegui, S. E., Leite Pitman, R., Mares, R., & Powell, G. (2008). An evaluation of camera traps for inventorying large-and medium-sized terrestrial rainforest mammals. Animal conservation, 11(3), 169-178. https://doi.org/10.1111/j.1469-1795.2008.00169.x

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. Biological Conservation, 233, 298–315. https://doi.org/10.1016/j.biocon.2019.01.023

Trofino-Falasco, C., Simoy, M. V., Aranguren, M. F., Pizzarello, M. G., Cortelezzi, A., Vera, D. G. & Berkunsky, I. (2023). How effective is camera trapping in monitoring grassland species in the southern Pampas ecoregion?. Revista mexicana de biodiversidad, 94. https://doi.org/10.22201/ib.20078706e.2023.94.5243

Troia, M. J., & McManamay, R. A. (2016). Filling in the GAPS: evaluating completeness and coverage of open-access biodiversity databases in the United States. Ecology and evolution, 6(14), 4654-4669. https://doi.org/10.1002/ece3.2225

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. Scientific Reports, 7(1), 9132–9132. https://doi.org/10.1038/s41598-017- 09084-6

Tulloch, V. J., Tulloch, A. I., Visconti, P., Halpern, B. S., Watson, J. E., Evans, M. C., Auerbach, N. A., Barnes, M., Beger, M., Chadès, I., Giakoumi, S., McDonald-Madden, E., Murray, N. J., Ringma, J., & Possingham, H. P. (2015). Why do we map threats? Linking threat mapping with actions to make better conservation decisions. Frontiers in Ecology and the Environment, 13(2), 91–99. https://doi.org/10.1890/140022

Usui, R., Sheeran, L. K., Li, J. H., Sun, L., Wang, X., Pritchard, A. J. & Wagner, R. S. (2014). Park rangers' behaviors and their effects on tourists and Tibetan macaques (Macaca thibetana) at Mt. Huangshan, China. *Animals*, *4*(3), 546-561. https://doi.org/10.3390/ani4030546

Van Berkel, T. (2014). Camera trapping for wildlife conservation: Expedition field techniques. Geography Outdoors. https://www.rgs.org/CMSPages/GetFile.aspx?nodeguid=8b5f8856-9572-4a11-8066-e4bf6ec57a97&lang=en-GB〉

Vesala, R., Niskanen, T., Liimatainen, K., Boga, H., Pellikka, P., & Rikkinen, J. (2017). Diversity of fungus-growing termites (Macrotermes) and their fungal symbionts (Termitomyces) in the semiarid Tsavo Ecosystem, Kenya. Biotropica, 49(3), 402-412. https://doi.org/10.1111/btp.12422

Walden-Schreiner, C., Leung, Y. F., & Tateosian, L. (2018). Digital footprints: Incorporating crowdsourced geographic information for protected area management. Applied Geography, 90, 44-54. https://doi.org/10.1016/j.apgeog.2017.11.004

Wachiye, S., Pellikka, P., Rinne, J., Heiskanen, J., Abwanda, S., & Merbold, L. (2022). Effects of livestock and wildlife grazing intensity on soil carbon dioxide flux in the savanna grassland of Kenya. Agriculture, Ecosystems & Environment, 325, 107713. https://doi.org/10.1016/j.agee.2021.107713.

Wangai, P., Muriithi, J. K., & Koenig, A. (2013). Drought related impacts on local people's socioeconomic life and biodiversity conservation at Kuku Group Ranch, Southern Kenya. International Journal of Ecosystem, 3(1), 1-6. https://doi.org/10.5923/j.ije.20130301.01

Western, D., Russell, S., & Cuthill, I. (2009). The status of wildlife in protected areas compared to non-protected areas of Kenya. PloS one, 4(7), e6140. https://doi.org/10.1371/journal.pone.0006140

Willemen, L., Cottam, A. J., Drakou, E. G., & Burgess, N. D. (2015). Using social media to measure the contribution of red list species to the nature-based tourism potential of African protected areas. PloS one, 10(6), e0129785. https://doi.org/10.1371/journal.pone.0129785

Williams, A. P., & Funk, C. (2011). A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa. Climate Dynamics, 37, 2417-2435. https://doi.org/10.1007/s00382-010-0984-y

Wilson, J. R., Procheş, Ş., Braschler, B., Dixon, E. S., & Richardson, D. M. (2007). The (bio) diversity of science reflects the interests of society. Frontiers in Ecology and the Environment, 5(8), 409-414. https://doi.org/10.1890/060077.1

Wolters, V., Bengtsson, J., & Zaitsev, A. S. (2006). Relationship among the species richness of different taxa. Ecology, 87(8), 1886-1895. https://doi.org/10.1890/0012-9658(2006)87[1886:RATSRO]2.0.CO;2

World Bank. (2024). Terrestrial protected areas (% of total land area) - Kenya. URL: https://data.worldbank.org/indicator/ER.LND.PTLD.ZS?locations=KE

Wu, X., Lindsey, G., Fisher, D., & Wood, S. A. (2017). Photos, tweets, and trails: Are social media proxies for urban trail use?. Journal of Transport and Land Use, 10(1), 789-804. https://www.jstor.org/stable/26211756

Yang, H., Viña, A., Tang, Y., Zhang, J., Wang, F., Zhao, Z., & Liu, J. (2017). Range-wide evaluation of wildlife habitat change: A demonstration using Giant Pandas. Biological Conservation, 213, 203-209. https://doi.org/10.1016/j.biocon.2017.07.010

Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M. & Culham, A. (2007). How global is the global biodiversity information facility?. PloS one, 2(11), e1124. https://doi.org/10.1371/journal.pone.0001124

Zheng, Y., Zhang, Y., Mou, N., Makkonen, T., Li, M., & Liu, Y. (2024). Selection biases in crowdsourced big data applied to tourism research: An interpretive framework. *Tourism Management*, *102*, 104874. https://doi.org/10.1016/j.tourman.2023.104874