Scientific support to the trial of Spoor AI at the European Offshore Wind Deployment Centre

Brighton, C.H., Clarke, J.A. & Boersch-Supan, P.H.

Birds Science People

Scientific support to the trial of Spoor AI at the European Offshore Wind Deployment Centre

C.H. Brighton, J.A. Clarke, P.H. Boersch-Supan

Report of work carried out by

the British Trust for Ornithology

under contract to

VATTENFALL VINDKRAFT AB
and
SPOOR AS

BTO Research Report 781

March 2025

British Trust for Ornithology, The Nunnery, Thetford, Norfolk, IP24 2PU
Registered Charity No. 216652

ACKNOWLEDGEMENTS
We are grateful to the entire team at Spoor for their open collaboration throughout the project, and to the project advisory group members for their input at various stages. Particular thanks is due to Aonghais Cook for leading the early stages of this project.

Table of contents

Executive Summary

1. This report assesses the capability of a Spoor AI camera system with both mono-vision (single-camera) and stereo-vision capabilities for bird monitoring deployed at the European Offshore Wind Deployment Centre in Aberdeen Bay using both theoretical considerations and onshore and offshore field trials.
2. A recent assessment of single-camera photogrammetry has highlighted natural body size variation as a fundamental driver of measurement error. We review body size variation in 20 species of seabirds and raptors that are of interest in a renewable energy context and assess the potential range errors caused by this natural variation.
    a. We find body size variation in wing length measurement that is generally on the order of 2-10 %, which translates into ranging errors of 5-10%.
    b. Ranging errors are particularly large in species that are sexually monomorphic in appearance, but exhibit large sexual size-dimorphism, in particular gulls. Large interspecific differences, especially among gulls, may result in particularly large range uncertainty when species identification is incomplete or when misidentifications are prevalent in datasets.
    c. Biometric measurements that are of interest for image analysis applications (wingspan, total length) differ from those that are routinely collected by ornithologists on live birds or museum specimens (wing chord, skeletal and bill measures). As a consequence, much of the existing biometric data is difficult to apply to image-based ranging techniques.
3. An onshore field trial was conducted to assess the measurement accuracy of mono-vision and stereo-vision systems under controlled conditions, using a drone with high-precision GPS positioning as an experimental target of known size at ranges up to 500m.
    a. For the stereo-vision system manual target trajectory reconstructions performed comparably to automated methods at shorter ranges, but achieved marginally smaller distance errors at longer ranges, particularly in lateral and vertical planes (c. 2% of the target range). Depth (range) errors were consistently larger than in the other two dimensions (c. 5% of target range).
    b. Positioning errors increased with greater distances and heights, emphasizing the sensitivity of accuracy to camera configuration and synchronization.
    c. For one investigated camera model (Avigilon) frame drops and time synchronization drift were observed, affecting data reliability.
    d. Mono-vision systems exhibited significantly larger distance errors compared to stereo-vision at c. 5% of the target range in lateral and vertical dimensions and 10-15% along the depth (range) axis.
4. An offshore trial at EOWDC involved the continuous collection of video footage over 18 months, and a concurrent human observer trial utilising a laser range finder (LRF) over three days in August 2023.
    a. Mono-vision systems produced in excess of 100,000 tracks across four cameras.
    b. Stereo-vision footage was analysed to manually track birds and reconstruct 3D flight paths. The calibration of stereo cameras relied on stationary points, landmarks, and turbine features.
    c. A total of 90 bird tracks were manually reconstructed in the period of the human observer trial. A limiting factor for this analysis workflow was the need for separate

calibrations each day of the measurement campaign. More rigid camera housings in future deployments would reduce this requirement.

d. Stereo-vision reconstructions were deemed more accurate based on estimated flight speeds and the shape of bird trajectories, even though calibrations were challenging because of a lack of reliable landmarks, resulting in an underestimation of known ranges by ~20%. Despite a short baseline distance stereo-reconstructions were considered to be reliable to a range of c. 500m. Increased frame rates and/or baseline distances in future deployments would likely expand the reliable stereo-range further towards the targeted turbine.
e. Mono-vision reconstructions focussed on birds flying above the horizon angle, as the employed tracking algorithm was not designed to detect birds in front of the sea surface. Mono reconstructions resulted in bird ranges that were c. 50% larger than the corresponding stereo-reconstructions. Mono tracks showed more variability and artefacts, such as jitter. Improved range estimates should be possible to obtain from a recalibration based on the stereo data.
f. LRF observations were intended to validate camera-derived data, although due to technical challenges the LRF data were not fit for quantitative validation. Out of 90 birds tracked via stereo cameras, 63 were matched with LRF observations. Only 35% of stereo tracks and 20% of mono tracks overlapped with LRF readings.

5. A statistical model was developed based on distance sampling methods to estimate the three-dimensional distribution of birds within the sampled volume of the camera system while accounting for imperfect detection, and in particular a drop-off in detection with distance to the camera. We implemented the model in the R programming language using maximum likelihood optimisation to integrate detectability gradients and bird density gradients within the pyramidal sampled volume of the camera.

a. Simulations tested various scenarios, including different detection functions, flight height distributions, field of view constraints, and observation errors. Results emphasized the importance of accurate detection function and flight height selection, adequate sample sizes (150–200 tracks), and accounting for errors in positional data.

b. Using c. 3,000 mono-vision tracks the model estimated species-specific flux and flight height distributions for European Herring Gull, Black-legged Kittiwake, and Northern Gannet.

c. Offshore facing cameras with a 48mm lens provided shorter detection ranges, but more reliable data due to a larger field of view compared to inshore cameras, which aided identification of the vertical bird density gradient.

d. Estimated detection functions varied by species and with camera focal length. Half-normal detection function standard deviations ranged from c. 250m for Kittiwake at 48mm focal length lens to c. 750m for Gannet at 70mm focal length (nominal distances from mono-vision reconstructions). At a range of 500m this translates to detections of c. 14% of available birds in the former case, and 80% of available birds in the latter.

e. Weather conditions significantly influenced flight height and detectability. For example, Herring Gulls flew higher during rain, while Gannets exhibited greater detectability in clear weather. Weather covariates improved model accuracy for certain species.

    f. In the mono-vision workflow, object tracking was generally possible at greater ranges than species identification. Species identification therefore constrained the effective detection range for single-species density and flight height models.
    g. At present the statistical model does not incorporate adjustments for (i) range measurements errors, nor (ii) does it explicitly address horizontal bird density gradients, such as those caused by meso-avoidance. Not accounting for range errors led to overestimated detectability parameters and underestimated densities. Correcting these errors using range calibrations based on stereo-reconstructions yielded more reliable results. Simulations indicated that the bias resulting from meso-avoidance, was minimal for the species analysed.
6. In conclusion we find that the Spoor AI system provides a promising tool for cost-effective scalable bird monitoring in the offshore environment. In particular, the system is capable of delivering the relatively large sample sizes (100s of tracks per species and environmental condition) that are required to estimate three-dimensional bird distributions in the presence of imperfect detection. Such in-situ estimates of bird distributions within OWFs have the potential to greatly reduce uncertainty in outputs of current collision risk models which are highly sensitive to assumptions around avoidance behaviour at macro- and meso-scales.
7. Our validation was limited to ranges shorter than distance of the targeted turbines. Further improvements to the technology, in particular the camera hardware, should enable better positional accuracy of reconstructed tracks at ranges beyond 500m.
8. There are fundamental trade-offs between mono-vision and stereo-vision approaches in terms of measurement capability, equipment costs, and analytical effort. Mono-vision systems are more limited in the precision of track reconstructions for wild birds, but stereo-vision systems come with added cost and complexity. A combination of both technologies may combine strengths of both systems, e.g. by scaling out monitoring with many mono-vision systems and using a smaller number of stereo systems for internal validation and calibration. In-situ validation of the offshore system proved challenging, and follow-up investigations highlighted fundamental limitations in the capability of human operated LRF devices to comprehensively collect validation data. Novel approaches are required to externally validate sensor-based offshore systems in-situ.
9. We recommend further development of stereo-vision capabilities in the Spoor AI workflow by improving frame synchronization, ensuring more rigid camera mounting, enlarging the baseline separation distance and/or conducting in-situ calibrations with drones or surface vessels. A more automated stereo-vision workflow could then provide a means of internal calibration for mono-vision range estimation. In addition, adding tracking capabilities for birds in front of the sea or land surface to the mono-vision system (as in other existing Spoor AI deployments) would likely greatly improve bird density and flight height estimation.

1 General Introduction

1.1 Background and objective

Offshore wind farms play a critical role in generating clean energy with minimal greenhouse gas emissions. However, their potential impacts on local populations of seabirds and migratory population of other bird groups raises environmental concerns, especially as increasing numbers of turbines are installed along key migratory routes and in the foraging ranges of seabird colonies (Croll et al., 2022). However, the scale and nature of interactions between birds and offshore wind farms remain uncertain, in part because observations of such interactions are technically and logistically challenging, and no existing monitoring approach is without sampling and or measurement imperfections. Further developments to monitoring technologies are necessary to obtain a more robust evidence base, and to allow scalable monitoring as the number and extent of OWFs continue to grow.

The overall objective of this study was to assess the utility of a Spoor AI system to monitor movements of birds around individual offshore wind turbines at the European Offshore Wind Deployment Centre (EOWDC) in Aberdeen Bay, with a particular focus on assessing the precision and accuracy of generated track reconstructions of individual birds, and the potential to quantify bird flux and to detect avoidance behaviour in the close vicinity of the monitored turbine.

To achieve this, we conducted theoretical (Section 2) and experimental (Sections 3,4) work to assess the measurement accuracy of the system, that is its potential of reliably reconstructing the true positions of imaged seabirds using both single-camera ('mono-vision') and stereo camera approaches.

To be able to conduct the estimation of bird flux, i.e. the true number of birds traversing the airspace of the OWF in a given time interval, we further assessed the sampling characteristics of the system and developed a statistical model to estimate bird density while accounting for both sampling and behavioural characteristics (Section 5).

1.2 Performance criteria for bird monitoring systems

To adequately characterise interactions between birds and offshore wind infrastructure, monitoring systems should deliver species-specific data that is representative of all individual characteristics (e.g., age, sex, body size), behaviours (e.g., foraging, commuting, resting), annual cycle stages (i.e., breeding, non-breeding, migration) and environmental conditions (e.g., temperature, windspeed, precipitation, food) that can be encountered within the area of interest (Feather et al., in press). Measurements should be free from bias, accurate and precise. The types of error and uncertainty associated with bird monitoring data can be broadly grouped into measurement and sampling processes, respectively (Borchers et al., 2002; Buonaccorsi, 2010). Measurement error here refers to the accuracy and precision of the positional information (range, flight height) and species identities of detected birds. Sampling error in this context emerges from the experimental design and the method of data collection and is shaped by device characteristics such as the sampling geometry and detection efficiency.

1.2.1 Measurement error

All monitoring methods incorporate some degree of positional error, which generally arises from multiple sources (e.g., equipment, operator, supplementary data) and at various stages (e.g., data collection, data analysis) of the sampling process. As the complexity of sampling methods increases, errors can arise and interact in increasingly complex and counterintuitive ways which further complicates all subsequent inferences.

Measurement errors that arise due to equipment characteristics (e.g., sensor accuracy, sampling frequency) are generally inherent to data collection and usually generate random noise around range or height estimates (Harwood et al., 2018; Lato et al., 2022). Errors that result from interactions between range / height measurements and supplementary data (e.g., sea level pressure, target size reference values) are introduced while data are processed and may introduce both random noise and systematic bias to height estimates (Boersch-Supan et al., 2024; Johnston et al., 2023; Schaub et al., 2023).

1.2.2 Sampling error

There is considerable uncertainty associated with each method's ability to adequately sample the population of interest and none of the methods discussed can provide species-specific flight height distributions that are fully representative of the populations (i.e., properties, constituents), environmental conditions (i.e., biotic, abiotic) and temporal scale (e.g., diel cycles, decades) they aim to describe.

Static sensor-based methods, such as camera systems, fundamentally sample a finite volume of airspace (hereafter 'sampled volume'), and the maximum sampled volume is typically determined by sensor characteristics, i.e. the geometry of the field of view. The sampled volume of camera-based approaches generally increases with increasing distance from the sensor due to the pyramidal shape of the surveyed volume, which is governed in the first instance by the aperture angles of the employed cameras (Feather et al., in press).

In addition to complex geometries of sampling volumes, the detection efficiency of monitoring methods is generally not uniform across the sampled volumes and effective sampling rates are therefore variable. Typically detectability decreases with increasing distance from the sensor due to limitations associated with visibility, optical resolution and / or targeting individuals (Barbraud and Thiebot, 2009; Borkenhagen et al., 2018; Harwood et al., 2018).

Both the sampled volume and detection probability may also vary considerably (temporal and spatial) in response to environmental conditions. The sampled volume and detection probability of imagery-based methods typically decreases with worsening weather conditions and decreasing light-levels. Detection probabilities are also fundamentally related to the species identity of birds - larger birds and/or those with plumage that contrasts the background can be detected at greater distances relative to smaller and/or more cryptically coloured ones (Barbraud and Thiebot, 2009; Cook et al., 2018; Schmid et al., 2019).

Image processing and target-tracking algorithms interact with these factors and typically have variable performance for different image-context (e.g. species identity, type of background).

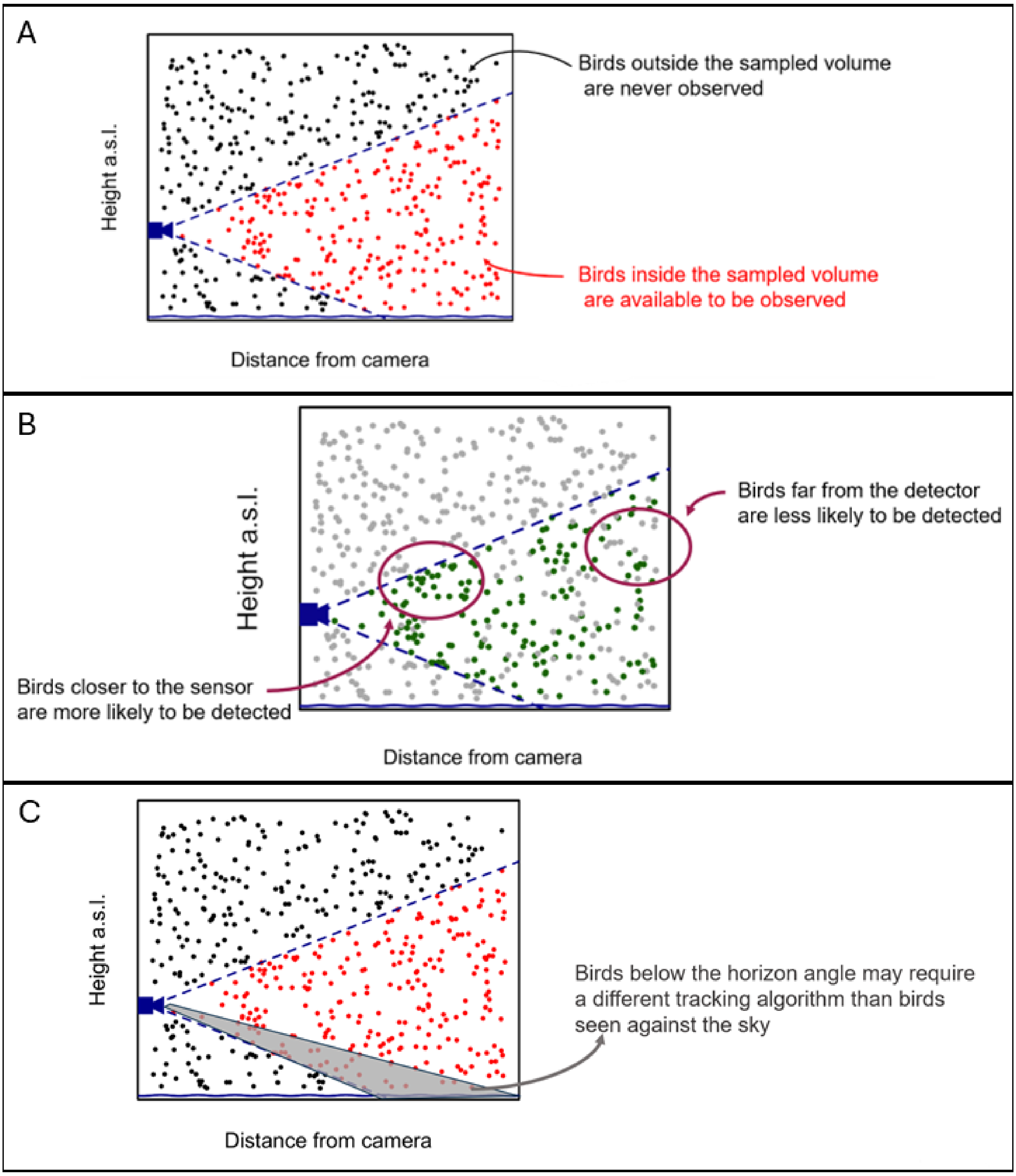

Figure 1-1: Conceptual schematic of the processes affecting sampling efficiency of a camera-based bird monitoring system (adapted from Feather et al. in press; a.s.l: above sea level). A: The geometry of the sampling volume governs which birds are available to be detected. B: The distance from the sensor generally governs the probability of an available bird being detected. C: Image backgrounds can have a strong effect on detectability, e.g. a dynamic background such as the sea surface may require different processing techniques than a more static background such as the sky.

1.3 Onshore and offshore field trials

We investigated the practicality of a Spoor AI-powered camera system by conducting onshore experimental work with a well-defined target – a drone with a high-precision GPS positioning system, as well as under field conditions at the deployment site at EOWDC in

Aberdeen Bay (Figure 1-2). At EOWDC two stereo camera pairs were deployed on a single turbine (AWF10), with the aim of reconstructing the three-dimensional flight paths of birds passing close to another turbine approximately 900m away. One pair was using 70mm focal length lenses aimed at turbine AWF05 (inshore/landward-facing) and one pair was using 48mm focal length lenses aimed at turbine AWF11 (offshore/seaward-facing). The high-definition video cameras were mounted on the turbine platform inside weatherproof housings approximately 20m above sea level. Strict space limitations meant that for each stereo pair the cameras were positioned only 4m apart, resulting in an almost parallel configuration.

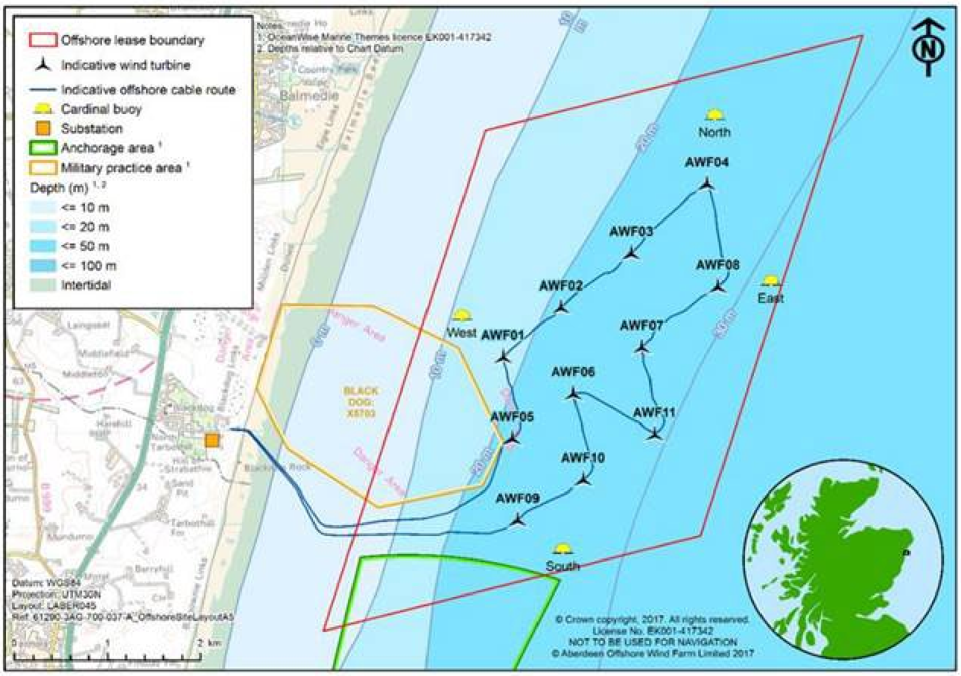

Figure 1-2: EOWDC at Aberdeen Bay, showing the location of the cameras (AWF10) and the two monitored turbines AWF05 (inshore) and AWF11 (offshore).

1.4 References

Barbraud C, Thiebot J-B. 2009. On the importance of estimating detection probabilities from at-sea surveys of flying seabirds. J Avian Biol. 40:584-590. https://doi.org/10.1111/j.1600-048X.2009.04653.x

Boersch-Supan PH, Brighton CH, Thaxter CB, Cook ASCP. 2024. Natural body size variation in seabirds provides a fundamental challenge for flight height determination by single-camera photogrammetry: a comment on Humphries et al. (2024). Mar Biol. 171:122. https://doi.org/10.1007/s00227-024-04396-4

Borchers DL, Buckland ST, Zucchini W. 2002. Estimating animal abundance. Springer, London

Borkenhagen K, Corman AM, Garthe S. 2018. Estimating flight heights of seabirds using optical rangefinders and GPS data loggers: a methodological comparison. Mar Biol. 165:17. https://doi.org/10.1007/s00227-017-3273-z

Buonaccorsi JP. 2010. Measurement error: models, methods, and applications. CRC Press, Boca Raton

Cook ASCP, Ward RM, Hansen WS, Larsen L. 2018. Estimating Seabird Flight Height using LiDAR. Scott Mar Freshw Sci. 9. https://doi.org/10.7489/12131-1

Croll DA, Ellis AA, Adams J, Cook ASCP, Garthe S, Goodale MW, Hall CS, Hazen E, Keitt BS, Kelsey EC, Leirness JB, Lyons DE, McKown MW, Potiek A, Searle KR, Soudijn FH, Rockwood RC, Tershy BR, Tinker M, VanderWerf EA, Williams KA, Young L, Zilliacus K. 2022. Framework for assessing and mitigating the impacts of offshore wind energy development on marine birds. Biol Conserv. 276:109795. https://doi.org/10.1016/j.biocon.2022.109795

Feather AP, Burton NHK, Johnston DT, Boersch-Supan PH. In press. A review of existing methods to collect data on seabird flight height distributions and their use in offshore wind farm impact assessments. BTO Research Report.

Harwood AJP, Perrow MR, Berridge RJ. 2018. Use of an optical rangefinder to assess the reliability of seabird flight heights from boat-based surveyors: implications for collision risk at offshore wind farms. J Field Ornithol. 89:372-383. https://doi.org/10.1111/jofo.12269

Johnston DT, Thaxter CB, Boersch-Supan PH, Davies JG, Clewley GD, Green RMW, Shamoun-Baranes J, Cook ASCP, Burton NHK, Humphreys EM. 2023. Flight heights obtained from GPS versus altimeters influence estimates of collision risk with offshore wind turbines in Lesser Black-backed Gulls Larus fuscus. Mov Ecol. 11:66. https://doi.org/10.1186/s40462-023-00431-z

Lato KA, Stepanuk JEF, Heywood EI, Conners MG, Thorne LH. 2022. Assessing the accuracy of altitude estimates in avian biologging devices. PLOS ONE. 17. https://doi.org/10.1371/journal.pone.0276098

Schaub T, Millon A, De Zutter C, Buij R, Chadœuf J, Lee S, Mionnet A, Klaassen RHG. 2023. How to improve the accuracy of height data from bird tracking devices? An assessment of high-frequency GPS tracking and barometric altimetry in field conditions. Anim Biotelemetry. 11:31. https://doi.org/10.1186/s40317-023-00342-1

Schmid B, Zaugg S, Votier SC, Chapman JW, Boos M, Liechti F. 2019. Size matters in quantitative radar monitoring of animal migration: estimating monitored volume from wingbeat frequency. Ecography. 42:931-941. https://doi.org/10.1111/ecog.04025

2 A review of intra-specific body size variation in seabirds and raptors and potential implications for imagery-based monitoring approaches in the renewable energy context

2.1 Introduction

The rapid growth of wind power generation in the UK and Europe, particularly offshore, is of increasing conservation concern due to the direct risk to birds from collisions. Many of the most impacted species have vulnerable conservation status; At offshore turbines, these include large gulls and other seabirds (Croxall et al., 2012; Furness et al., 2013), and at onshore turbines, these include raptors. While raptor fatalities are generally low in absolute terms, low species abundances cause raptors to be disproportionately affected by wind turbine collisions (Allison et al., 2019; Thaxter et al., 2017).

While reducing carbon dioxide emissions remains imperative, sustainable development must also aim to minimize impacts on nature and wildlife. Wind farm developers in the UK and elsewhere are obliged to assess bird densities within their proposed site to inform their Environmental Impact Assessments (EIAs). Most assessments use collision risk models (CRMs: Band, 2012; Masden and Cook, 2016; McGregor et al., 2018) to estimate i) the number of birds flying through the rotor-swept zone (RSZ), ii) the probability of an individual bird colliding with an individual wind turbine and iii) overall collision rates across the proposed wind farm. Higher proportions of birds at collision risk height result in proportional increases in predicted collision rate (Masden et al., 2021).

CRMs require data on both turbine specifications (such as rotor speed, turbine height) and bird characteristics (such as flux, flight height, flight speed, avoidance rates) – and are especially sensitive to the flight parameter inputs. These flight characteristics are known to vary with different behaviours (Cleasby et al., 2015; Fijn and Gyimesi, 2018) and environmental conditions (Hanssen et al., 2020), thus collision risk may vary across behaviours. This variability is poorly characterized, and not consistently incorporated into CRMs, and as a result, collision risk estimates are based on potentially incomplete or low-quality data and overly simplistic model assumptions. Currently in the UK, there is no systematic effort to quantify or validate collision rates predicted by CRMs (Ballester et al., 2024). Consequently, our ability to assess the true level of impact of wind farms on seabird and raptor populations is restricted.

At present, the flight characteristics are determined using a variety of approaches. The recommended method for collecting flight height data is observer-based visual estimation (Strickland et al., 2011), while speed data are taken from existing values in the scientific literature which may in turn be based on estimates from stationary observers or animal-borne sensors (comprising ground speed rather than air speed - e.g. Alerstam et al., 2007).

A more scalable and reliable method to quantify flight parameters, and in turn collision risk, is to expand the use of sensor-based technologies (Chilson et al., 2012; Shepard et al., 2016; Thaxter et al., 2016).

Understanding the accuracy and precision of these new tools is crucial to assess whether they can support the generation of a more robust evidence base around collision risks than the status quo, and - where precision is limited – whether they allow a robust quantification of uncertainty. Both aspects are required to provide confidence in the evidence used to inform development consenting decisions (Searle et al., 2023).

Here, we quantify the level of natural intra-specific variation in seabird and raptor body size, to highlight how this variation may result in substantial uncertainties and possible biases in bird position and/or trajectory reconstruction based on image analysis.

2.2 Image-based ranging and trajectory reconstruction

Monitoring birds in wind farms is challenging which is why a wide range of approaches are available for measuring bird positions, including flight heights (Largey et al., 2021). To build behaviour-based CRMs three-dimensional flight data are required (Masden and Cook, 2016), which can only be reliably obtained from sensor-based measurement. As a result of recent advances in digital technologies several image-based approaches to ranging/positioning have been proposed, both in research settings and for commercial application (e.g. Prinsloo et al., 2021; Nicholls et al., 2022; Humphries et al., 2023; Feather et al., in press). While the equipment cost may be higher than traditional human observer-based visual methods, there is a greater capacity to quantify the measurement characteristics of these sensor-based devices, allowing for model-based corrections of sampling and measurement errors. In addition, such systems offer the potential to quickly increase the number of monitoring data, as they can be deployed in remote locations, collect large volumes of flight data on temporal and spatial variability, and create a permanent record of observations that can be re-analysed. Different technologies have different uses, and there are often trade-offs between cost and image resolution, affecting how well birds can be identified in images (Nicholls et al., 2022). Ultimately, for the technology to be of value in terms of providing accurate data for bird monitoring, the image processing must be carefully considered, and appropriate choices made.

Image-based ranging and flight height determination, or more generally, 3D trajectory reconstruction generally falls into one of two overarching approaches: Single-image (also known as single-camera or mono-vision) and stereo-image (stereo-vision) techniques.

Stereo-based approaches are attractive on theoretical grounds, as they are in principle capable of reconstructing trajectories of arbitrary targets without requiring any auxiliary data about the targets themselves. Stereo-vision approaches rely on triangulation for rangefinding and the geometry of the camera pair or camera array is sufficient for track reconstruction, and the accuracy of the reconstruction is linked to device characteristics such as image resolution, temporal synchronization of stereo-image pairs, and the accuracy of the knowledge of the system geometry. Sub-metre spatial accuracy at ranges of up to several hundred metres can be achieved (Brighton et al., 2019, 2022, Prinsloo et al., 2021). However, to reliably achieve such precision, instrumentation requirements not only involve at least twice as many cameras as mono-vision approaches, but also additional features such as large baseline distances between cameras, electronically synchronized shutters and/or high image resolution and frame rates. This generally leads to increased instrumentation costs, increased image processing requirements, and may decrease deployability, e.g. on moving survey platforms.

For better scalability and economy, a number of mono-vision approaches have been proposed (e.g. Lyon, 1994; Willisch et al., 2013; Humphries et al., 2023). Single-image photogrammetry relies on the fact that the perceived size of an object decreases as distance to the camera increases (Fig. 2-1).

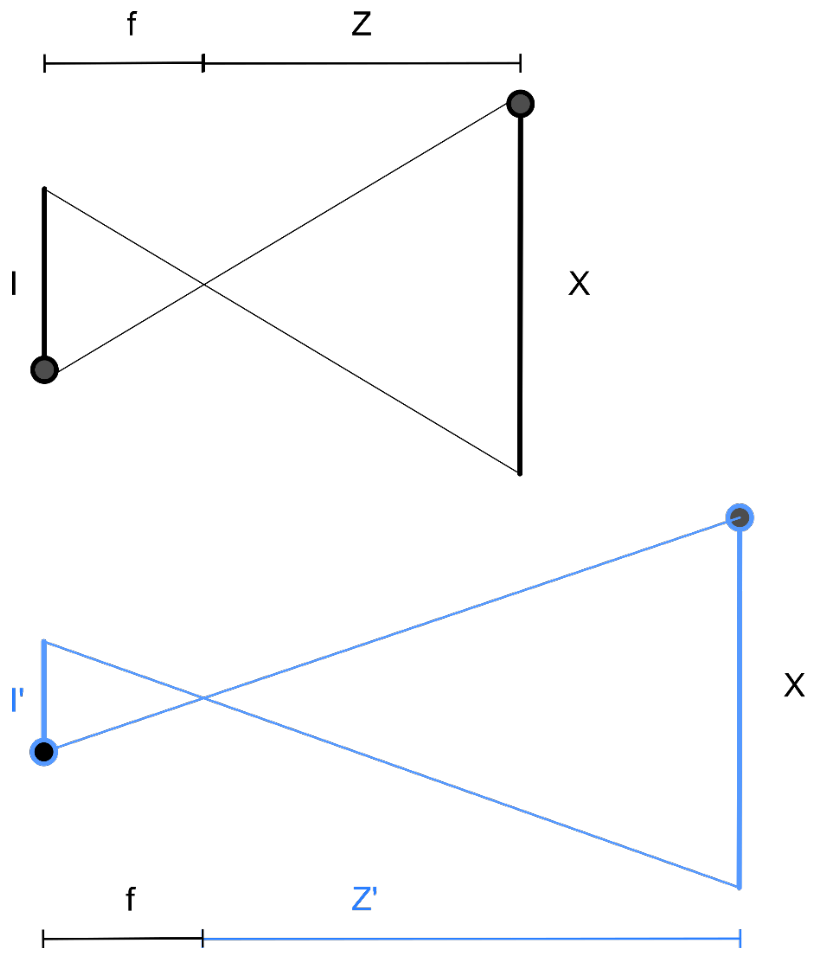

Figure 2-1: Simplified pinhole geometry of photogrammetric relationships. For a fixed focal length f the apparent size I of a target object X decreases to I' as the range Z increases to Z'. When X is known the target range can be estimated as Z = fX/I as per the angle-angle-angle similarity theorem.

However, for objects of unknown size, distance and size cannot be estimated at the same time from a single-image alone. Single-camera photogrammetry therefore requires either auxiliary information on the range to the target (Bergeron, 2007; Jaquet, 2006; Lyon, 1994), or information about target sizes, i.e. individual or species biometric measurements (Willisch et al., 2013; Humphries et al., 2023). Any uncertainty in these auxiliary data add to the sources of uncertainty in the imaging process (e.g. image resolution), and it is therefore crucial to adequately quantify and propagate these uncertainties into the estimated flight heights or track reconstructions.

Error sources arising from the imaging process (e.g. image resolution, variation in camera spec) have been covered in depth elsewhere (e.g. Scherz, 1974). At very long ranges, i.e. when the apparent size approaches that of a single pixel of the image, the uncertainties from the imaging process dominate the overall uncertainty. However, we here focus on the biological component of uncertainty, i.e. the role of bird body size variation (and hence target size variation) as a source of uncertainty in single-camera photogrammetry applications, as recent work has demonstrated that this may be the dominant source of photogrammetry error at short to intermediate ranges (i.e. where the image size is considerably larger than the image resolution).

A fundamental property of the photogrammetric relationships is that substituting an unknown target length with a mismatched reference length, such as the mean length of the population of targets, will result in biased range estimates. E.g. any bird that is shorter than the average size of its species will appear further away and hence be assigned a positively biased range estimate. Conversely, every bird that is larger than average will appear closer to the camera than an average sized bird at the same distance (Fig. 2-2). Consequentially, when using an average body size as reference length, the inferred distribution of individual range estimates is more dispersed than the underlying true distribution of range estimates.

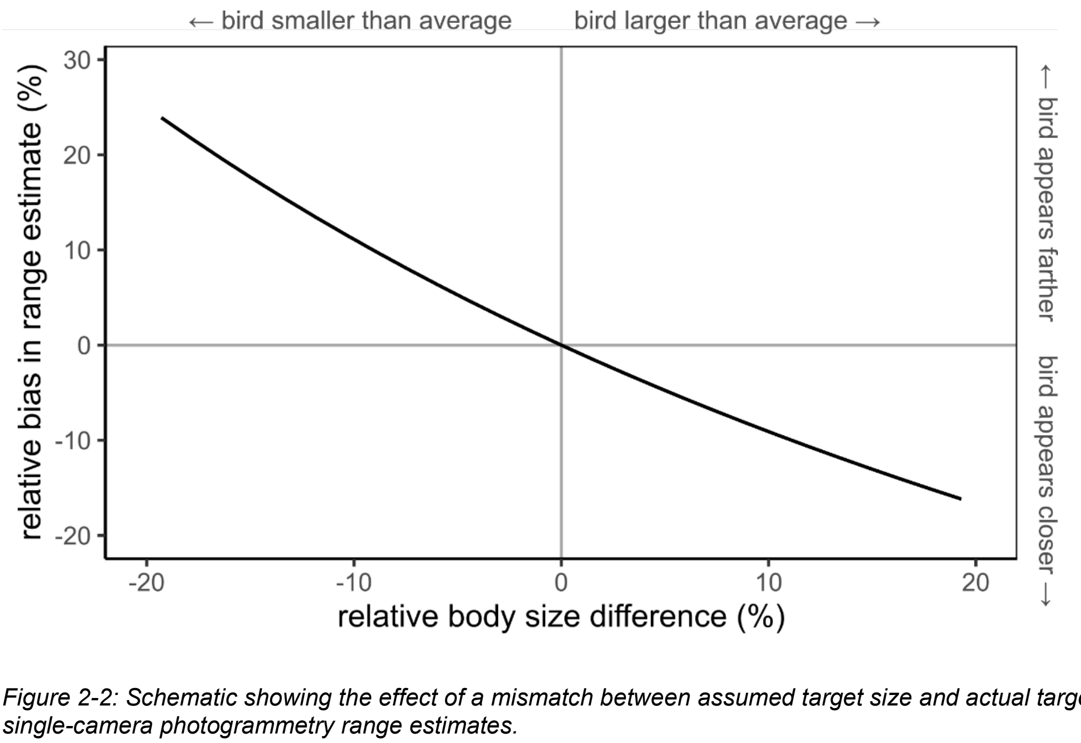

Figure 2-2: Schematic showing the effect of a mismatch between assumed target size and actual target size on single-camera photogrammetry range estimates.

In a three-dimensional case any biases in range estimates incur related biases in all three axes. For example, for targets above the central camera axis positive range biases will lead to positive elevation biases, and for targets below the central axis positive range biases will lead to negative elevation biases.

2.3 Bird measurements

2.3.1 Measuring live birds

Most bird ringing schemes encourage the collection of biometric data whenever captured birds are handled for ringing or ring reading. Measurements on live birds must be taken with the utmost consideration for animal welfare. Measurements must be quick to avoid keeping birds captive for longer than necessary, and measurements must not cause undue distress to the bird in the hand. Additionally, such measurements are generally collected under field conditions, which provides challenges such as working in remote locations, suboptimal illumination, and exposure to the elements. This limits which body parts can be measured. Generally, weight (body mass) and one wing measurement are conducted as a minimum, although additional measurements are routinely recorded, including measurements of certain flight feathers, and/or external measurements of the hind limbs, bill, or head.

2.3.2 Measuring museum specimens

Dead bird specimens are usually prepared by a taxidermist, who preserves their skins by drying them. Generally, absolute length measurements from dried specimens are 0.5 – 2.5% lower due to tissue shrinkage, and it has been suggested that a shrinkage factor be applied to all data derived from skins (Williams, 2017; Wilson and McCracken, 2008). However, any correction factor needs to be species-specific, and shrinkage has also been shown to vary within a species (Ewins, 1985). Therefore, other studies caution this approach, as measurement error depends heavily on the species, the measure required and the method (Barrett et al., 1989; Eck et al., 2012) and hence may be greater sources of error than any accrued during the drying process. Nevertheless, all methods – whether on live birds or skins – are heavily influenced by the experience of the measurer.

2.3.3 Other sources of measurements

A recent study in UK waters derived flight height data from digital aerial surveys, using single-camera photogrammetry (Humphries et al., 2023). Their estimates of flight height relied on reference lengths of each bird species flying at the sea surface (identified as such by their reflections on the water).

2.4 Bird biometrics

2.4.1 Wing measurements

The most frequent and most important biometric measurement taken for birds – whether live, freshly dead, or a preserved skin – is a measure of wing length. Total body size, especially in birds, is difficult to quantify in a single measure, and so wing length serves as a proxy since it is highly correlated with body mass, and thus can be used to determine sex and/or age in some species (Dawson, 2005). Taking wing measurements is problematic, due to differences between observers and a lack of consistency in technique/preparation (Ewins, 1985). In addition, wing length also has at least two interpretations dependent on the field of study –ornithologists use the distance between the carpal joint (bend in the wing at the wrist joint) to the tip of the longest primary on a closed wing (wing chord, Figure 2-2), whereas in biomechanics the wing length measure used for assessing flight performance is from the shoulder joint (where the wing meets the body) to the wing tip on an outstretched wing, (which generally translates to wingspan as 2x wing length + body width). For live birds, the most widely published measure (recommended by the European Union for Bird Ringing, EURING) is wing chord, using a ruler with a zero-stop, however there are further variations in how it is measured. In Europe, the wing camber (natural arc of the primaries) is flattened against the measuring device and the primaries are straightened as much as possible (Svensson, 1992), while in the US, the measurement is taken on an unflattened wing across the natural camber (Ralph, 1993). For museum specimens, measurements will be affected by a small shrinkage due to the removal of tissue from the carpalia, or if the wing is sewn or positioned too closely to the body during drying this may cause exaggerated curvature of the primaries and decrease the wing length (Jenni and Winkler, 1989). While the lengths of individual primary feathers do not suffer shrinkage (Jenni and Winkler, 1989). All wing length measurements are rounded to nearest mm (so any measure is ±0.5mm).

Total wingspan is a measurement usually only taken on live or freshly killed birds, and involves putting the bird on its back, on top of a ruler, with outstretched but relaxed wings (i.e. not stretched). Even though this measure is of high utility for image analysis, it is rarely taken due to the difficulties and animal welfare risks involved, or not taken in a single measure (i.e. take a half-span measurement which is then multiplied by two and added to the body width). In principle wingspan measurements can be derived from imagery taken on free-flying birds (e.g. Humphries et al., 2023), but this is expected to incur additional uncertainty from postural variation and/or optical foreshortening when the bird is not aligned with the image plane.

2.4.2 Tarsus

Usually measured with callipers, from the notch on the back of the intertarsal joint (knee) to the edge of the last scale before the toes separate, or where the foot bends backwards. All tarsus measurements are rounded to nearest 0.1mm (or 1mm if larger than 100mm)

2.4.3 Bill length

Several types of bill measurements are found in the literature. The most widespread is the measurement from the distal edge of the nostril to the tip of the bill (Winker, 1998). Other methods include the total culmen length which measures from the bill tip to the base of the skull (notch on the forehead), and exposed culmen length which measures from the bill tip to the start of the forehead feathers covering the bill. The latter is the most variable measurement since the feather edge is less of a distinct point. All bill measurements are rounded to nearest 0.1mm (or 1mm if larger than 100mm).

2.4.4 Bill depth

This is the maximum measurement between top and bottom mandible either at the base of the bill where the feathering starts or at the nostril. In skin specimens the mandibles can dry in unnatural positions – specimens with any gap between the cutting edges, which cannot be closed, should be disregarded.

2.4.5 Tail

The tail is measured by placing a ruler under the tail and pushing it gently against the base of the central pair of tail feathers. The measurement is taken to the tip of the longest tail feather when the tail is folded naturally. Shrinkage is not one-directional and for the tail, it can result in larger measurements than compared to live specimens, as the contact point of measurement may shift anteriorly (Eck et al., 2012).

2.4.6 Total length

Recent image analysis techniques have identified species from measurements of head-to-tail length (Humphries et al., 2023), however, this is rarely taken from live specimens as it is difficult to obtain. It involves holding the bird by its legs, laying it dorsally on top of a ruler with the tip of the tail touching the zero-stop while gently stretching its crown to rest on the ruler. The measurement is read to the bill tip and can generally not be taken on skins.

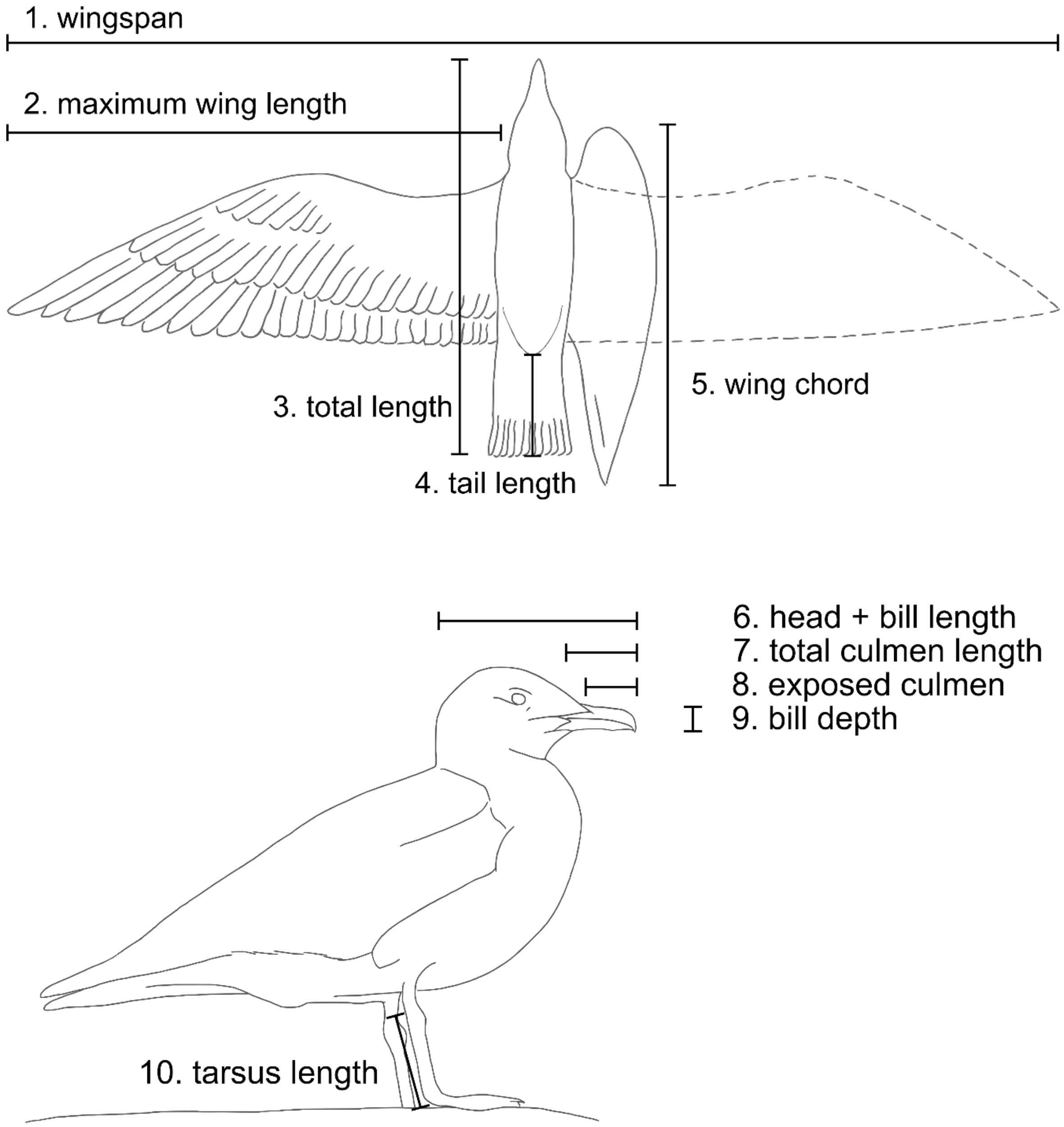

Figure 2-3: Biometric measurements that are of interest for image analysis applications (1-3) differ from those that are routinely collected by ornithologists on live birds or museum specimens (4-10). In addition, terminology differs between disciplines and subfields, with both maximum wing length (2) and wing chord (5) being referred to as "wing length". Illustration: Ruth Walker/ BTO

2.5 Bird sexing

Not only is sexing individuals important ecologically, for establishing life history traits and behaviours between the sexes (Roff, 1993; Seyer et al., 2019), but it is particularly critical in the context of monitoring birds around wind energy installations, where differences in body size has serious effects (Willisch et al., 2013). This generally involves monitoring free-living animals of multiple species, where it impossible to ground truth the biometrics of individual targets and because many bird species, including seabirds and raptors exhibit sexual dimorphism in body size, as well as large inter-individual variation within sexes and across geographical ranges. It can be extremely challenging to sex individuals in the field. Many seabirds and raptors are monomorphic by external appearance (exhibit little or no differences in plumage between the sexes; Boersma and Davies, 1987; Fairbairn and Shine, 1993), and even where species do show sexual dimorphisms, the perception of differences in colouration or other features can be subjective and hard to observe.

2.6 Methods

Here, we quantify the level of natural intra-specific variation in seabird and raptor body size, to highlight how this variation may result in substantial uncertainties and possible biases in inferences about flight heights.

We assessed the body size variation of 20 bird species associated with offshore and onshore wind developments, predominantly in Europe: 14 seabirds – Great Black-backed gull (*Larus marinus*, GB), Lesser Black-backed gull (*Larus fuscus*, LB), Herring Gull (*Larus argentatus*, HG), Kittiwake (*Rissa tridactyla*, KI), Gannet (*Morus bassanus*, GX), Fulmar (*Fulmarus glacialis*), Manx shearwater (*Puffinus puffinus,* MX), Guillemot (*Uria aalge,* GU), Common tern (*Sterna hirundo*, CN), Sandwich tern (*Thalasseus sandvicensis*, TE), Arctic skua (*Stercorarius parasiticus,* AC), Great skua (*Stercorarius skua*, GX), European storm petrel (*Hydrobates pelagicus*), Leach's storm petrel (*Oceanodroma leucorhoa*); and six birds of prey – Red kite (*Milvus milvus*, KT), Kestrel (*Falco tinnunculus*, K.), Buzzard (*Buteo buteo*, BZ), Osprey (*Pandion haliaetus*, OP), Golden eagle (*Aquila chrysaetos*, EA), White-tailed eagle (*Haliaeetus albicilla*, WE). We reviewed two data sources *i*) the scientific literature and *ii*) the British and Irish Ringing Scheme.

For most seabird species, we restricted our literature search to studies of bird populations in countries surrounding UK and European waters, however, for Fulmar, Common tern and Sandwich tern we included data from the US to increase sample sizes. For all raptor species, except Buzzard, there was an absence of measurements for full grown birds, so we were required to use published data from museum specimens (Cramp et al., 1988). Literature searches were conducted on Google Scholar using combinations of species names and the relevant biometric measures (e.g. "wing length") as search terms. For all data derived from skins, we did not apply any shrinkage correction factors since species-specific values were not available for the vast majority of species and universal correction factors are untrustworthy (Eck et al., 2012). In total, we collated biometric data from 58 references for 6 characters (wing length, bill length, bill depth, head plus bill length, tail length, and tarsus length). We did not consider body weight due to its high variability with breeding status and body condition (Croxall, 1995), and because linear measurements are the primary metrics that can be extracted from imagery. We focus on wing chord length (further "wing length") in the main text, as this is the most comparable measure to the body length measure employed by Humphries et al. (2023). In total we analysed wing chord data from 19,560 individual birds from the literature. We filtered out any anomalous wing measurements from the ringing data by using a conservative threshold of ±3 times the standard deviation, and removed all species, age and sex groupings which had less than five records (excluding WE and OP for which there were very few records). In total, we analysed wing length data from 110,306 individual birds from the ringing data (having removed duplicate records of the same bird).

For the literature data, we took the assignments of adult and non-adult age classes from the original sources, whereas for the ringing data, we classified birds as adults as soon as their plumage was akin to the adult plumage – given that features required for exact ageing, such as leg or bill colouration, are likely to not be visible in relatively low-resolution camera imagery. In gull species and Gannets moult patterns are more noticeable in the field, and so for KI we classified adults as birds of ≥ 3 calendar years (cy); for GB, LB and HG adults were ≥ 4 cys, and for GX they were ≥ 5 cy. For all other species, we classified adults as birds that were ≥ 1cy. Fledged birds below these ages were classified as non-adults. Unfledged birds were excluded from the analysis.

For each literature source, we recorded the means and standard deviations of the six biometric characters listed above, for each age and sex class (where possible). For the ringing data, we calculated the mean wing length and standard deviation for each age and sex class for each species. To integrate the data across studies and obtain summary statistics for pooled age or sex classes, we calculated weighted means μ as

$$\mu = \sum_i p_i \mu_i.$$

and weighted estimates of population standard deviations σ as

$$\sigma = \sqrt{\sum_i p_i \sigma_i^2 + \sum_i p_i \mu_i^2 - \left(\sum_i p_i \mu_i\right)^2}$$

where μ_i and σ_i are the means and standard deviations from a study and age or sex class, and p_i are weights calculated as the relevant relative sample size (e.g. single-study sample size divided by total sample size across all studies). We then obtained summary coefficients of variation (CV) from the weighted means and standard deviations, as a means of comparing body size variability across metrics and species (Figs. 2-6 - 2-11). For the purposes of visualisation and simulation, we used the weighted means and standard deviations to approximate the expected population distributions assuming normal distributions (Figs. 2-12, 2-13, 2-17, 2-18).

Since males and females are generally indistinguishable on imagery, we aggregated the sexes into a single category per age class, assuming an even sex ratio (Conover and Hunt Jr., 1988; Parkes, 1989). We therefore obtained overall estimates of body size coefficients of variation for visually distinguishable groups within each species (Figs. 2-10, 2-15). Furthermore, to show the effect of possible species misidentification, we also aggregated measurements of the four gull species (GB, LB, HG, KI) into a single category per age class, assuming even sex ratios and even occurrence frequencies (Figs. 2-11, 2-16).

2.7 Results

Overall, we found intra-specific coefficients of variation in biometrics to range between 2-10%, except for tail length and bill length in common terns which was >10% in certain age/sex classes (Figs. 2-4 – 2-11). Wing length and head plus bill length showed less variation in general than other linear measurements (Figs. 1, 4). Non-adult birds often showed more size variation than adult birds, but not in the four gull species combined (Figs. 2-11, 2-16). The distribution plots show that sexual dimorphism was pronounced in gulls, fulmars, skuas and raptors, but not in shearwaters, Gannet, guillemots, terns or petrels. The interspecific size differences in gull species resulted in very large size variations for the hypothetical populations of “unidentified gulls” in both data sources (wing length CV 15-17%, Figs. 2-11, 2-16).

The observed variation in natural body size imposes expected ranging errors on the scale of 5-10% of the target range for birds with known species identity (Figure 2-19).

2.7.1 Literature data

Figure 2-4: Coefficients of variation of wing chord (where data available) for visually distinguishable groups of 14 seabird species (AC, Arctic skua; CN, common tern; F., Fulmar; GB, Great Black-backed gull; GU, Guillemot; GX, Gannet; HG, Herring Gull; KI, Kittiwake; LB, Lesser Black-backed gull; MX, Manx shearwater; NX, Great skua; TE, Sandwich tern; TM, European storm petrel; TL, Leach’s storm petrel), and six raptor species (pink: BZ, Buzzard; EA, Golden eagle; K., Kestrel; KT, Red kite; OP, Osprey; WE, White-tailed eagle).

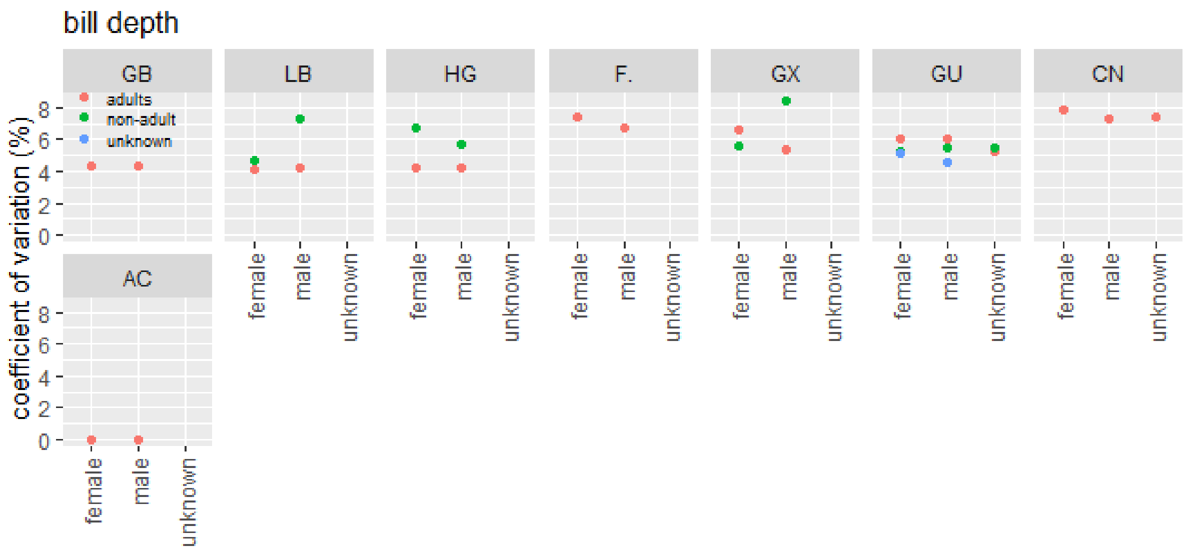

Figure 2-5: Coefficients of variation of bill depth measurements (where data available) for visually distinguishable groups of 14 seabird species (see Figure 2-4 for species codes)

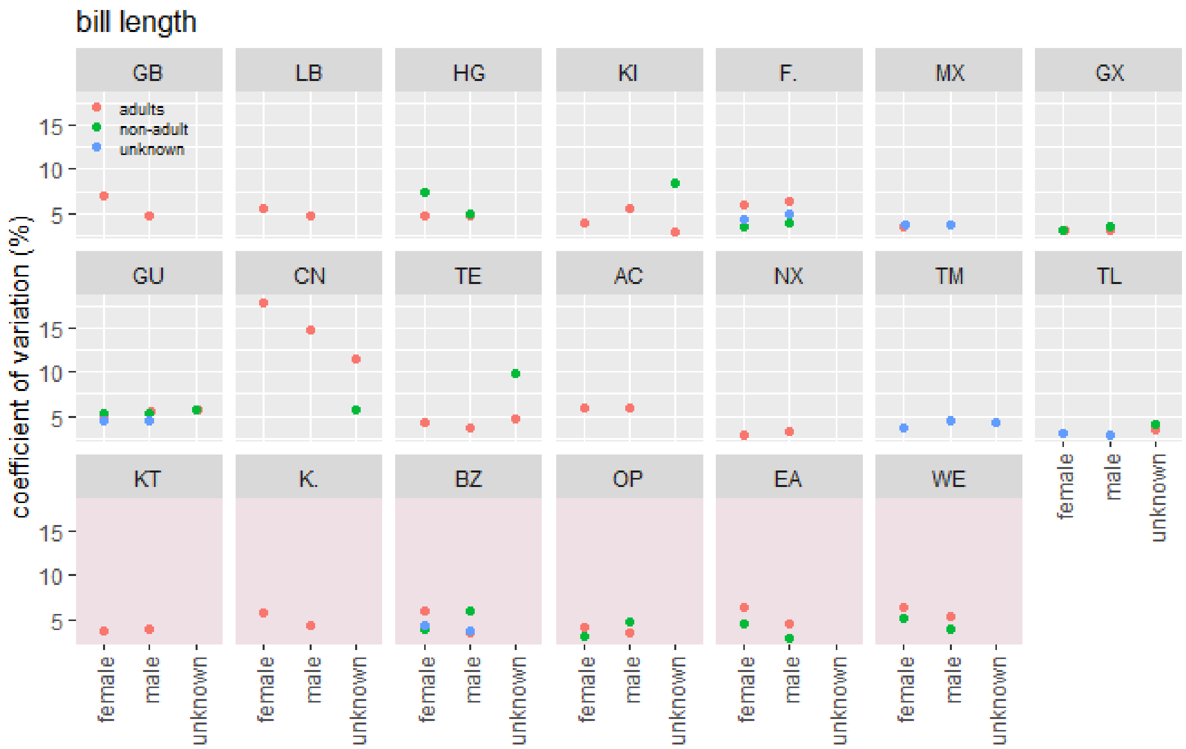

Figure 2-6: Coefficients of variation of bill length measurements (where data available) for visually distinguishable groups of 14 seabird species (see Figure 2-4 for species codes).

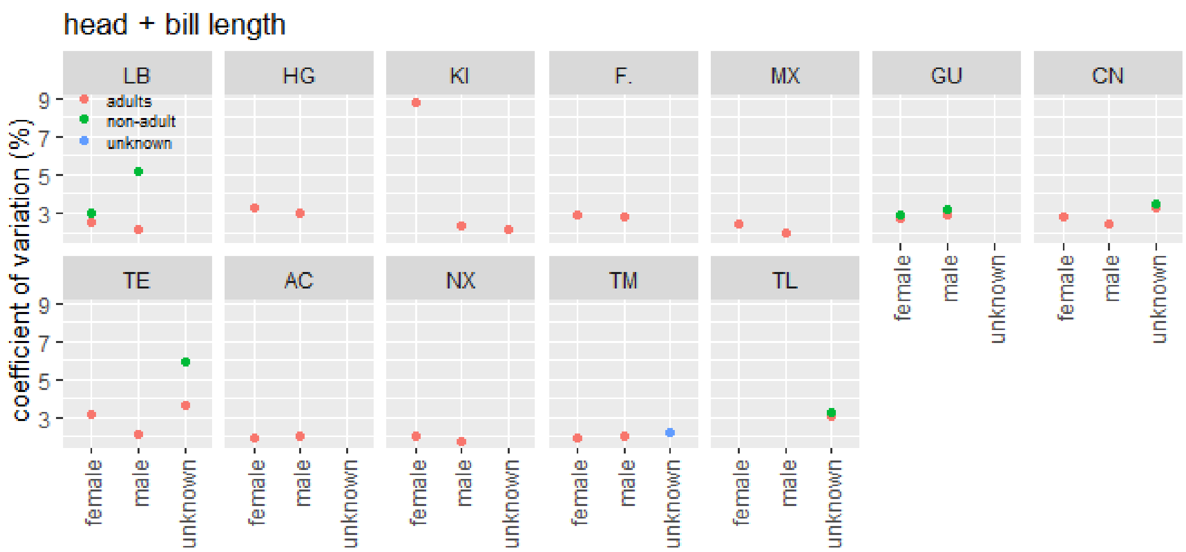

Figure 2-7: Coefficients of variation of head + bill length measurements (where data available) for visually distinguishable groups of 14 seabird species (see Figure 2-4 for species codes).

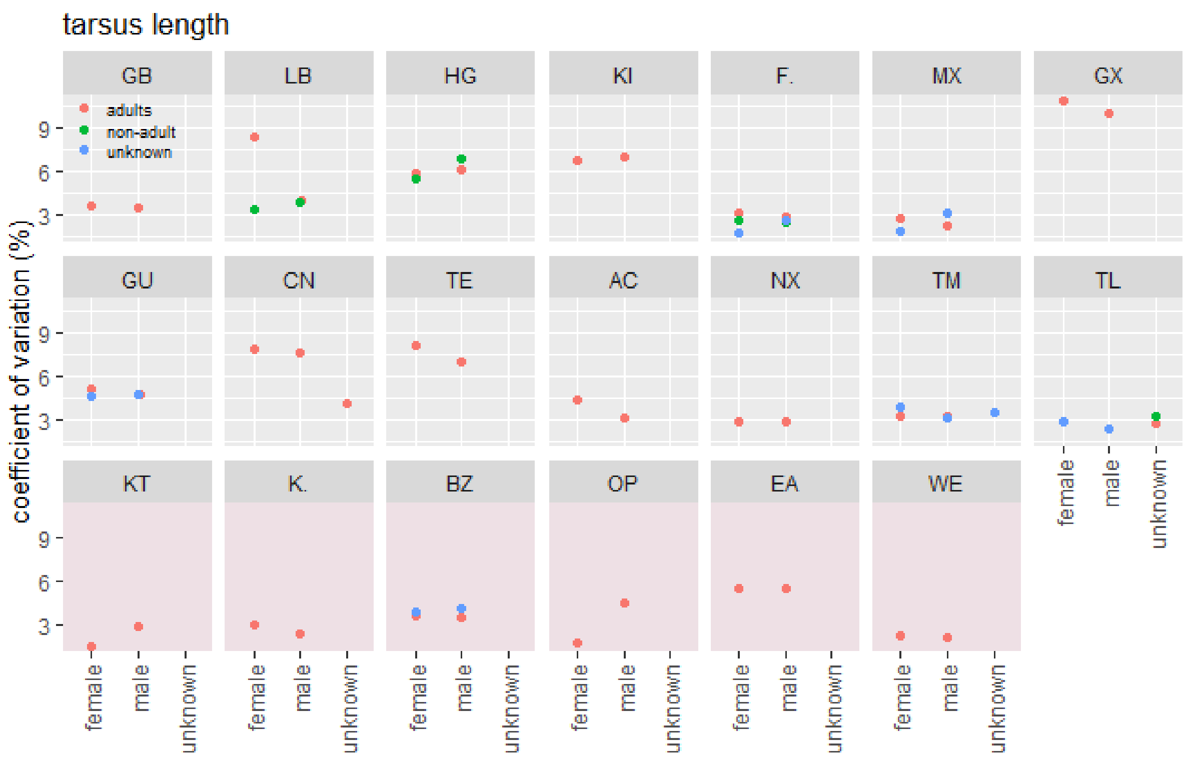

Figure 2-8: Coefficients of variation of tarsus length measurements (where data available) for visually distinguishable groups of 14 seabird species (see Figure 2-4 for species codes).

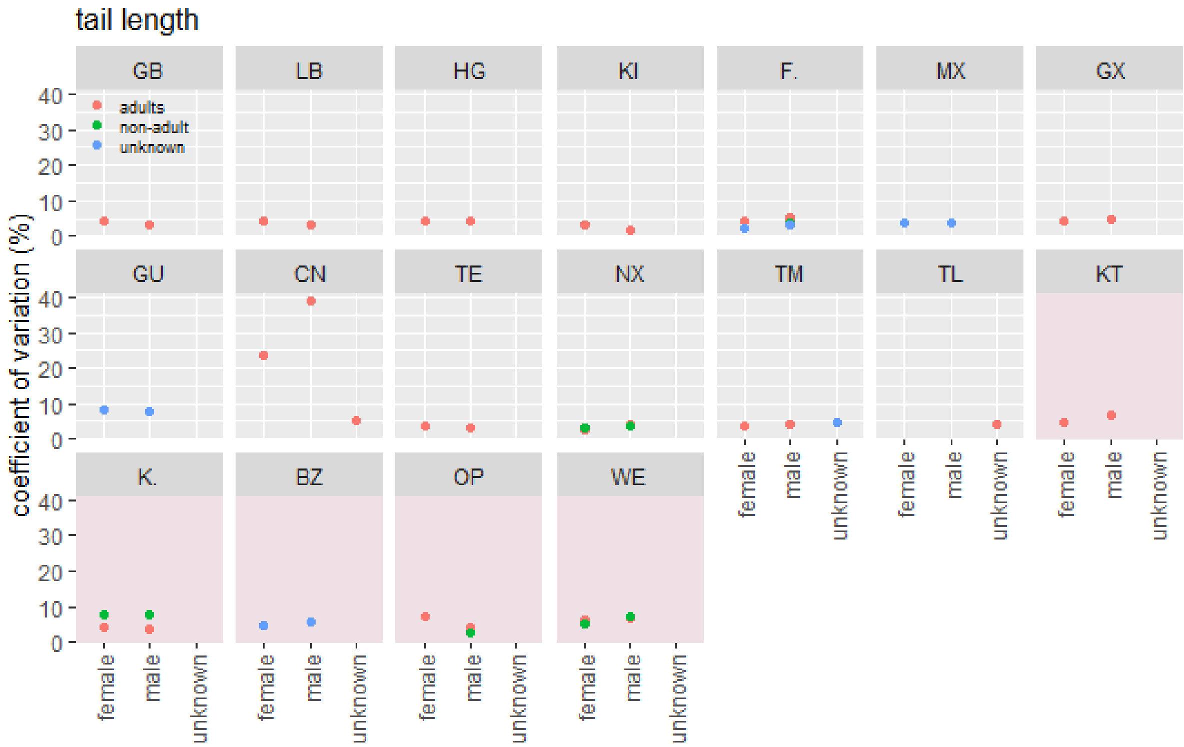

Figure 2-9: Coefficients of variation of tail length measurements (where data available) for visually distinguishable groups of 14 seabird species (see Figure 2-4 for species codes).

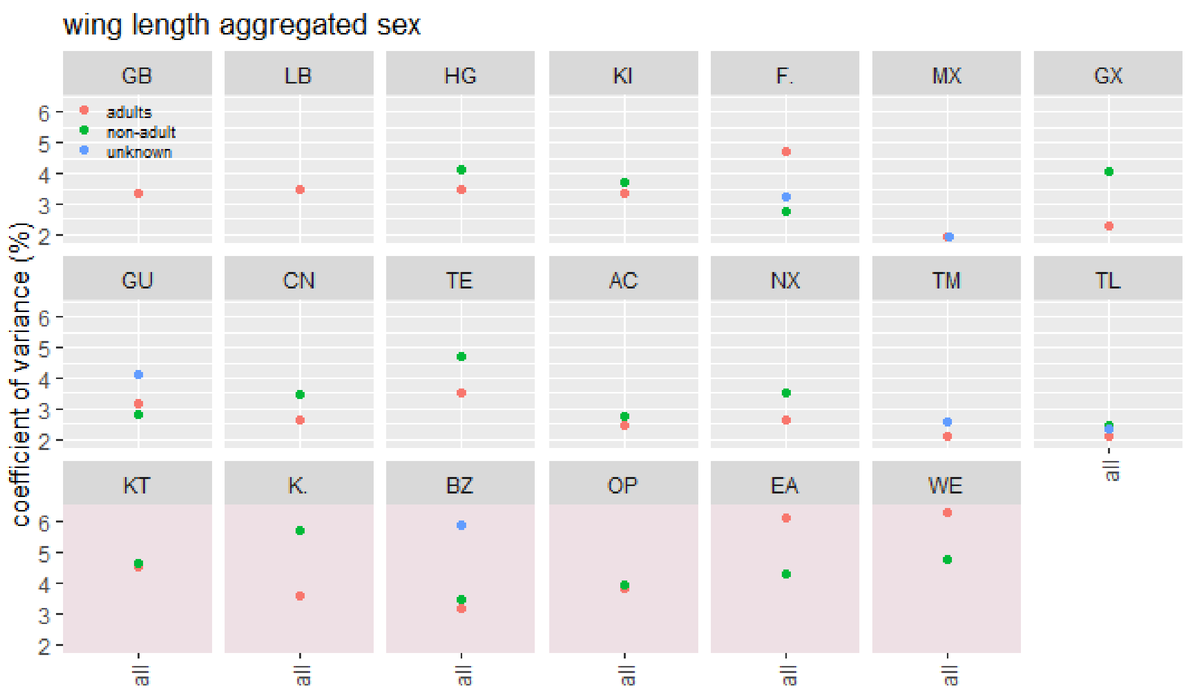

Figure 2-10: Coefficients of variation of wing length measurements, with sexes pooled, for visually distinguishable groups of 14 seabird species (AC, Arctic skua; CN, common tern; F., Fulmar; GB, Great Black-backed gull; GU, Guillemot; GX, Gannet; HG, Herring Gull; KI, Kittiwake; LB, Lesser Black-backed gull; MX, Manx shearwater; NX, Great skua; TE, Sandwich tern; TM, European storm petrel; TL, Leach's storm petrel), and six raptor species (pink: BZ, Buzzard; EA, Golden eagle; K., Kestrel; KT, Red kite; OP, Osprey; WE, White-tailed eagle)

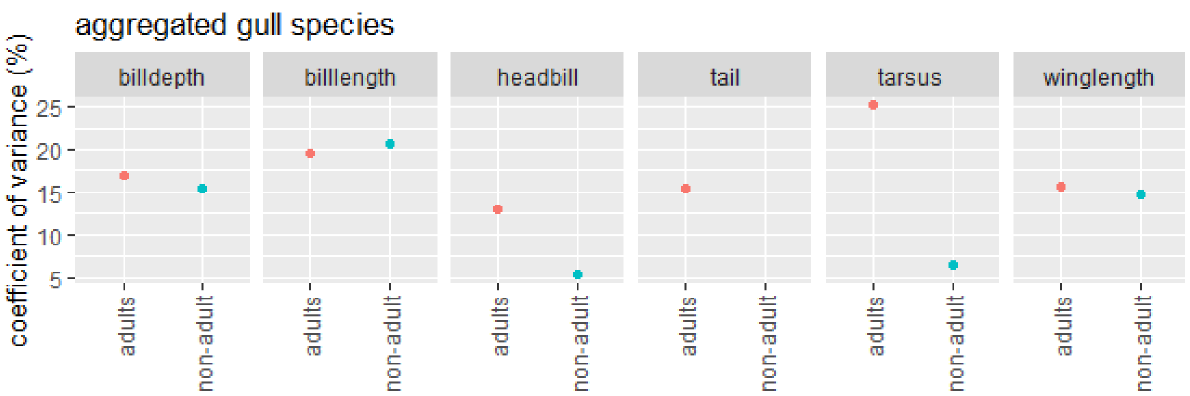

Figure 2-11: Coefficients of variation of six linear measurements, with sexes and gull species pooled, for visually distinguishable groups of four gull species (GB, Great Black-backed gull; HG, Herring Gull; KI, Kittiwake; LB, Lesser Black-backed gull).

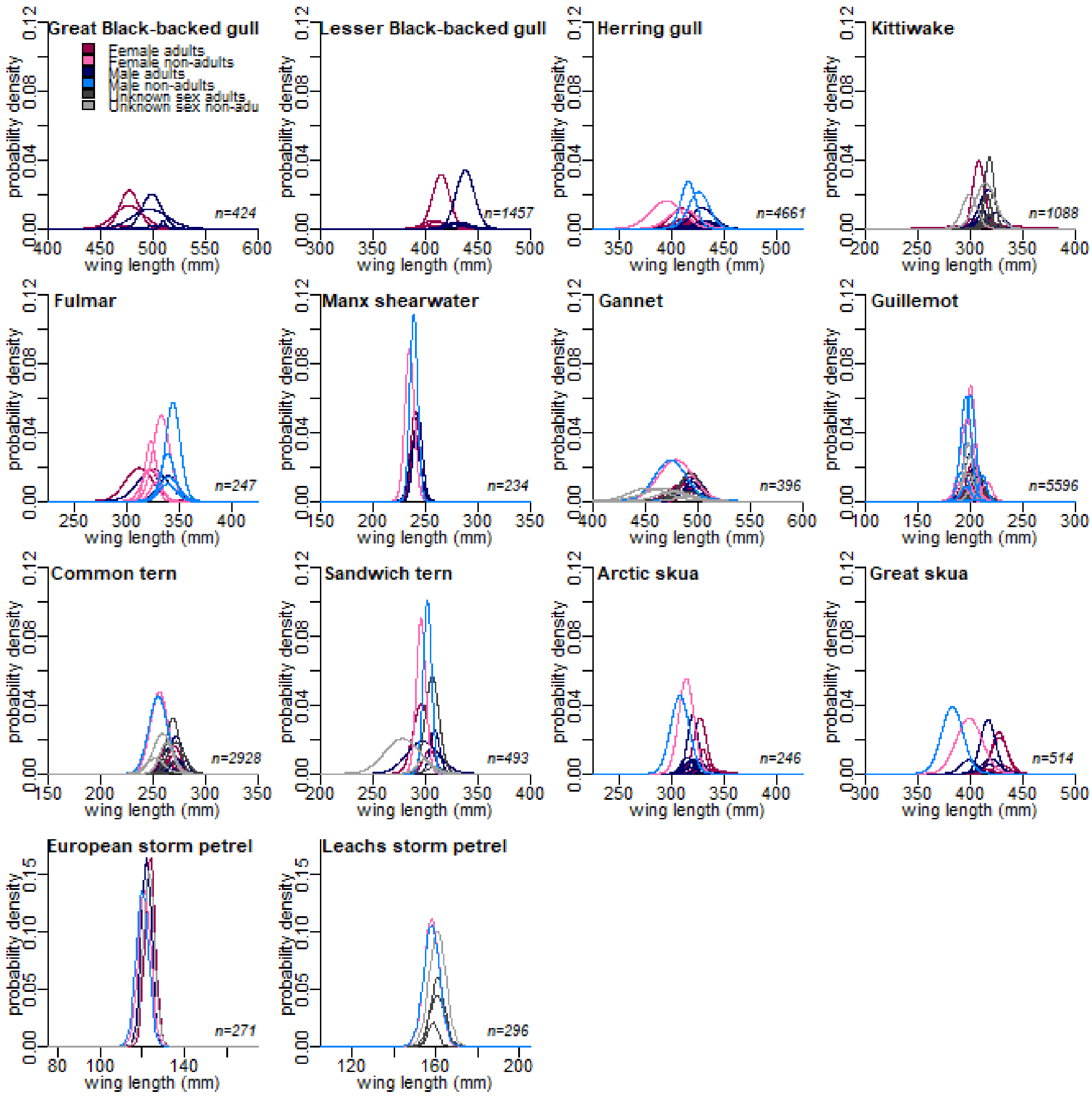

Figure 2-12: Wing length (measured as wing chord) distributions of 14 seabird species from published means and standard deviations of age/sex classes in each population, scaled by their respective sample sizes. Literature sources are given in Table S1.

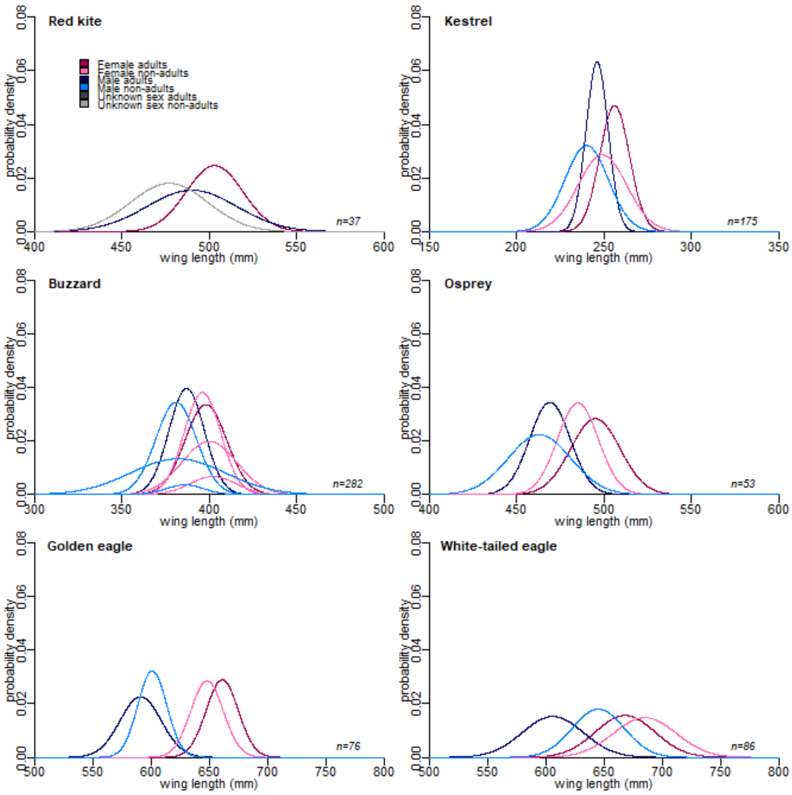

Figure 2-13: Wing length distributions (measured as wing chord) of six raptor species from published means and standard deviations of each population, scaled by their respective sample sizes. Literature sources are given in Table S1.

2.7.2 Ringing data

Figure 2-14: Coefficients of variation of wing length measurements (measured as wing chord) for visually distinguishable groups of 14 seabird species (AC, Arctic skua; CN, common tern; F., Fulmar; GB, Great Black-backed gull; GU, Guillemot; GX, Gannet; HG, Herring Gull; KI, Kittiwake; LB, Lesser Black-backed gull; MX, Manx shearwater; NX, Great skua; TE, Sandwich tern; TM, European storm petrel; TL, Leach's storm petrel), and six raptor species (BZ, Buzzard; EA, Golden eagle; K., Kestrel; KT, Red kite; OP, Osprey; WE, White-tailed eagle).

Figure 2-15: Coefficients of variation of wing length measurements (measured as wing chord), with sexes pooled, for visually distinguishable groups of 14 seabird species (AC, Arctic skua; CN, common tern; F., Fulmar; GB, Great Black-backed gull; GU, Guillemot; GX, Gannet; HG, Herring Gull; KI, Kittiwake; LB, Lesser Black-backed gull; MX, Manx shearwater; NX, Great skua; TE, Sandwich tern; TM, European storm petrel; TL, Leach's storm petrel), and six raptor species (pink: BZ, Buzzard; EA, Golden eagle; K., Kestrel; KT, Red kite; OP, Osprey; WE, White-tailed eagle).

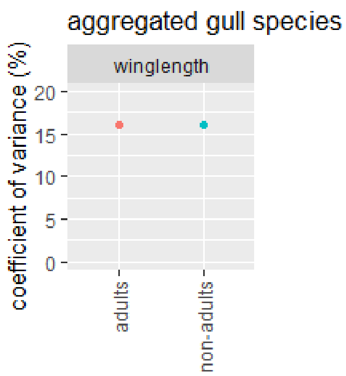

Figure 2-16: Coefficients of variation of six linear measurements, with sexes and gull species pooled, for visually distinguishable groups of four gull species (GB, Great Black-backed gull; HG, Herring Gull; KI, Kittiwake; LB, Lesser Black-backed gull).

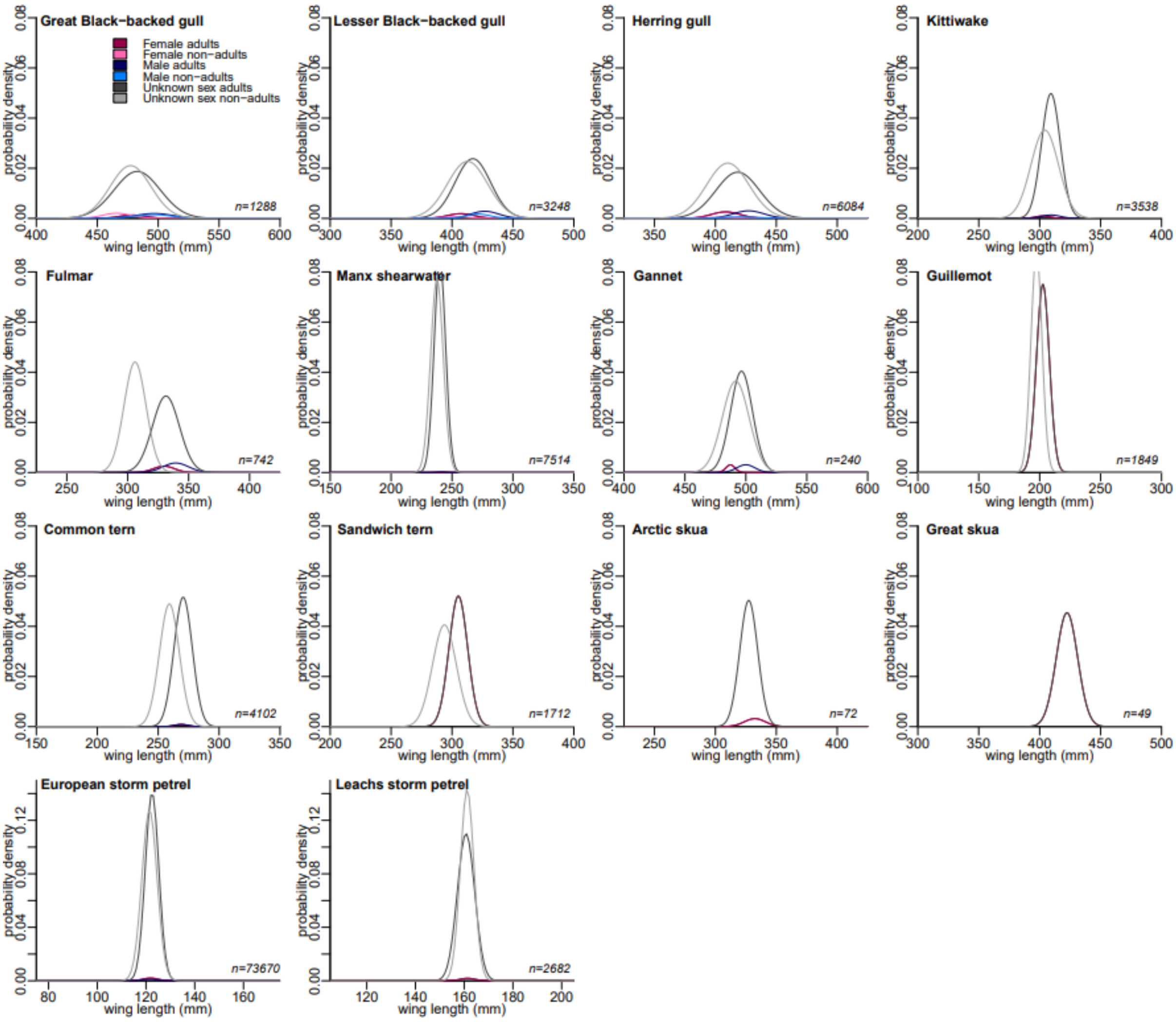

Figure 2-17: Wing length distributions (measured as wing chord) of 14 seabird species and six raptor species from the means and standard deviations of ringing data, scaled by their respective sample sizes.

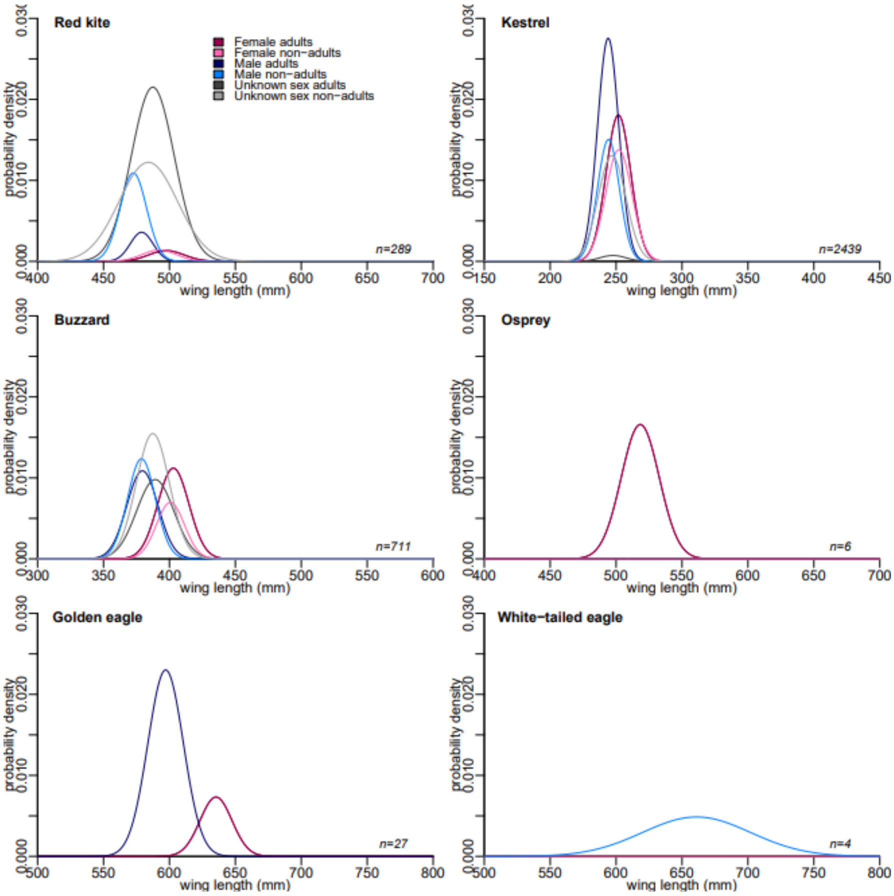

Figure 2-18: Wing length distributions (measured as wing chord) of six raptor species from the means and standard deviations of ringing data, scaled by their respective sample sizes.

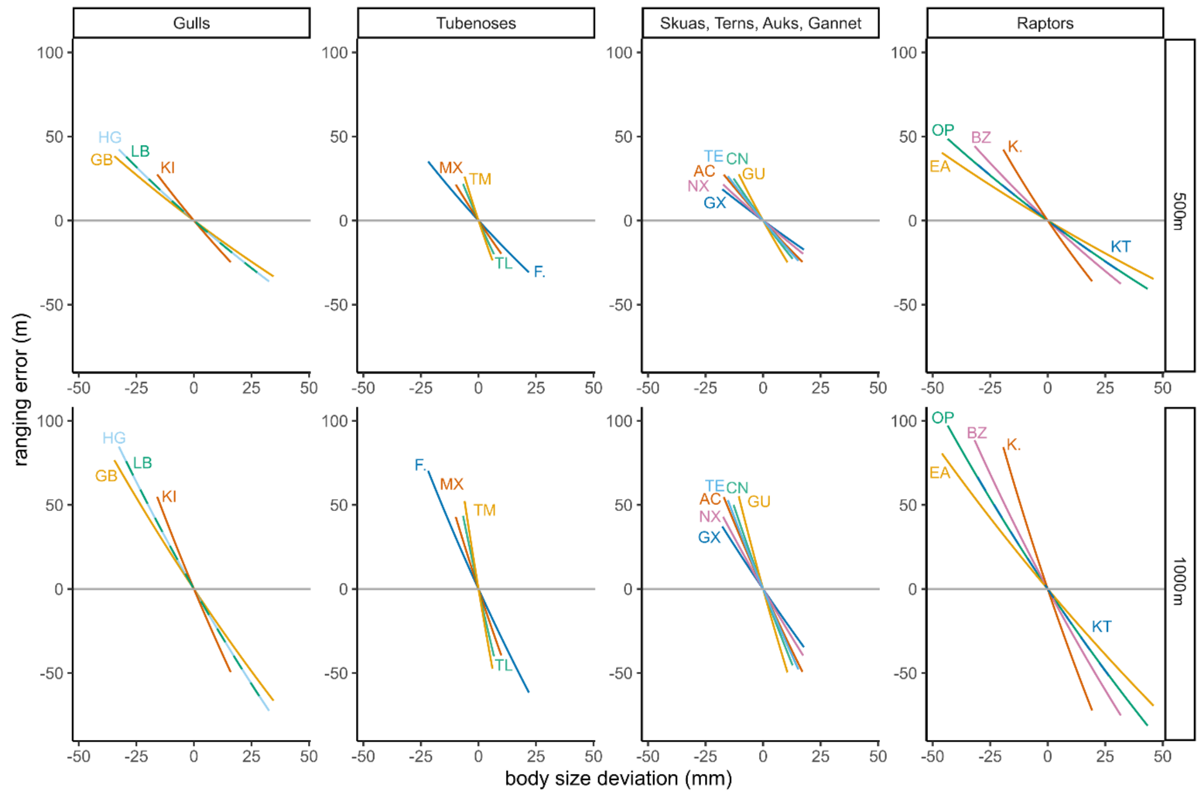

Figure 2-19: : Single-camera photogrammetry ranging errors are caused by the mismatch between mean body size for a species and the actual body size of an individual. For the seabird and raptor species investigated here they amount to 5%-10% of the true target range and affect smaller than average individuals more heavily than larger than average ones. Lines for each species are drawn for the 95% interval of body sizes, as represented by wing chord measurements from ringing data. (AC, Arctic skua; CN, common tern; F., Fulmar; GB, Great Black-backed gull; GU, Guillemot; GX, Gannet; HG, Herring Gull; KI, Kittiwake; LB, Lesser Black-backed gull; MX, Manx shearwater; NX, Great skua; TE, Sandwich tern; TM, European storm petrel; TL, Leach's storm petrel; BZ, Buzzard; EA, Golden eagle; K., Kestrel; KT, Red kite; OP, Osprey; WE, White-tailed eagle).

2.8 Discussion

Most, if not all, methods currently used to reconstruct bird trajectories are affected by sampling or measurement errors, or both (Feather et al., in press). Quantifying these errors as well as propagating them into any inferences about three-dimensional space-use remains an important research priority to ensure bird monitoring data can fulfil its intended purpose, e.g. in offshore regulatory contexts (Searle et al., 2023).

We demonstrate above that natural body size variation within species is non-negligible (2-10% CV), and this provides a fundamental challenge to how precisely positions and trajectories of free-flying birds can be estimated using single-camera approaches.

The observed levels of natural body size variation impose expected ranging errors on the scale of 5-10% of the target range. This translates to range errors on the order of 50-100 metres – roughly the equivalent of the blade length of many current offshore turbines - at target ranges of interest in offshore monitoring applications (e.g. inter-turbine distances) and may hinder achieving robust inferences at the desired ecological scales.

Our estimates of ranging errors are based on measurements of living specimens, as routinely collected by field ornithologists, in particular wing chord, skeletal and bill measurements. However, such measurements are not directly applicable to most real-world image analysis applications. Unfortunately measures that are easily obtained from images, such as total wingspan and total length are difficult, if not impossible to collect on living birds. Ornithological handbooks do provide wingspan and/or total length measures for many species, typically obtained from dead birds or via approximations, but these are often without source attribution and typically as ranges without useable statistical measures of spread. Although the literature features various approximations to get from single-wing measures to wingspan (e.g. Baumgart et al., 2021, Fu et al., 2023), but these generally require secondary measurements such as body width or feather dimensions which are similarly challenging to obtain.

Single-camera technologies are attractive from an operational perspective, given the relative simplicity of their deployment, and their scalability, which may provide cost-effective monitoring where other technologies currently fall short.

To ensure the data from single-camera technologies form a valid and reliable evidence base it is thus imperative that such single-camera approaches — like all methods for determining ranges — are adequately validated, and that any uncertainty in their measurements is correctly quantified and propagated into subsequent analyses (Searle et al., 2023). This extends to derived quantities such as flight height, for which in downward looking systems the error is similar or equal to the range error (cf. Boersch-Supan et al., 2023), whereas for forward-looking systems the range error propagates as a function of the elevation angle of the target relative to the camera (cf. Fig. 5-19).

Other existing observation technologies are in principle robust to imperfect knowledge of seabird body size and have been demonstrated to produce range estimates with considerably lower uncertainty. This includes stereo photogrammetry (e.g. Brighton et al., 2019, Prinsloo et al., 2021), and methods employing direct range estimation of birds such as handheld laser rangefinding equipment (Harwood, Perrow and Berridge, 2018). Although, these approaches are not without their own implementation, operational and/or analytical challenges, they may complement single-camera approaches and aid the calibration and/or validation of bird trajectory reconstruction in the face of natural body-size variation and other sources of uncertainty.

2.9 References

Alerstam T, Rosén M, Bäckman J, Ericson PGP, Hellgren O. 2007. Flight speeds among bird species: Allometric and phylogenetic effects. Plos Biol. 5(8):e197. https://doi.org/10.1371/journal.pbio.0050197

Allison TD, Diffendorfer JE, Baerwald EF, Beston JA, Drake D, Hale AM, Hein CD, Huso MM, Loss SR, Lovich JE. 2019. Impacts to wildlife of wind energy siting and operation in the United States. Issues in Ecology. 21(1):2-18.

Ballester C, Dupont SM, Corbeau A, Chambert T, Duriez O, Besnard A. 2024. A standardized protocol for assessing the performance of automatic detection systems used in onshore wind power plants to reduce avian mortality. J Environ Manage. 354:120437. https://doi.org/10.1016/j.jenvman.2024.120437

Band W. 2012. Using a collision risk model to assess bird collision risks for offshore wind farms. Report to strategic ornithological support services programme, project SOSS-02.

Barrett R, Peterz M, Furness R, Durinck J. 1989. The variability of biometric measurements. Ringing Migration. 10(1):13-16. https://doi.org/10.1080/03078698.1989.9676001

Baumgart SL, Sereno PC, Westneat MW. 2021. Wing shape in waterbirds: morphometric patterns associated with behavior, habitat, migration, and phylogenetic convergence. Integrative Organismal Biology. 3(1): obab011. https://doi.org/10.1093/iob/obab011.

Bergeron P. 2007. Parallel lasers for remote measurements of morphological traits. The Journal of wildlife management. 71(1):289-292. https://doi.org/10.2193/2006-290

Boersch-Supan PH, Brighton CH, Thaxter CB, Cook ASCP. 2024. Natural body size variation in seabirds provides a fundamental challenge for flight height determination by single-camera photogrammetry: a comment on Humphries et al. (2024). Mar Biol. 171:122. https://doi.org/10.1007/s00227-024-04396-4

Boersma PD, Davies EM. 1987. Sexing monomorphic birds by vent measurements. The Auk. 104(4):779-783. https://doi.org/10.1093/auk/104.4.779

Brighton CH, Kloepper LN, Harding CD, Larkman L, McGowan K, Zusi L, Taylor GK. 2022. Raptors avoid the confusion effect by targeting fixed points in dense aerial prey aggregations. Nature Communications. 13. https://doi.org/10.1038/s41467-022-32354-5

Brighton CH, Taylor GK. 2019. Hawks steer attacks using a guidance system tuned for close pursuit of erratically manoeuvring targets. Nature Communications. 10. https://doi.org/10.1038/s41467-019-10454-z

Chilson PB, Bridge E, Frick WF, Chapman JW, Kelly JF. 2012. Radar aeroecology: Exploring the movements of aerial fauna through radio-wave remote sensing. The Royal Society. https://doi.org/10.1098/rsbl.2012.0384

Cleasby IR, Wakefield ED, Bearhop S, Bodey TW, Votier SC, Hamer KC. 2015. Three-dimensional tracking of a wide-ranging marine predator: Flight heights and vulnerability to offshore wind farms. J Appl Ecol. 52(6):1474-1482. https://doi.org/10.1111/1365-2664.12529

Conover MR, Hunt Jr GL. 1988. Sex ratios of North American gulls based on museum collections. Colon Waterbird.38-45. https://doi.org/10.2307/1521168

Cramp S, Simmons K, Ferguson-Lees I, Gillmor R, Hollom P, Hudson R, Nicholson E, Ogilvie M, Olney P, Voous K. 1988. Birds of Europe, the Middle East and North Africa. VV New York.

Croxall JP. 1995. Sexual size dimorphism in seabirds. Oikos.399-403. https://doi.org/10.2307/3545964

Croxall JP, Butchart SH, Lascelles B, Stattersfield AJ, Sullivan B, Symes A, Taylor P. 2012. Seabird conservation status, threats and priority actions: A global assessment. Bird Conserv Int. 22(1):1-34. https://doi.org/10.1017/S0959270912000020

Dawson A. 2005. The scaling of primary flight feather length and mass in relation to wing shape, function and habitat. Ibis. 147(2):283-292. https://doi.org/10.1111/j.1474-919x.2005.00400.x

Eck S, Fiebig J, Fiedler W, Heynen I, Nicolai B, Töpfer T, Van den Elzen R, Winkler R, Woog F. 2012. Measuring birds–Vögel vermessen. Ostrich. 83:117.

Ewins PJ. 1985. Variation of black guillemot wing lengths post-mortem and between measurers. Ringing Migration. 6(2):115-117. https://doi.org/10.1080/03078698.1985.9673867

Fairbairn J, Shine R. 1993. Patterns of sexual size dimorphism in seabirds of the southern hemisphere. Oikos.139-145. https://doi.org/10.2307/3545319

Fijn RC, Gyimesi A. 2018. Behaviour related flight speeds of sandwich terns and their implications for wind farm collision rate modelling and impact assessment. Environ Impact Asses. 71:12-16. https://doi.org/10.1016/j.eiar.2018.03.007

Fu HY, Su M, Chu JJ, Margaritescu A, Claramunt S. 2023. New methods for estimating the total wing area of birds. Ecology and Evolution. 13: e10480. https://doi.org/10.1002/ece3.10480.

Furness RW, Wade HM, Masden EA. 2013. Assessing vulnerability of marine bird populations to offshore wind farms. J Environ Manage. 119:56-66. https://doi.org/10.1016/j.jenvman.2013.01.025

Hanssen F, May R, Nygård T. 2020. High-resolution modeling of uplift landscapes can inform micrositing of wind turbines for soaring raptors. Environ Manage. 66:319-332. https://doi.org/10.1007/s00267-020-01318-0

Humphries GR, Fail T, Watson M, Houghton W, Peters-Grundy R, Scott M, Thomson R, Keogan K, Webb A. 2023. Aerial photogrammetry of seabirds from digital aerial video images using relative change in size to estimate flight height. Mar Biol. 170(2):18. https://doi.org/10.1007/s00227-022-04161-5

Jaquet N. 2006. A simple photogrammetric technique to measure sperm whales at sea. Mar Mammal Sci. 22(4):862-879. https://doi.org/10.1111/j.1748-7692.2006.00060.x

Jenni L, Winkler R. 1989. The feather-length of small passerines: A measurement for wing-length in live birds and museum skins. Bird Study. 36(1):1-15. https://doi.org/10.1080/00063658909476996

Largey N, Cook AS, Thaxter CB, McCluskie A, Stokke BG, Wilson B, Masden EA. 2021. Methods to quantify avian airspace use in relation to wind energy development. Ibis. 163(3):747-764. https://doi.org/10.1111/ibi.12913

Lyon BE. 1994. A technique for measuring precocial chicks from photographs. The Condor. 96(3):805-809. https://doi.org/10.2307/1369486

Masden E, Cook A. 2016. Avian collision risk models for wind energy impact assessments. Environ Impact Asses. 56:43-49. https://doi.org/10.1016/j.eiar.2015.09.001

Masden EA, Cook AS, McCluskie A, Bouten W, Burton NH, Thaxter CB. 2021. When speed matters: The importance of flight speed in an avian collision risk model. Environ Impact Asses. 90:106622. https://doi.org/10.1016/j.eiar.2021.106622

McGregor RM, King S, Donovan CR, Caneco B, Webb A. 2018. A stochastic collision risk model for seabirds in flight. Marine Scotland Science.

Nicholls AJ, Barker M, Armitage M, Votier S. 2022. Review of seabird monitoring technologies for offshore wind farms. SBMon project report to ORJIP Offshore Wind Programme. Carbon Trust, London.

Parkes KC. 1989. Sex ratios based on museum collections: a caution. Colon Waterbird. 12(1):130-131. https://doi.org/10.2307/1521326

Prinsloo ND, Postma M, Ryan PG, Coetzee M De Bruyn P. 2021. Estimating bird flight height using 3-D photogrammetry. Journal of Zoology, 314, 174-186. https://doi.org/10.1111/jzo.12879

Ralph CJ. 1993. Handbook of field methods for monitoring landbirds. Pacific Southwest Research Station.

Roff D. 1993. Evolution of life histories: Theory and analysis. Springer Science Business Media.

Scherz JP. 1974. Errors in photogrammetry. Int J Rock Mech Min Sci Geomech Abstr. 11:A179.

Searle K, O'Brien S, Jones E, Cook A, Trinder M, McGregor R, Donovan C, McCluskie A, Daunt F, Butler A. 2023. A framework for improving treatment of uncertainty in offshore wind assessments for protected marine birds. Ices J Mar Sci.fsad025. https://doi.org/10.1093/icesjms/fsad025

Seyer Y, Gauthier G, Bernatchez L, Therrien J-F. 2019. Sexing a monomorphic plumage seabird using morphometrics and assortative mating. Waterbirds. 42(4):380-392. https://doi.org/10.1675/063.042.0403

Shepard EL, Williamson C, Windsor SP. 2016. Fine-scale flight strategies of gulls in urban airflows indicate risk and reward in city living. Philosophical Transactions of the Royal Society B: Biological Sciences. 371(1704):20150394. https://doi.org/10.1098/rstb.2015.0394

Strickland M, Arnett E, Erickson W, Johnson D, Johnson G. 2011. Comprehensive guide to studying wind energy/wildlife interactions. National wind coordinating collaborative website.

Svensson L. 1992. Identification guide to European passerines. Stockholm.

Thaxter CB, Buchanan GM, Carr J, Butchart SHM, Newbold T, Green RE, Tobias JA, Foden WB, O'Brien S, Pearce-Higgins JW. 2017. Bird and bat species' global vulnerability to collision mortality at wind farms revealed through a trait-based assessment. P Roy Soc B-Biol Sci. 284(1862). https://doi.org/10.1098/rspb.2017.0829

Thaxter CB, Ross-Smith VH, Cook A. 2016. How high do birds fly?: A review of current datasets and an appraisal of current methodologies for collecting flight height data; literature review. British Trust for Ornithology.

Williams M. 2017. From field to museum tray: Shrinking of the mallard Anas platyrhynchos. Wildfowl. 67(67):123-129.

Willisch CS, Marreros N, Neuhaus P. 2013. Long-distance photogrammetric trait estimation in free-ranging animals: A new approach. Mamm Biol. 78(5):351-355. https://doi.org/10.1016/j.mambio.2013.02.004

Wilson RE, McCracken KG. 2008. Specimen shrinkage in cinnamon teal. The Wilson Journal of Ornithology. 120(2):390-392. https://doi.org/10.1676/07-040.1

Winker K. 1998. Suggestions for measuring external characters of birds. Ornitol Neotrop. 9:23-30.

2.10 Supplementary tables

Supplementary data tables are available at https://doi.org/10.5281/zenodo.14975387.

3 Assessment of image-based 3D tracking systems for monitoring bird movements

3.1 Introduction

While the benefits of transitioning to wind energy generation are clear, in terms of reducing CO_2 emissions, the rapid growth of the sector has increasing ecological implications for avian wildlife. Birds are susceptible to lethal collisions with wind turbines, and many at risk species are already high conservation priorities. Determining the true level of impact depends on accurate and reliable methods to quantify flight parameters – in particular, vertical habitat use relative to turbine height. The various monitoring technologies are suited to different purposes, and there are often trade-offs between cost and image resolution, affecting how well birds can be identified. Recent approaches have used sensor-based technologies (Chilson et al., 2012; Thaxter et al., 2016; Cole et al., 2019), but currently, there is no systematic effort to quantify or validate collision rate data (Ballester et al., 2024).

Image-based approaches offer the potential to collect large volumes of flight data on temporal and spatial variability, which can be deployed in remote locations and create a permanent record of observations. Image-based flight height determination generally falls into one of two approaches: Single-image (also known as single-camera or mono-vision) for better scalability and economy, and stereo-image (stereo-vision) for better accuracy. Single-image photogrammetry systems rely on auxiliary information on the range to the target (Bergeron, 2007, Jaquet, 2006, Lyon, 1994) or information about target size, i.e. species biometric measurements (Willisch et al., 2013). Any uncertainty in these auxiliary data add to the sources of uncertainty in the image processing. Stereo-based systems do not require any auxiliary information about the target, and instead rely on the principle of motion parallax and the geometry of the camera pair or camera array. The accuracy of the reconstruction requires high image resolutions, temporal synchronization of devices, and exact measurements of the setup and camera extrinsic parameters. Sub-metre spatial accuracy at ranges of up to several hundred metres can be achieved (Brighton et al., 2022; Brighton and Taylor, 2019; Prinsloo et al., 2021), but the increased instrumentation costs, increased image processing requirements, and (in some cases) reduced deployability (e.g. on moving survey platforms) have meant this approach has received little attention. Furthermore, setting up cameras at offshore turbines is very space-limited, meaning that optimal geometries for stereo configurations cannot be achieved.

Therefore, to assess the feasibility of mono- and stereo- vision systems, under compromised field conditions, we aimed to validate a camera system deployed at an offshore wind farm in Aberdeen Bay. We undertook onshore field trials, by replicating the setup as closely as possible to the offshore configuration (which was restricted to a very short baseline distance between the cameras) and used a drone as our reference object. By comparing the drone's 3D position as measured by the onboard GPS, to the 3D position obtained from semi-automated image analysis (for both mono and stereo) techniques, we can quantify the reconstruction error and hence the accuracy and precision of the AI systems. The mono reconstructions were created using the Spoor AI workflow. The stereo-reconstructions were based on an established photogrammetry technique (Brighton et al., 2022, Brighton and Taylor, 2019) which has previously achieved accuracies on the order of cm when tracking birds at ranges <500m. Because this approach relies heavily on manual image processing – which is not viable for large or long-term datasets – we use it as a 'best model' for testing the performance of the AI software.

3.2 Methods

3.2.1 Drone flights

We recorded video of a DJI Mavic 3E drone flying through a series or predefined transects on a cricket field in Ekebergparken, Oslo, Norway, on June 20th, 2023. The drone flight path was programmed so that it covered a large area inside the camera field of view. Transects consisted of (i) horizontal "lawnmower" left to right flight paths at different heights and distances from the cameras (Fig. 3-1A) and (ii) 30-40m length "tetrahedrons" at different distances to the cameras (Fig. 3-1B). The drone's position was recorded by the onboard RTK-enabled GPS.

We used two different pairs of high-definition video cameras: (i) "Axis" (Axis Q1798-LE Network) and (ii) "Avigilon" (Avigilon H5 Pro) to reconstruct the three-dimensional flight paths of the drone, setting the camera lenses to their widest zoom setting. For the Axis cameras, we recorded 25 Hz video at 3,840 × 2,160 pixels (4k), whereas for the Avigilon cameras, we recorded 9.63 Hz video at 7,143 × 4,624 pixels (7k). In turn, each camera pair was set in stereo configuration with a baseline distance of 4 m. The cameras were mounted on tripods and were adjusted to the same height using a tape. The cameras were turned on and left to record for the duration of each drone trial.

Trial (internal code)	Camera	Transect	Drone GPS	Mono recon automated	Stereo recon automated	Stereo recon manual
1 (05)	Avigilon	Lawnmower	Yes – RTK	Yes	Partial	Partial
2 (06,07)	Avigilon	Tetrahedral	Yes – RTK	Yes	Partial	No
3 (08)	Axis	Lawnmower	Yes (no RTK)	Yes	Yes	Yes
4 (09)	Axis	Tetrahedral	Yes (no RTK)	Yes	Yes	Yes

Table 3-1. Summary of drone trials and 3D reconstructions.

3.2.2 Photogrammetry for mono camera tracking

Mono camera track reconstructions were provided by Spoor.

3.2.3 Photogrammetry for stereo camera tracking

We synchronized the videos using the DLTdv8 video tracking toolbox (Hedrick, 2008) in MATLAB R2023b (MathWorks Inc., Natick, MA) by matching the arm motions of a fieldworker clapping their hands above their head in front of the cameras. We then applied the relevant frame offset to synchronize them to the nearest frame. As the cameras' shutters were not electronically synchronized, this post hoc procedure can only guarantee synchronization of the frames to within ±0.02 s at the 25 Hz frame rate, and to within ±0.05 s at the 9.63 Hz frame rate. The fully manual method also used the DLTdv8 toolbox to identify the 2D pixel coordinates of the drone in both cameras within a pair, manually tracking the visual centre of the drone's body from the point at which it appeared in both cameras until it was too distant to see (too few pixels). For the automated method, we used the 2D coordinates obtained from the Spoor AI mono tracking outputs. The auto-derived 2D data did not retain the frame numbers containing no drone detections, so we were required to interpolate the frame number to restore the correct timings. Under both methods, we were able to track the drone transects up to ~450m from the cameras.

We calibrated the cameras by matching 20-30 points across both frames, including background features and points on the hovering drone (when it was effectively stationary), to cover as much of the capture volume as possible. To get a better scaling for each calibration, we took distance measurements in the field using a laser range finder (Leica) between two clearly identifiable points in each camera view. Using MATLAB's Optimisation Toolbox R2023b and custom-written code (Walker et al., 2009), the known camera intrinsic parameters (sensor size, focal length, frame rate), extrinsic parameters (spacing, tilt) and the 2D positions of the calibration points were used to solve the camera collinearity equations by means of a nonlinear least squares bundle adjustment. The optimisation routine identified the jointly optimal estimates of the position and pose of the cameras, and spatial coordinates of the calibration points, by minimising the sum of the squared reprojection error of the matched image points.

We assumed no lens distortion and no offset of the principal point with respect to the centre of the camera sensor. We used focal lengths of 10,237 pixels for the Axis cameras and 13,465 pixels for the Avigilon cameras. These values were initially estimated using the manufacturers specifications (focal length in pixels = focal length (mm) / sensor width (mm) * image width (px)) and then refined in an optimisation during the calibration process. To test the sensitivity of our focal length estimation for the Axis cameras, we ran reconstructions using focal lengths of 1,850 – 11,000 pixels (in 250 px increments) and looked at the effects on the estimated lengths of two scale references (calibration pole and tetrahedron length). We found the minimum distance discrepancies between the estimated reference lengths and the measured lengths at 10,250 px, which is very close to our optimised value of 10,237 px (Fig. S3-1).

The calibration reconstructs the spatial coordinates of the matched image points in a Cartesian coordinate system, aligned with the sensor axes of one of the cameras. To compare to the real-world drone GPS data, we were required to transform the spatial coordinates of the drone into an Earth axis system in which the z axis was vertical. To do so, we held a pole with a post-level in front of the cameras and filmed and reconstructed its image coordinates. The resulting vertical reference was used to calculate the real-world rotation that was applied to the data.

3.2.4 Aligning the datasets

To spatially match the drone GPS data and all the photogrammetry reconstructions, we converted the drone data to a Universal Transverse Mercator (UTM) coordinate system. All datasets were then shifted to be relative to ground level and the array centre (midpoint between the cameras). Due to some rotational offset between the GPS and photogrammetry datasets, we were required to rotate the GPS data using the measured bearing of the left camera (-168°). To temporally match the drone GPS data to the photogrammetry reconstructions, we plotted the X axis against time in milliseconds and manually identified when the maximum value of the first 'peak' occurred in the datasets and shifted the photogrammetry data so the peaks aligned with the GPS data. For the Axis cameras, we upsampled the GPS data from 5 Hz (200 ms) to 0.04 Hz (40 ms). For the Avigilon cameras, because of the decimal frame rate, we had to crop the data to match the start and end points of the GPS data (by visually matching the X axis vs time plot) and create a new timestamp which matched the length of the drone data in milliseconds.

We report the root mean square (RMS) reprojection error (standard deviation of the residuals, or prediction errors) as a check on the accuracy of the calibrations and reconstructions. We assess the mean distance error for each camera system and reconstruction type, defined as the mean Euclidean distance between corresponding 3D drone points and photogrammetry points. We also assess the mean distance error for the X, Y and Z axes separately, and for different distance bands (50m increments away from the cameras) and height bands (10m increments above ground level).

3.3 Results

For the Avigilon cameras, we found that the raw videos frequently dropped frames during processing and therefore the time-synch between cameras drifted considerably over the duration of the recordings. This affected our ability to reconstruct the drone position in stereo, using either the automated or manual methods, since these methods depend on accurate time-synching. Therefore, we only report the results from the Axis cameras for the stereo analyses, results are reported as means and their standard errors.

3.3.1 Stereo system - manual vs auto

The manual and auto systems used the same calibration – for the Axis cameras, the RMS reprojection error of the calibration was 0.20 px. Using manual 2D tracking of the drone, the RMS error of the lawnmower transect was 0.35 px, and 1.20 px for the tetrahedral transect. Using automated 2D tracking of the drone, the RMS error of the lawnmower transect was 0.98 px, and 2.98 px for the tetrahedral transect (Table S3-1). The sub-pixel reprojection error that we achieved in the calibrations is appropriate to the method. The higher reprojection error of the drone flights reconstructions is also expected because any spatiotemporal error in the matching of points across camera frames will manifest as reprojection error in the reconstructions.

Overall, both systems performed better in the X (lateral) and Z (height) planes, but the Y (depth) plane had much larger distance errors (Table 3-2). For the lawnmower transect, the 3D mean distance error of the fully manual technique was 8.24±6.48 m (mean±SD), compared to 10.84±7.40 m for the semi-automated. For each axis separately, the mean 2D distance error of the manual and semi-auto techniques respectively, were 1.97±6.48 m and 2.20±1.17 m for the X axis; 7.15±6.89 m and 10.21±7.62 m for the Y axis; and 2.19±1.19 m and 1.65±0.89 m for the Z axis. For the tetrahedral transect, the fully manual technique generally performed marginally better, with a 3D mean distance error of 6.70±2.04 m, compared to 11.85±5.54 m for the semi-automated, though the latter was inflated by the Y axis error. For each axis separately, the mean 2D distance error of the manual and semi-auto techniques respectively, were 1.67±0.49 m and 1.32±0.62 m for the X axis; 5.20±2.58 m and 11.3±5.66 m for the Y axis; and 3.36±1.01 m and 2.60±1.20 m for the Z axis. These differences were of little practical importance, and statistically significant at larger ranges only ($p<0.05$; linear mixed model for correlated data; Table S3-2)

3.3.2 Mono systems – left vs right

Since the analyses of the left and right cameras are independent, we can compare the precision of corresponding data from the left and right cameras. Overall, the mono systems also performed better in the X (lateral) and Z (height) planes, with the Y (depth) plane showing much larger distance errors, except for one case (Avigilon left camera, tetrahedral transect) where the X axis error was greatest (Table 3-3). In addition, the deviations from the mean were generally much greater than seen in the stereo systems. For the Axis lawnmower transect, the 3D mean distance error of the left camera data was 20.28±11.37 m, compared to 20.90±31.23 m for the right camera data. For each axis separately, the mean 2D distance error of the left camera and right camera respectively, were 4.92±4.31 m and 8.01±5.93 m for the X axis; 18.83±11.71 m and 17.38±31.45 m for the Y axis; and 3.32±1.77 m and 2.83±3.76 m for the Z axis. For the Axis tetrahedral transect, the 3D mean distance error of the left camera data was 20.86±29.89 m, compared to 27.54±9.28 m for the right camera data. For each axis separately, the mean 2D distance error of the left camera and right camera respectively, were 11.2±5.51 m and 5.13±3.71 m for the X axis; 30.8±39.6 m and 33.7±13.0 m for the Y axis; and 2.53±5.32 m and 2.30±1.85 m for the Z axis. For the Avigilon lawnmower transect, the 3D mean distance error of the left camera data was 33.7±39.7 m, compared to 34.5±12.8 m for the right camera data. For each axis separately, the mean 2D distance error of the left camera and right camera respectively, were 11.2±5.51 m and 5.13±3.71 m for the X axis; 30.8±39.6 m and 33.7±13.0 m for the Y axis; and 2.53±5.32 m and 2.30±1.85 m for the Z axis. For the Avigilon tetrahedral transect, the 3D mean distance error of the left camera data was 22.37±7.37 m, compared to 19.48±20.15 m for the right camera data. For each axis separately, the mean 2D distance error of the left camera and right camera respectively, were 16.71±6.22 m and 2.58±3.80 m for the X axis;

10.97±7.94 m and 17.92±20.43 m for the Y axis; and 6.34±3.64 m and 3.59±3.59 m for the Z axis.

3.3.3 Effect of distance on error

We compared the distance error between the reconstructed drone tracks and the drone GPS tracks for different height and depth bands for the stereo data (Figs. 3-2A, B) and the mono data (Figs. 3-3A, B). For the Axis stereo system, the Y axis error shows a strong increase as the depth and the height increases, while the X and Z axis errors remain relatively constant. For the Axis mono system, the Y axis error is consistently high over all depth and height bands, while the X axis and Z axis errors remain relatively constant or increase only slightly. For the Avigilon mono system, the Y axis error shows a strong increase as the depth increases but is consistently high over all height bands. For the X axis, the error remains constant across depths or increases slightly, except for the right camera lawnmower transect, where the error decreases with height. The Z axis errors remain relatively constant for height and depth but show a slight increase at the maximum depth distances (>400m).

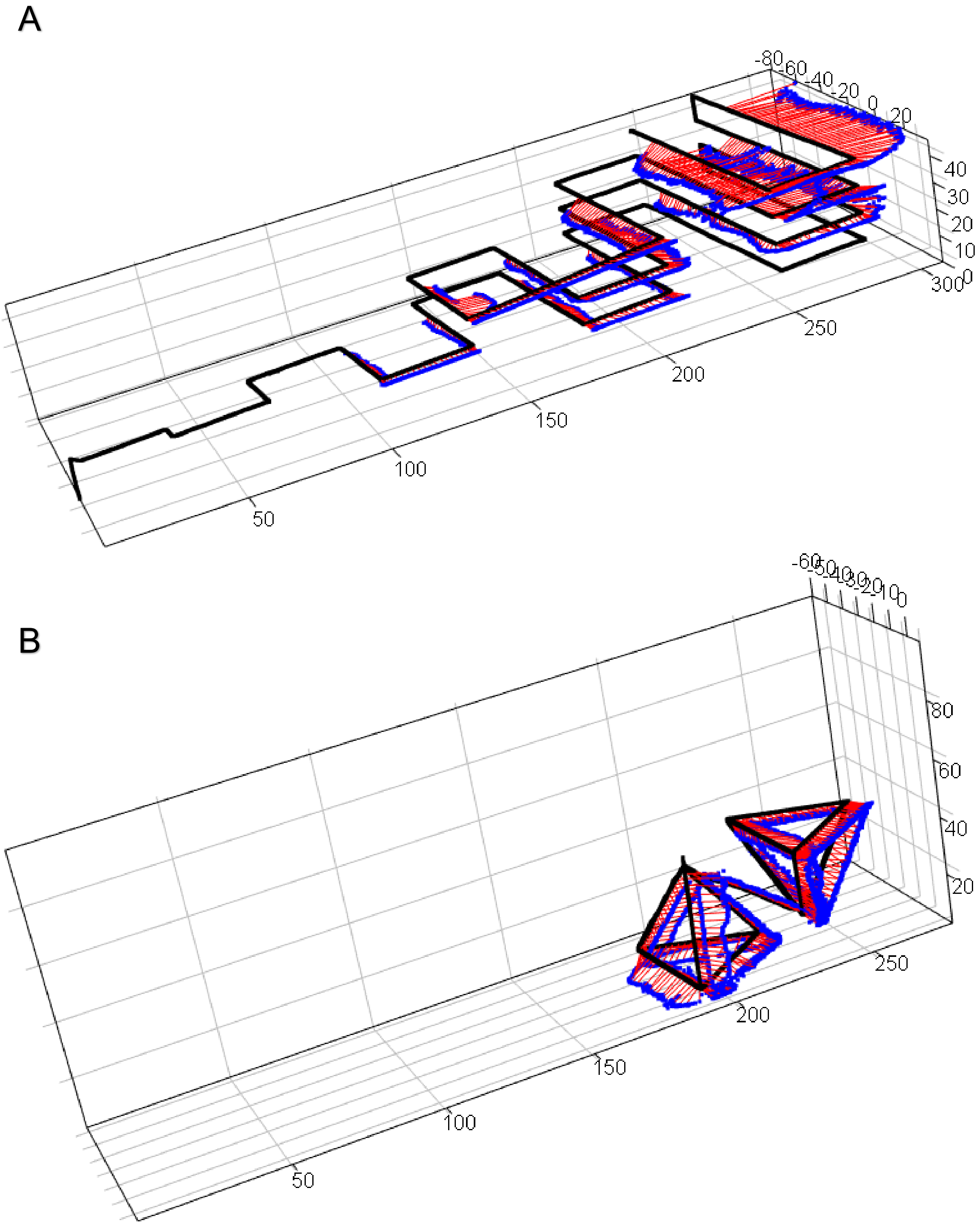

Figure 3-1. Shows the lawnmower transect (A) and tetrahedral transect (B) flown by the drone (black) during the Axis camera trials, and the manual stereo-reconstruction points (blue). Red lines connect every 10th corresponding point to show the position offset. Axes are distance in metres from the midpoint between the cameras.

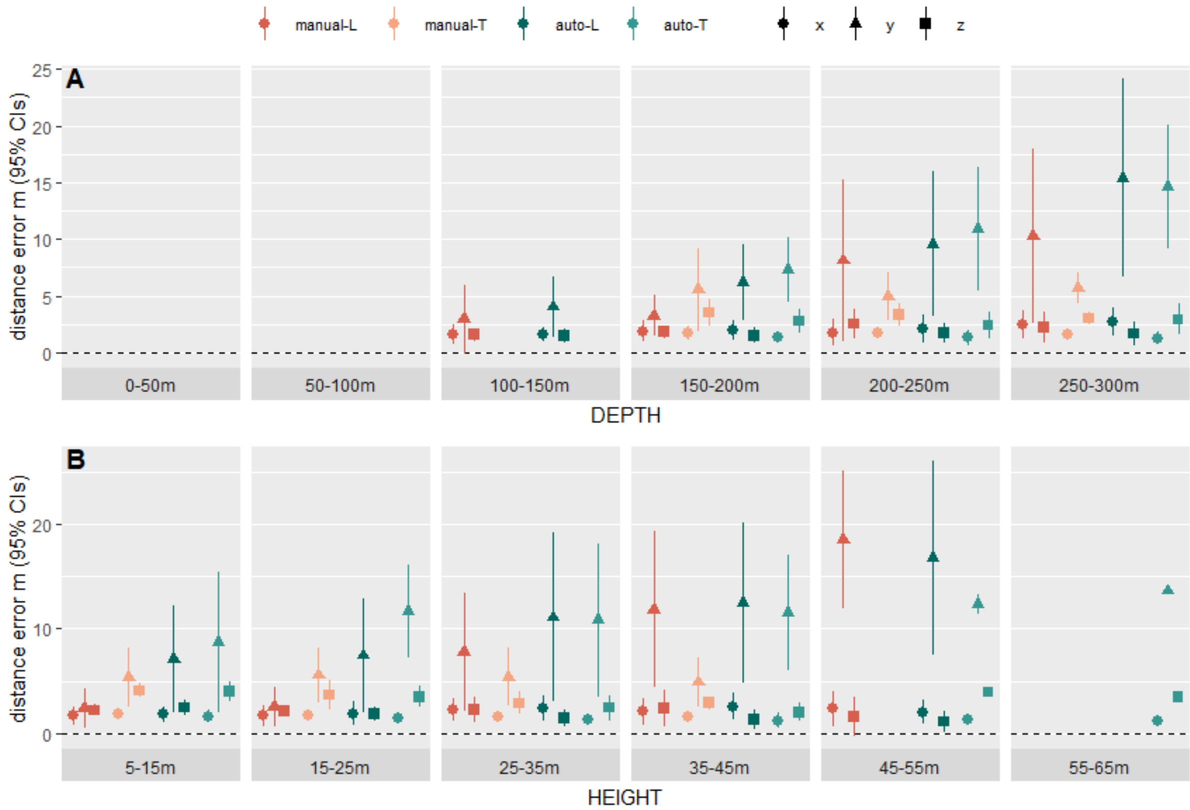

Figure 3-2A. Stereo camera reconstructions. Distance error of the Axis stereo-reconstructions relative to the drone GPS; comparing the automated 2D tracking method (auto) and manual 2D tracking method (manual) for lawnmower transects (L) and tetrahedral transects (T). The errors are shown for each axis (x,y and z) with 95% confidence intervals, and are divided into (A) longitudinal and (B) vertical distance bands.

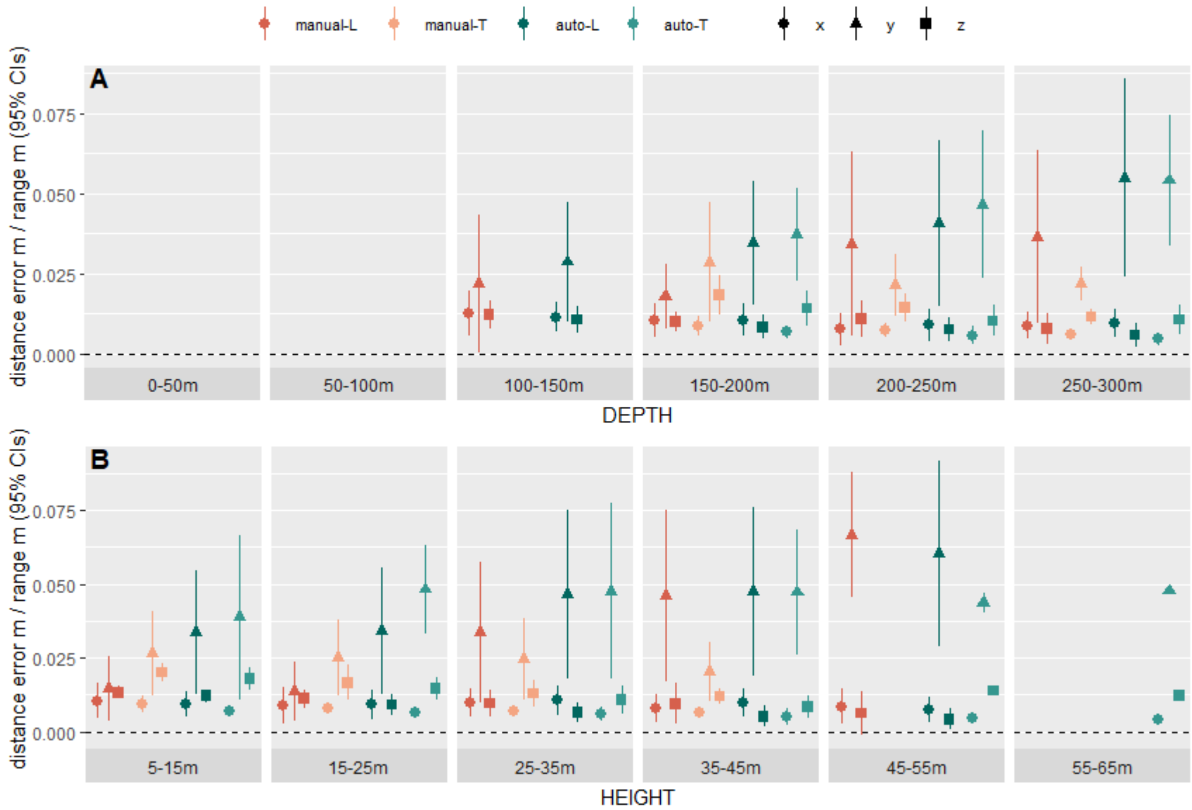

Figure 3-2B. Stereo camera reconstructions. Relative distance error (error divided by the straight-line distance from the midpoint between cameras), of the Axis stereo-reconstructions relative to the drone GPS; comparing the automated 2D tracking method (auto) and manual 2D tracking method (manual) for lawnmower transects (L) and tetrahedral transects (T). The errors are shown for each axis (x,y and z) with 95% confidence intervals, and are divided into (A) longitudinal and (B) vertical distance bands.

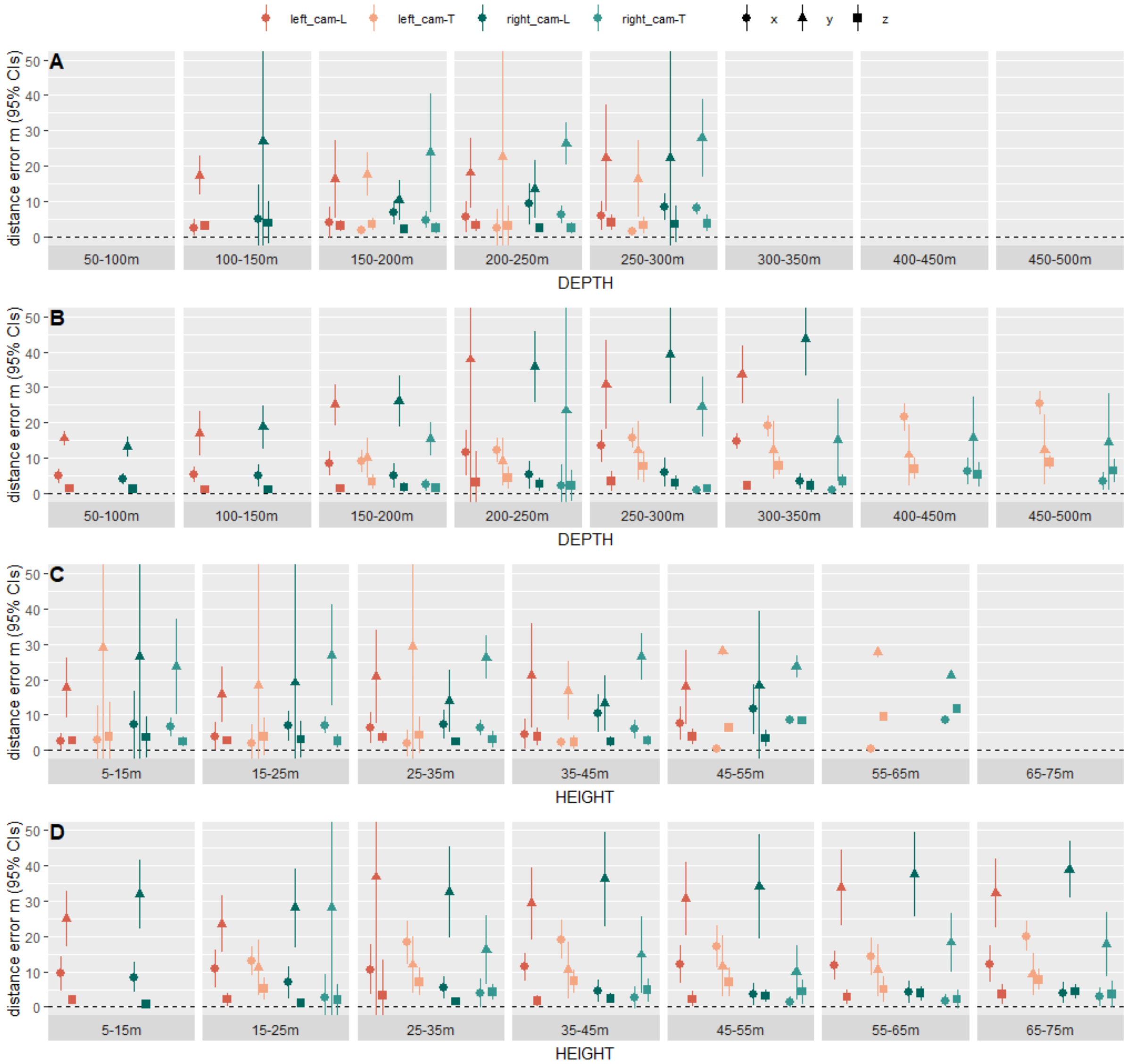

Figure 3-3A. Mono camera reconstructions. Distance error of the Axis (A,C) and Avigilon (B,D) mono reconstructions relative to the drone GPS; comparing the left and right cameras for lawnmower transects (L) and tetrahedral transects (T). The errors are shown for each axis (x,y and z) with 95% confidence intervals, and are divided into (A,B) longitudinal and (C,D) vertical distance bands. NB some of the CIs extend off the plotting limits.

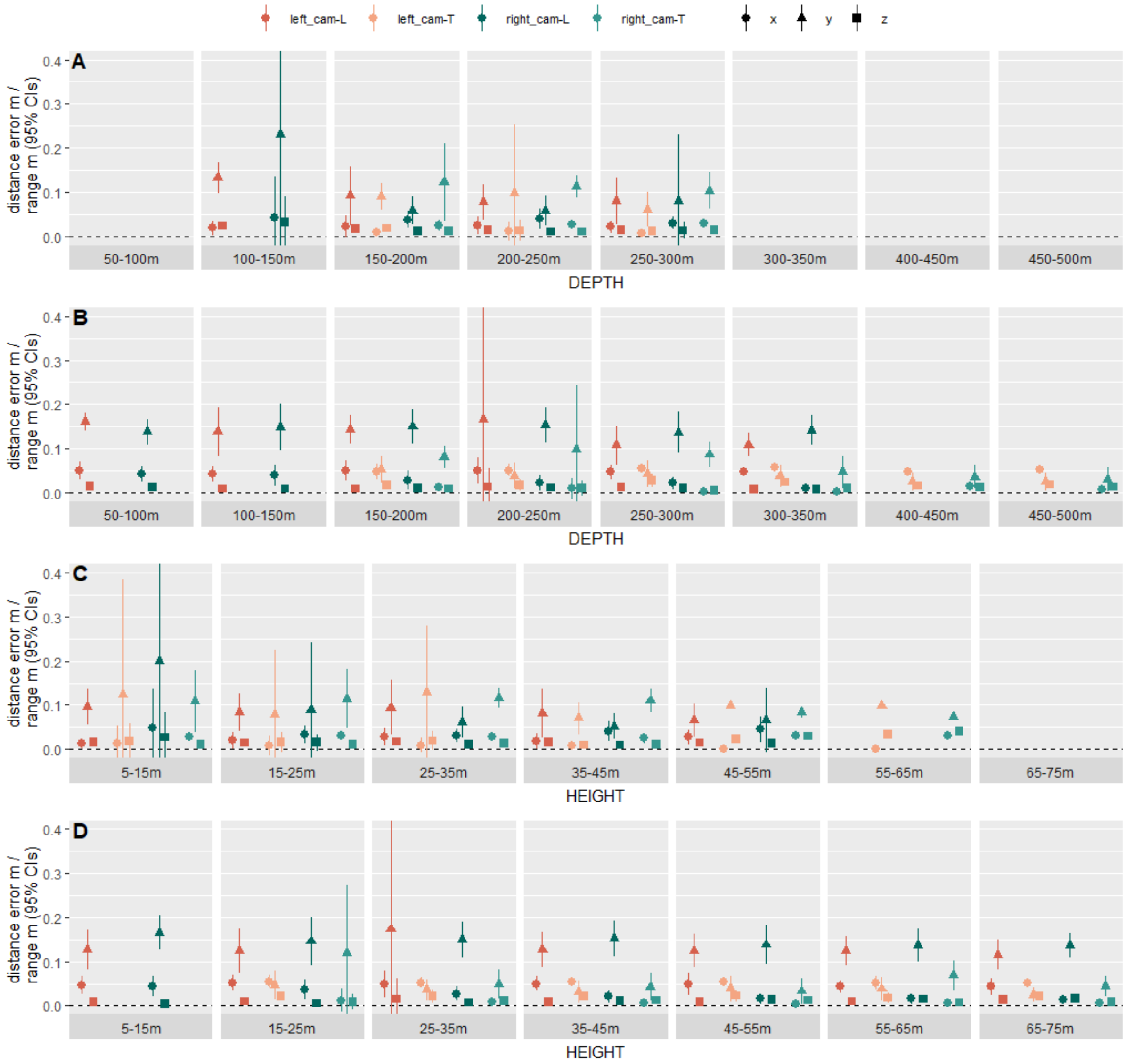

Figure 3-3B. Mono camera reconstructions. Relative distance error (error divided by the straight-line distance from the midpoint between cameras), of the Axis (A,C) and Avigilon (B,D) mono reconstructions relative to the drone GPS; comparing the left and right cameras for lawnmower transects (L) and tetrahedral transects (T). The errors are shown for each axis (x,y and z) with 95% confidence intervals, and are divided into (A,B) longitudinal and (C,D) vertical distance bands. NB some of the CIs extend off the plotting limits.

Table 3-2. Summary of mean distance errors for the stereo 3D reconstructions compared to the drone data.

measure	system	transect	Stereo - Manual			Stereo - Automated		
			mean dist err (m)	SD (m)	SE (m)	mean dist err (m)	SD (m)	SE (m)
XYZ	Axis	Lawnmower	8.24	6.48	0.08	10.84	7.40	0.09
X (left-right)	Axis	Lawnmower	1.97	1.16	0.01	2.20	1.17	0.01
Y (depth)	Axis	Lawnmower	7.15	6.89	0.08	10.21	7.62	0.09
Z (height)	Axis	Lawnmower	2.19	1.19	0.01	1.65	0.89	0.01
XYZ	Axis	Tetrahedral	6.7	2.04	0.03	11.85	5.45	0.08
X (left-right)	Axis	Tetrahedral	1.67	0.49	0.01	1.32	0.62	0.01
Y (depth)	Axis	Tetrahedral	5.2	2.58	0.03	11.3	5.66	0.08
Z (height)	Axis	Tetrahedral	3.36	1.01	0.01	2.6	1.20	0.02

Table 3-3. Summary of mean distance errors for the mono 3D reconstructions compared to the drone data.

measure	system	transect	Mono - Left camera			Mono - Right camera		
			mean dist err (m)	SD (m)	SE (m)	mean dist err (m)	SD (m)	SE (m)
XYZ	Axis	Lawnmower	20.48	11.37	0.13	20.90	31.23	0.35
X (left-right)	Axis	Lawnmower	4.92	4.31	0.05	8.01	5.93	0.07
Y (depth)	Axis	Lawnmower	18.83	11.71	0.13	17.38	31.45	0.35
Z (height)	Axis	Lawnmower	3.32	1.77	0.02	2.83	3.76	0.04
XYZ	Axis	Tetrahedral	20.86	29.89	0.42	27.54	9.28	0.13
X (left-right)	Axis	Tetrahedral	2.07	4.23	0.06	6.39	2.56	0.04
Y (depth)	Axis	Tetrahedral	20.20	29.45	0.41	26.30	9.69	0.14
Z (height)	Axis	Tetrahedral	3.13	4.60	0.07	2.78	1.90	0.03
XYZ	Avigilon	Lawnmower	33.7	39.7	0.67	34.5	12.8	0.21
X (left-right)	Avigilon	Lawnmower	11.2	5.51	0.09	5.13	3.71	0.06
Y (depth)	Avigilon	Lawnmower	30.8	39.6	0.66	33.7	13.0	0.21
Z (height)	Avigilon	Lawnmower	2.53	5.32	0.09	2.30	1.85	0.03
XYZ	Avigilon	Tetrahedral	22.37	7.37	0.13	19.48	20.15	0.33
X (left-right)	Avigilon	Tetrahedral	16.71	6.22	0.11	2.58	3.80	0.06
Y (depth)	Avigilon	Tetrahedral	10.97	7.94	0.14	17.92	20.43	0.33
Z (height)	Avigilon	Tetrahedral	6.34	3.64	0.07	3.59	3.59	0.06

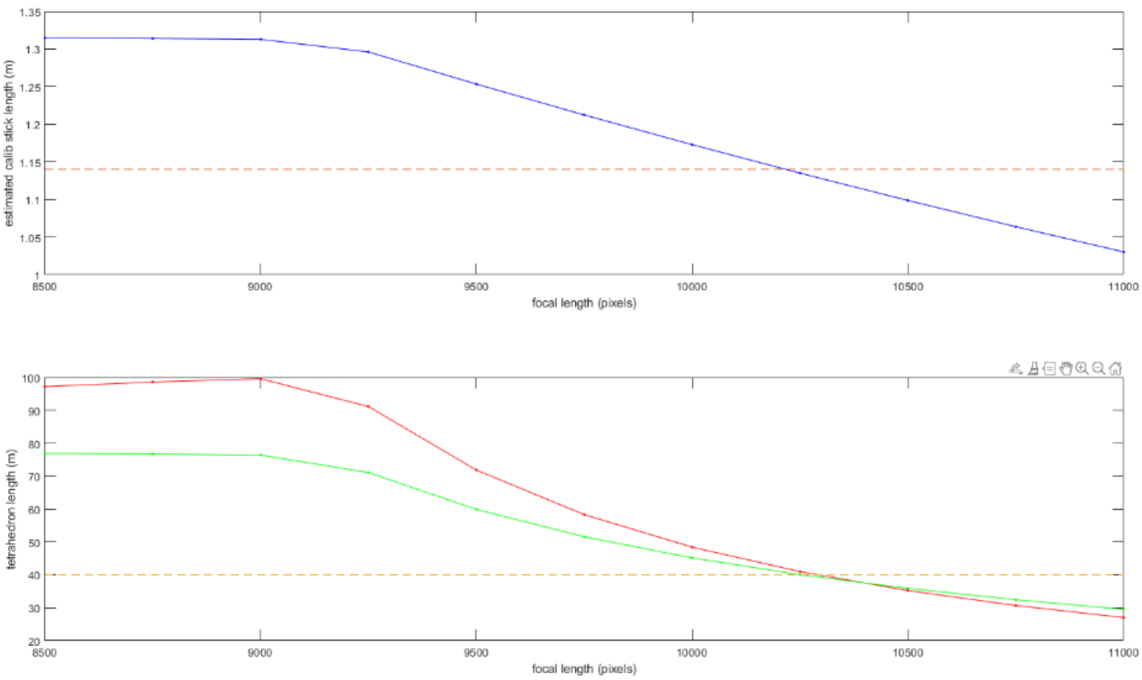

Figure S3-1. Focal length sensitivity check. Top panel plots the change in estimated length of the calibration stick (actual length = 1.14m; dashed line) when running the optimisation with different focal lengths (blue line). Lower panel plots the change in estimated length of two sides of the tetrahedron flown by the drone (actual length = 40m; dashed line) when running the optimisation with different focal lengths (red, green).

Table S3-1. Summary of stereo-reconstruction errors

			RMS error (px)
AXIS Stereo	Calibration		0.21
	Manual	Drone trial 3 lawnmower	0.35
		Drone trial 4 tetrahedral	1.20
	Auto	Drone trial 3 lawnmower	0.98
		Drone trial 4 tetrahedral	2.98

Table S3-2: Comparison of stereo-reconstruction methods. We compared spatial reconstruction errors by method (manual/automatic) and transect type (lawnmower/pyramid) using gaussian linear mixed models with identity links which accounted for the serial correlation in the reconstructed tracks using an AR(1) correlation structure. The models took the general form: distance_error ~ method*transect + distance*method + ar1(seq + 0 | method:transect); where distance represented the distance between the drone and centre of the camera array.

Effects on X error	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	**Significance**
Intercept (manual, lawnmower)	1.955887	0.328320	5.957	2.56e-09	***
automatic	-0.127802	0.464977	-0.275	0.78343	
pyramids	-1.425639	0.483781	-2.947	0.00321	**
distance	-0.003527	0.004403	-0.801	0.42310	
automatic:pyramids	0.205344	0.702977	0.292	0.77021	
automatic:distance	0.026773	0.005619	4.765	1.89e-06	***
Effects on Y error	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	**Significance**
Intercept (manual, lawnmower)	-1.838072	0.150447	-12.22	<2e-16	***
automatic	0.130897	0.213225	0.61	0.5393	
pyramids	-0.088681	0.215577	-0.41	0.6808	
distance	1.088762	0.005119	212.71	<2e-16	***
automatic:pyramids	0.168731	0.314177	0.54	0.5912	
automatic:distance	0.013382	0.006552	2.04	0.0411	*
Effects on Z error	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	**Significance**
Intercept (manual, lawnmower)	3.285089	0.408787	8.036	9.27e-16	***
automatic	-0.938705	0.578847	-1.622	0.1049	
pyramids	1.012827	0.602057	1.682	0.0925	.
distance	-0.125419	0.005490	-22.844	< 2e-16	***
automatic:pyramids	0.044680	0.874434	0.051	0.9592	
automatic:distance	0.048362	0.007004	6.905	5.01e-12	***

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

3.4 References

Ballester C, Dupont SM, Corbeau A, Chambert T, Duriez O, Besnard A. 2024. A standardized protocol for assessing the performance of automatic detection systems used in onshore wind power plants to reduce avian mortality. Journal of Environmental Management. 354: 120437. https://doi.org/10.1016/j.jenvman.2024.120437

Bergeron P. 2007. Parallel lasers for remote measurements of morphological traits. The Journal of Wildlife Management. 71: 289-292. https://doi.org/10.2193/2006-290

Brighton CH, Kloepper LN, Harding CD, Larkman L, McGowan K, Zusi L, Taylor GK. 2022. Raptors avoid the confusion effect by targeting fixed points in dense aerial prey aggregations. Nature Communications. 13. https://doi.org/10.1038/s41467-022-32354-5

Brighton CH, Taylor GK. 2019. Hawks steer attacks using a guidance system tuned for close pursuit of erratically manoeuvring targets. Nature Communications. 10. https://doi.org/10.1038/s41467-019-10454-z

Chilson PB, Bridge E, Frick WF, Chapman JW, Kelly JF. 2012. Radar aeroecology: exploring the movements of aerial fauna through radio-wave remote sensing. The Royal Society. https://doi.org/10.1098/rsbl.2012.0384

Cole E-L, Waggitt JJ, Hendenström A, Piano M, Holton MD, Börger L, Shepard ELC. 2019. The Ornithodolite as a tool to quantify animal space use and habitat selection: a case study with birds diving in tidal waters. Integrative Zoology. 14: 4-16. https://doi.org/10.1111/1749-4877.12327

Hedrick TL. 2008. Software techniques for two- and three-dimensional kinematic measurements of biological and biomimetic systems. Bioinspiration Biomimetics. 3. https://doi.org/10.1088/1748-3182/3/3/034001

Jaquet N. 2006. A simple photogrammetric technique to measure sperm whales at sea. Marine Mammal Science. 22: 862-879. https://doi.org/10.1111/j.1748-7692.2006.00060.x

Lyon BE. 1994. A technique for measuring precocial chicks from photographs. The Condor. 96: 805-809. https://doi.org/10.2307/1369486

Prinsloo ND, Postma M, Ryan PG, Coetzee M, De Bruyn P. 2021. Estimating bird flight height using 3-D photogrammetry. Journal of Zoology. 314: 174-186. https://doi.org/10.1111/jzo.12879

Thaxter CB, Ross-Smith VH, Cook A. 2016. How high do birds fly?: A review of current datasets and an appraisal of current methodologies for collecting flight height data; literature review. British Trust for Ornithology. https://doi.org/10.1098/rspb.2017.0829

Walker SM, Thomas ALR, Taylor GK. 2009. Photogrammetric reconstruction of high-resolution surface topographies and deformable wing kinematics of tethered locusts and free-flying hoverflies. Journal of the Royal Society Interface. 6: 351-366. https://doi.org/10.1098/rsif.2008.0245

Willisch CS, Marreros N, Neuhaus P. 2013. Long-distance photogrammetric trait estimation in free-ranging animals: a new approach. Mammalian Biology. 78: 351-355. https://doi.org/10.1016/j.mambio.2013.02.004

4 Offshore validation trials of Spoor AI systems at Aberdeen Bay

4.1 Introduction

The flexibility and cost-effectiveness of mono- and stereo- vision systems mean they have the potential to be an excellent tool for monitoring birds in offshore environments. We sought to investigate their practicality by testing a camera system deployed at an offshore wind farm in Aberdeen Bay. We compared the distance measurements of birds observed using a laser range finder (LRF) with those obtained from image analysis techniques (both in mono and stereo).

4.2 Methods

Two stereo camera pairs were deployed on a single turbine in an offshore wind farm (AWF10, Fig. 4-1), to reconstruct the three-dimensional flight paths of birds passing close to another turbine approximately 900m away. One pair was aimed at turbine AWF05 (landward-facing) and one pair was aimed at turbine AWF11 (seaward-facing). The high-definition video cameras were mounted on the turbine platform inside weatherproof housing (approximately 20m above mean tide sea level). Cameras were installed by the wind farm operator. Strict space limitations meant that for each stereo pair the cameras were positioned only 4m apart, resulting in an almost parallel configuration. The camera lenses were set to their widest zoom setting (48mm focal length for the seaward-facing pair, 70mm focal length for the landward-facing pair) and recorded 8 Hz video at 5,472 × 3,648 pixels (6k). The cameras were left to record for the duration of the project (1 year), automatically saving the video data as .mkv files in 5-minute segments.

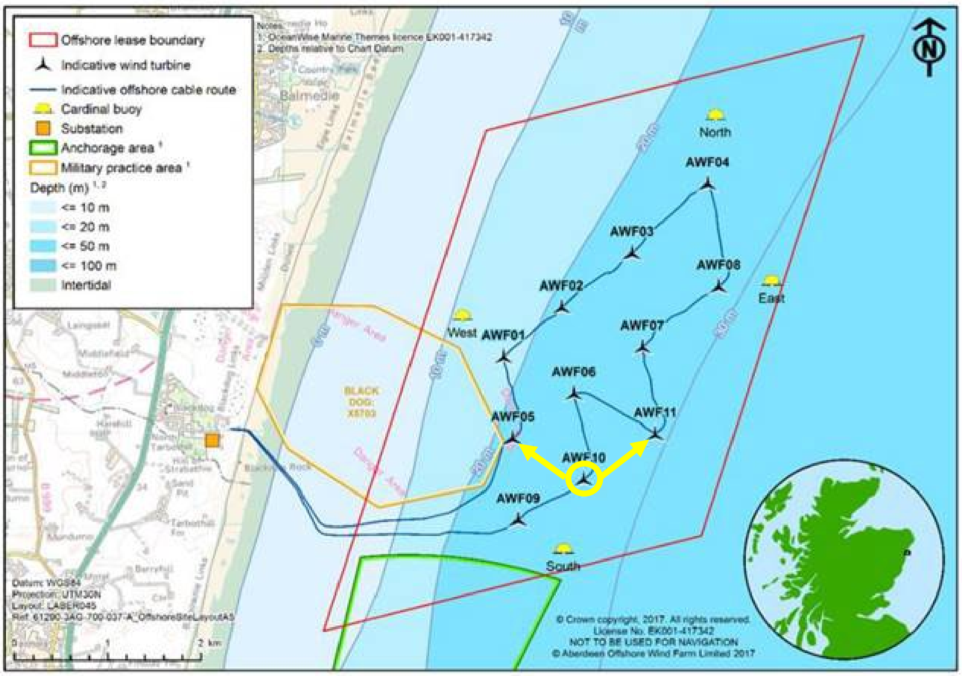

Figure 4-1. Image of Aberdeen Bay offshore wind farm, showing the location of the cameras (yellow circle) and the two monitored turbines.

4.2.1 Photogrammetry for mono camera tracking

Mono camera track reconstructions were provided by Spoor.

4.2.2 Photogrammetry for stereo camera tracking

To calibrate the 3D volume for each stereo pair, we matched 20-30 points across both frames. Points included distant landmarks (for the landward cameras only), stationary points on the turbine and cloud points (for the seaward cameras only), to cover as much of the capture volume as possible (Fig. 4-2). We conducted scale checks for each calibration using known distances on the turbine – tower height, blade length and platform width (Tables, 4-1, 4-2). For the land-facing cameras we also used the straight-line distance to certain points in the landscape, identified and measured on an OS map (Fig. 4-2, Table 4-1). The turbine tower height was also used as the vertical reference for calculating the real-world rotation that was applied to the data.

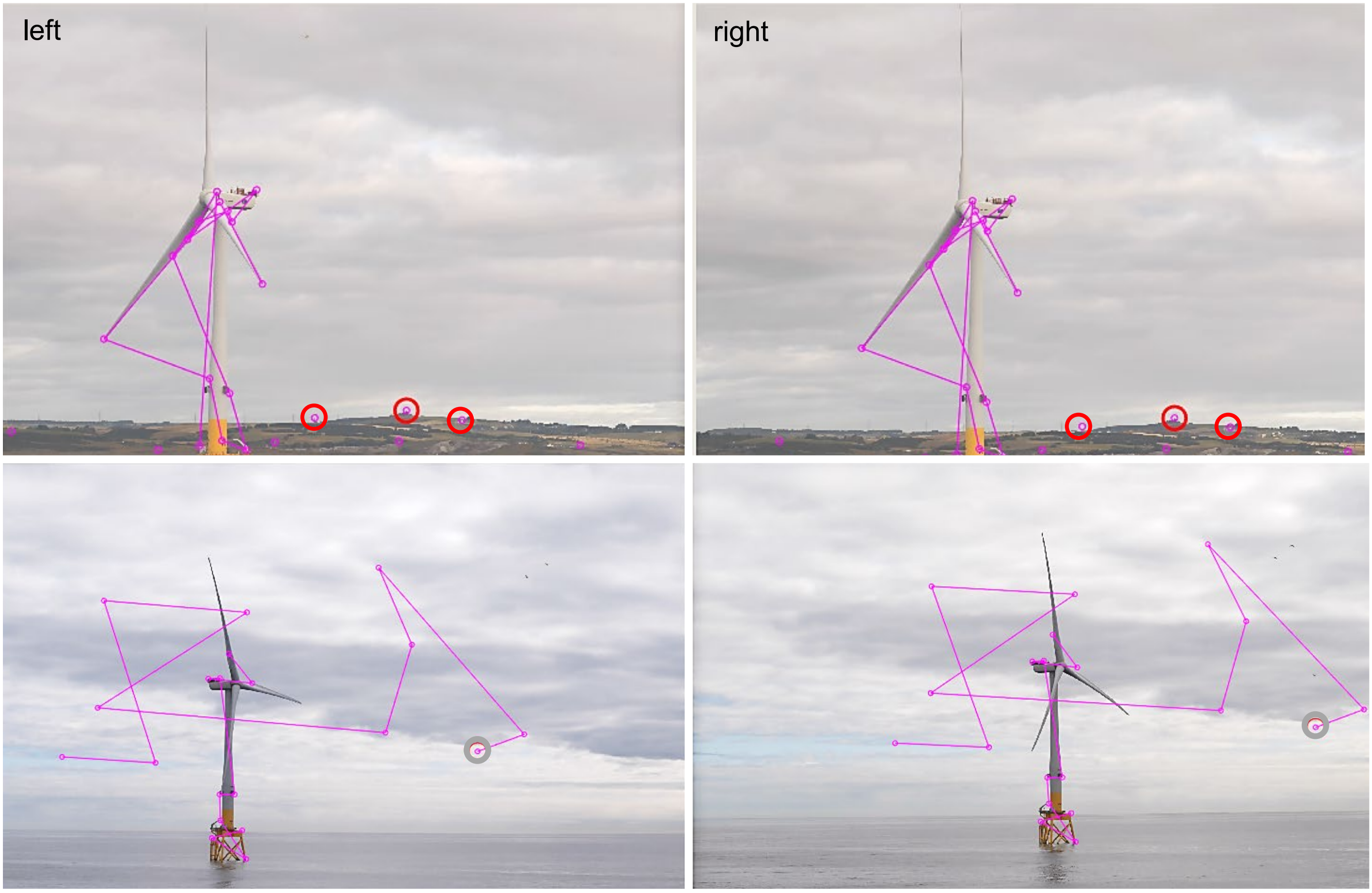

Figure 4-2. Top row: Matched calibration points for the landward cameras (15 stationary turbine points + eight landmarks; red = points identified on OS map). Bottom row: Calibration points for the seaward cameras (ten stationary turbine points + ten cloud points; grey = one of ten cloud points, with a line linking the others).

We synchronized the videos using the timestamp from the video metadata to get an initial estimate of the time offset, before using the wing motions of birds flying through the images to get the exact offset (NB this only guarantees synchronization of the frames to within ±0.0625 s at the 8 Hz frame rate). We obtained the 2D pixel coordinates of any birds in the scene by manually tracking the visual centre of the bird's body from the point at which it appeared in both cameras until it passed out of the field of view (Fig. 4-3).

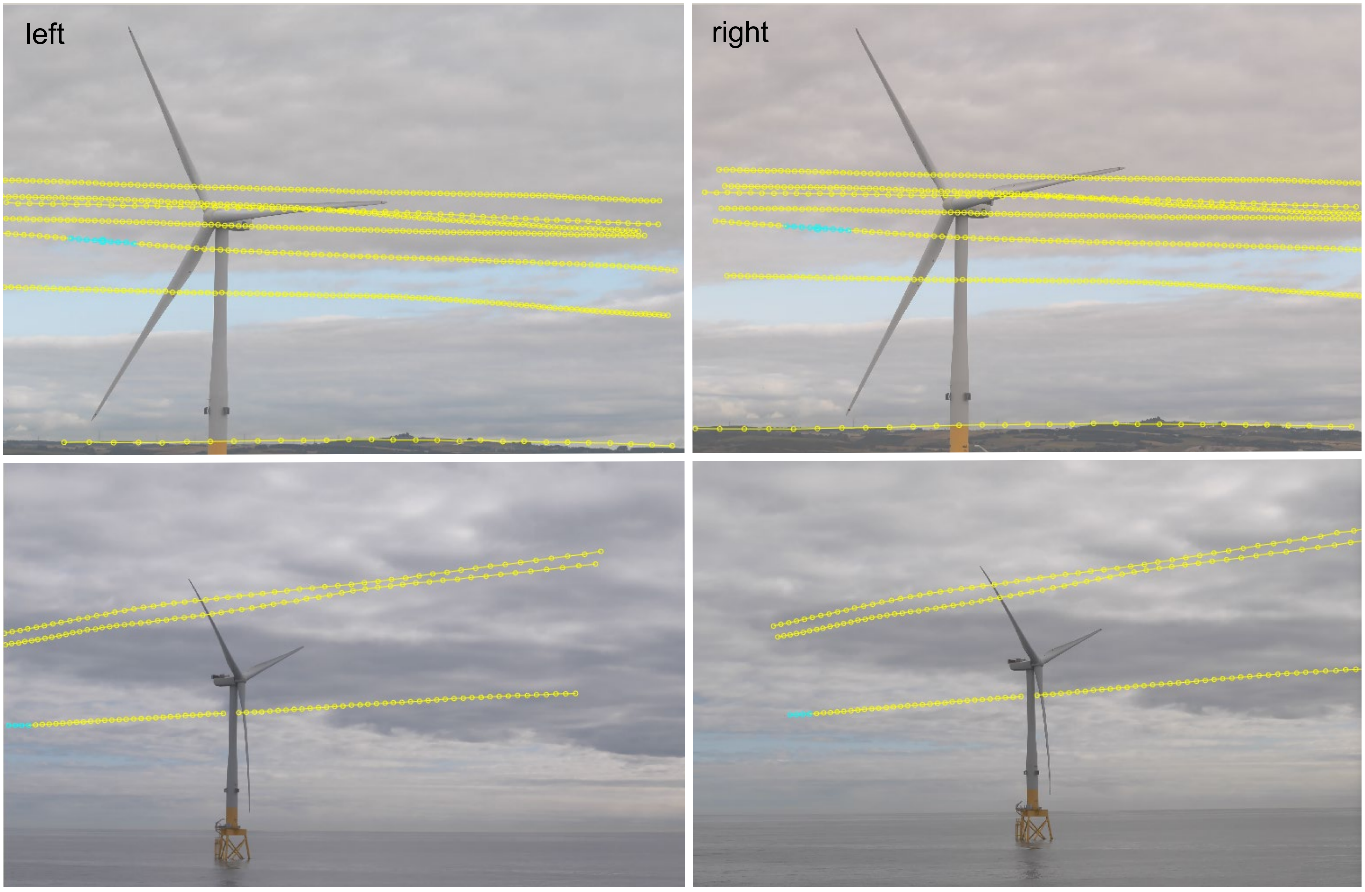

Figure 4-3. Top row: 2D tracked bird points in landward cameras (seven birds flying left to right). Bottom row: 2D tracked bird points in seaward cameras (three birds flying right to left).

To reconstruct the 3D bird trajectories, we used the same optimisation routine that was used for the drone validation trials (fully manual technique), which identified the jointly optimal estimates of the position and pose of the cameras, and spatial coordinates of the calibration points, by minimising the sum of the squared reprojection error of the matched image points (Walker et al., 2009). We used focal lengths of 19,543 pixels for the landward cameras and 12,000 pixels for the seaward cameras. These values were calculated using the manufacturers specifications (focal length in pixels = focal length (mm) / sensor width (mm) * image width (px)). To test the sensitivity of our focal length estimations, we ran reconstructions using focal lengths of 19,000 – 20,000 pixels for the landward pair and 11,600 – 12,400 pixels for the seaward pair (in 100 px increments) and looked at the effects on the estimated lengths of the scale references and landmark ranges. We found that over the tested ranges the scale distances were relatively stable and so we didn't apply any further refinement to the focal lengths (Figs. S4-1, S4-2).

The tilt of the cameras was measured during installation as ~9.5° for the land-facing pair, and ~9° for the sea-facing pair, however – as evidence by the shift in the video images – they have moved considerably from their initial positions. Therefore, we manually estimated the tilt using the value which caused the turbine in the reconstruction to be closest to its actual height above sea level. This resulted in tilts of 5° for the land-facing and 5.5° for the sea-facing cameras.

4.2.2.1 Land-facing calibration metrics:

Table 4-1. Land-facing stereo camera calibration metrics – comparing the estimated distances against known measured distances. RMS is the root mean square error in pixels.

3D points	RMS (px)	Estimated distance (m)	Measured distance (m)	Distance discrepancy (%)
Calibration points	0.47	n/a	n/a	n/a
Vertical reference	0.15	68.6	76	-9.7
All scale references	0.69	60.6	82	-26.1
		63.8	82	-22.2
Turbine distance from camera		793	880	-9.9
Turbine height above mean sea level		29	33	-12.1
Landmark 1 distance from camera		7741	7900	-2.0
Landmark 1 height above mean sea level		137	~150	-8.7
Landmark 2 distance from camera		6696	6700	-0.1
Landmark 2 height above mean sea level		96	~85	+12.9
Landmark 3 distance from camera		5895	6040	-2.4
Landmark 3 height above mean sea level		91	85+mast (>100?)	-9.0

4.2.2.2 Sea-facing calibration metrics:

Table 4-2. Sea-facing stereo camera calibration metrics – comparing the estimated distances against known measured distances. RMS is the root mean square error in pixels.

3D points	RMS (px)	Estimated spacing (m)	Measured distance (m)	Distance discrepancy (%)
Calibration points	2.06	n/a	n/a	n/a
Vertical reference	1.18	63.8	76	-16.1
All scale references	0.33	16.5	14	+17.9
		12.8	21	-39.0
Turbine distance from camera		725	910	-20.3
Turbine height above mean sea level		13	33	-60.6

4.2.3 LRF validation and matching

To assess the accuracy of the 3D reconstructed bird trajectories, a field observer was stationed on turbine AWF10 with the aim of measuring any birds flying across the camera field of views with a laser range finder (LRF; Vector Aero). Over three days (Aug 9th to 11th

2023) the observer recorded 625 fixes (range and bearing) of 190 different birds – species identified were Herring Gull, Kittiwake, Gannet, Fulmar, Common tern, Arctic skua, Cormorant and Manx shearwater. Technical difficulties meant that the observer was unable to record bird height or azimuth and so we used the bearing and time interval of the LRF observations to match the tracks to the video data. However, this proved to be challenging as magnetic interference from the metal turbine meant the LRF bearings were inaccurate and the times of the LRF fixes were only recorded to the nearest minute. Therefore, to limit the matching uncertainty between birds, we filtered the LRF data to only include discernible species which were easy to identify in the videos (e.g. Gannets, juvenile Kittiwakes) and any birds which were temporally separated from any other bird (i.e. only one individual flying through the image with no other birds for at least 1-2 minutes either side). We then identified these individuals in the video streams, using the associated timestamps, and manually tracked them. While we cannot rule out the possibility that the LRF readings may have been taken outside the camera field of view, the fact that most birds were on straight flight paths means we can be reasonably confident that at least some of the fixes should correspond with our photogrammetry birds.

4.3 Results

In total we stereo-tracked 90 birds (67 from turbine AWF11 and 23 from turbine AWF5; Fig. 4-4) from the three observer days and found potential LRF matches for 63 (Table S4-1). Of these 90 birds we were able to track 54 using mono-vision, which consisted of 89 trajectories (left and right cameras). Because of the spatial and temporal uncertainty in aligning the camera data with the LRF data, it would not be appropriate to compare these data quantitatively. Therefore, we visually compared the flight distances of the LRF tracked birds with the corresponding bird distances estimated via photogrammetry by overlaying the reconstructed trajectories onto the arc given by the min-max LRF range from the observer/camera position (Fig. 4-5). Overall, 19 out of 54 stereo tracks (35%) overlapped the LRF min-max, whereas only 18 out of 89 mono tracks (20%) were found to overlap.

We compared the measurement difference between each stereo and mono trajectory by calculating the mean Euclidean distance between each 3D stereo bird point and 3D mono bird point (Fig. 4-6) and found a wide distance discrepancy in many of the flights. We assessed each axis separately (Fig. 4-7) and compared the instantaneous stereo and mono positions relative to the cameras. Overall, the mono tracks were shown to overestimate the heights and ranges compared to the stereo, but the distances in the lateral axes were similar. To assess specific flight metrics – tortuosity, path length, amount of turning and speed – we had to interpolate any gaps in the stereo data and smooth the trajectories using quintic splines (the mono tracks were already smoothed by the Spoor AI software during the reconstruction process). We then estimated the straightness of the flight paths (tortuosity) in the horizontal plane by dividing the total path length by the straight-line distance between the first and last points in the trajectory (i.e. the arc-chord ratio). This revealed that in general the mono reconstructions were longer and less straight than the corresponding stereo tracks (Fig. 4-8 A,B), and in particular apparent changes of direction in mono reconstructions were not apparent in most stereo-reconstructions . We also calculated the total amount of turning (track azimuth) between successive trajectory points in the horizontal plane, which did not identify major differences between mono and stereo track reconstructions. (Fig. 4-8C). The mean flight speed for each trajectory showed a much wider distribution for the mono tracks with some very extreme and unrealistic values, compared to the stereo tracks which had a much tighter distribution around 40-50 km/h (10-14 m/s; Fig. 4-8D), which is consistent with existing flight speed estimates for seabirds (e.g. Pennycuick, 1987; Cook et al., 2023).

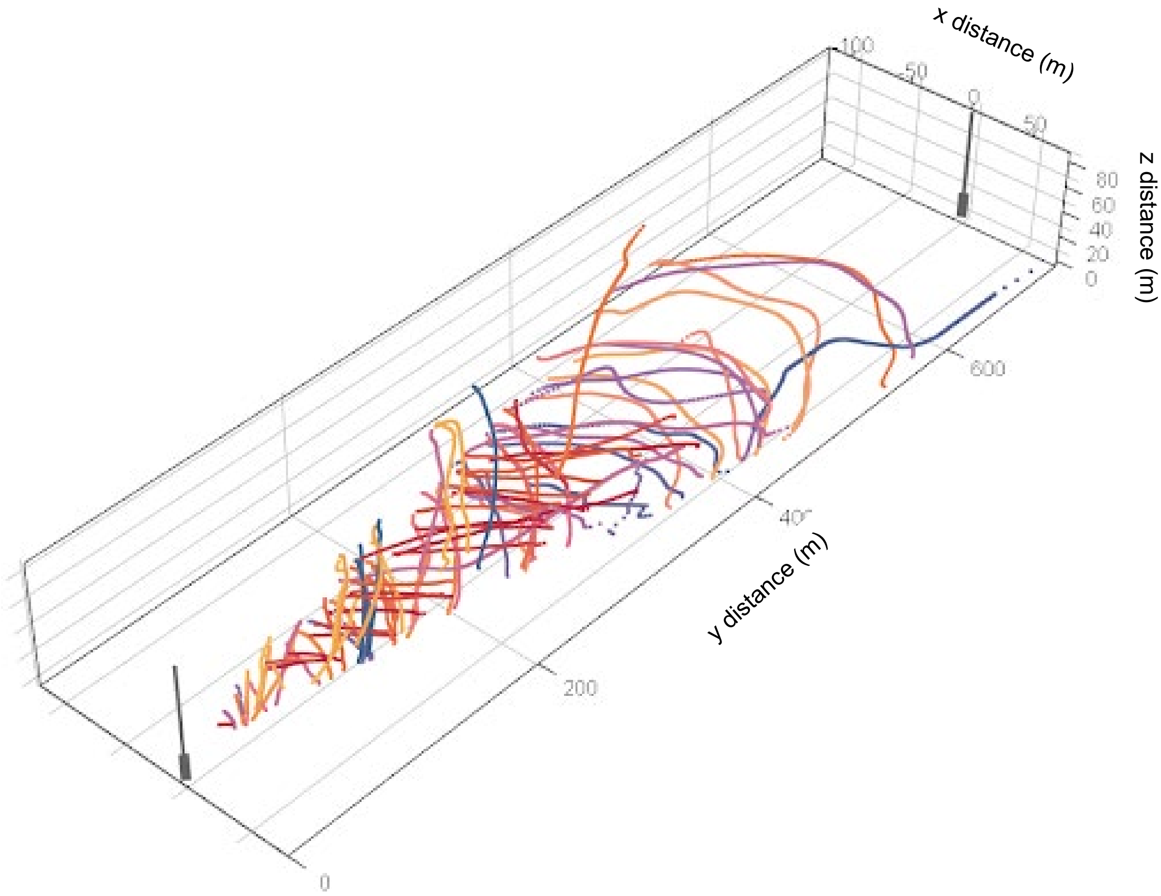

Figure 4-4. The 3D reconstructed tracks of 90 birds flying through the wind farm during 3 days in August. The grey lines represent the approximate positions of the camera turbine and the focal turbine. NB the cone shape produced by the combined trajectories reflects the area of overlap between the cameras.

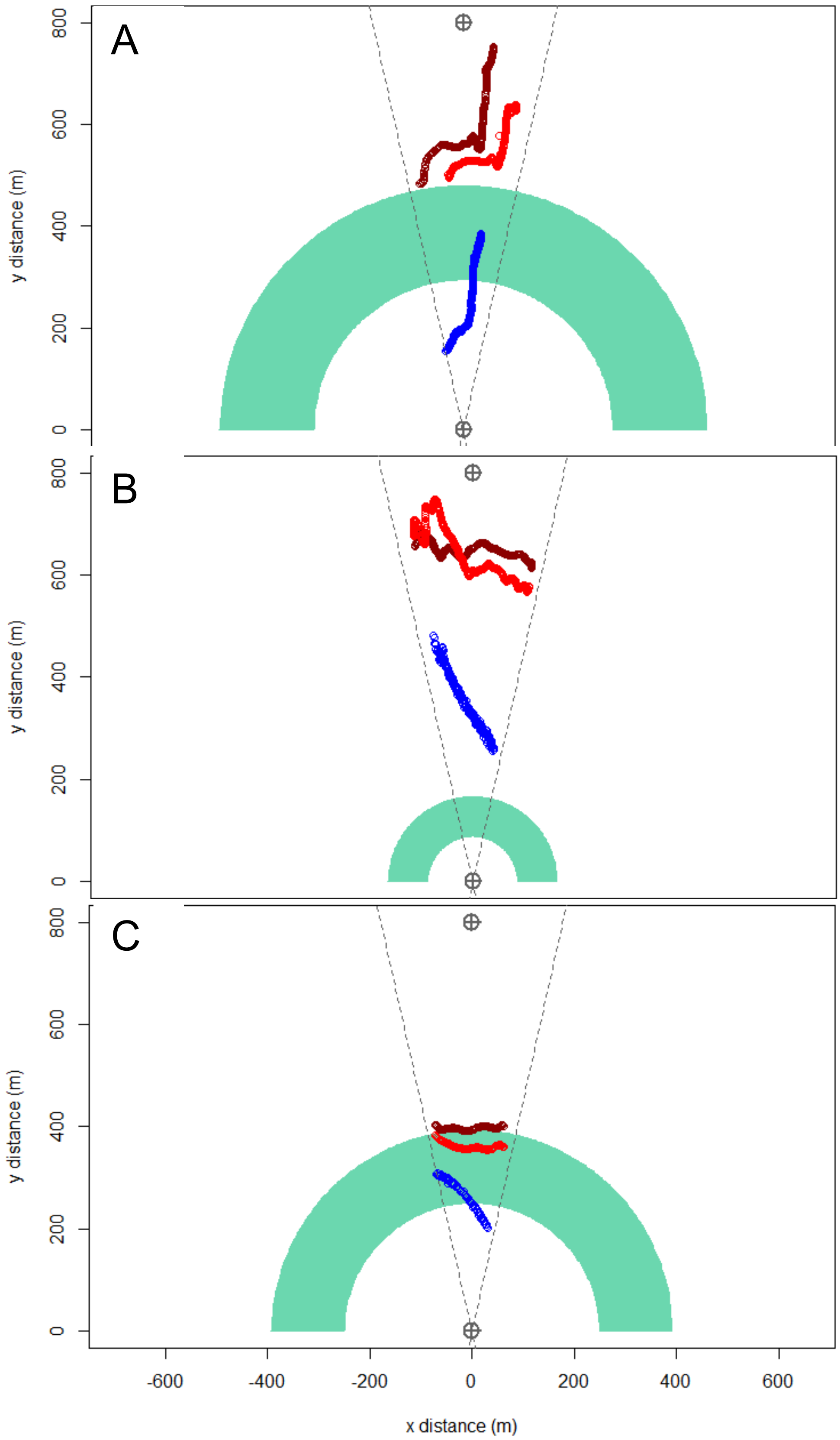

Figure 4-5. Top view of the camera field of view (dashed lines) in relation to the turbines (circles) with three stereo reconstructed bird trajectories (blue), 6 mono trajectories (red and dark red) and the min-max range of the LRF readings (green) which were identified as matches. Panels show examples of A) a stereo-LRF overlap but mono distance is over estimated, although trajectory shapes agree between stereo and mono; B) a poor LRF match for stereo and mono, and diverging track shapes with much more tortuous mono tracks; and C) a good

match where all datasets overlap with the LRF distances, although directionality of mono and stereo tracks differs.

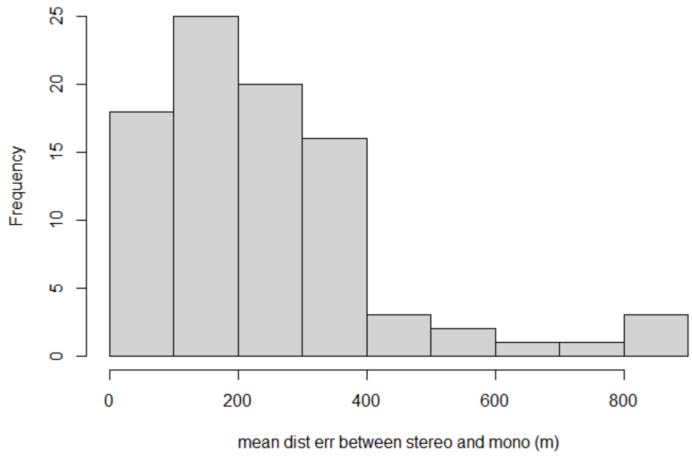

Figure 4-6. Histogram of measurement difference between the stereo and mono trajectories. Mean distance error is calculated using the Euclidean distance between each 3D stereo bird point and mono bird point.

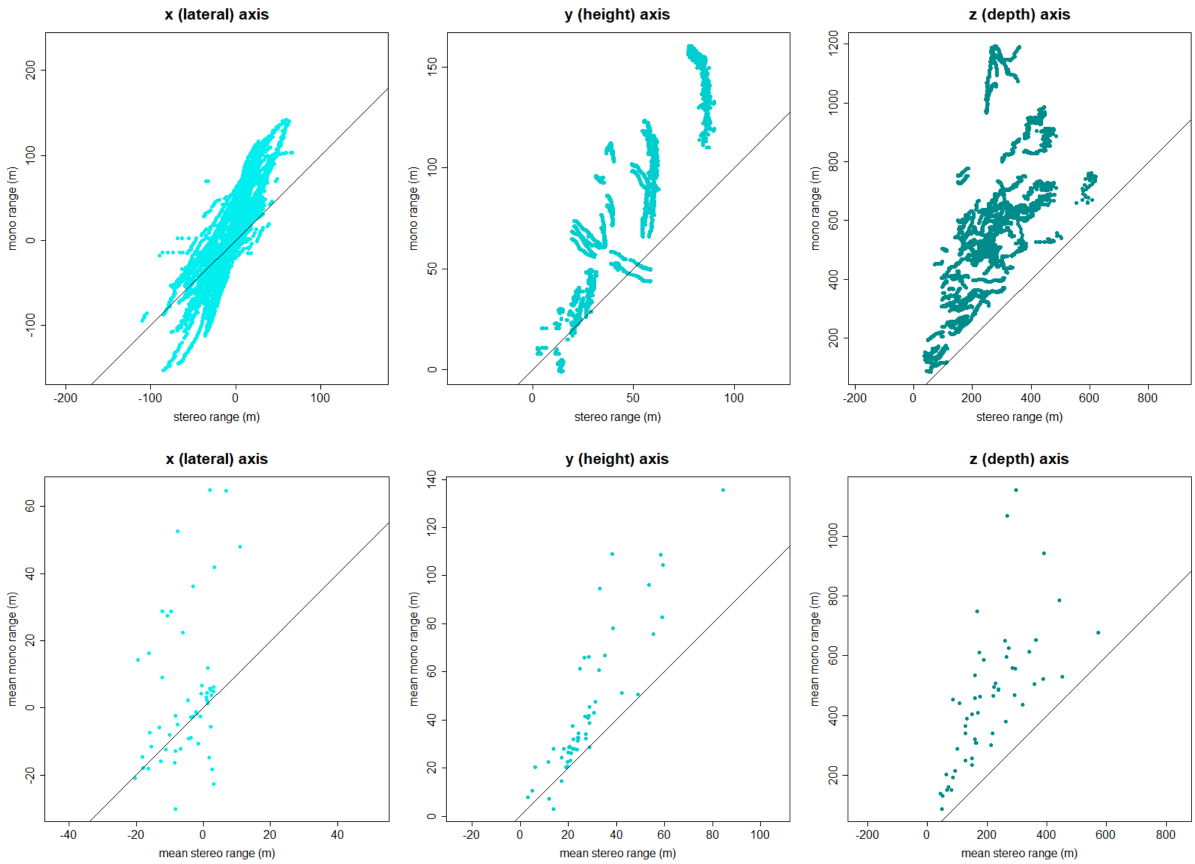

Figure 4-7. Scatter plots of point-by-point (top row) and means (bottom row) of stereo x, y and z distances versus mono x, y and z distances. The diagonal line represents the 1:1 line.

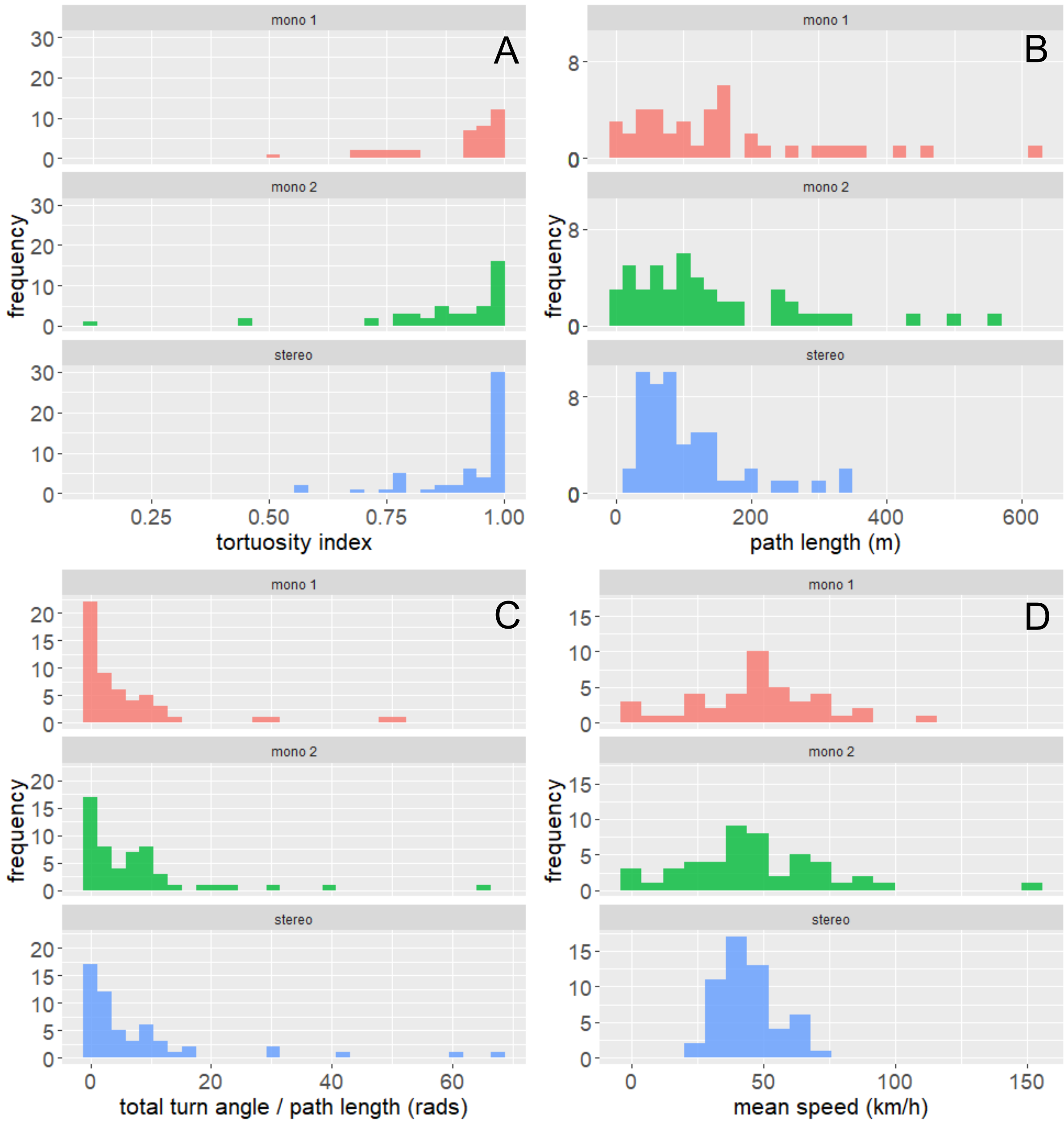

Figure 4-8. Comparison of four different flight metrics – tortuosity (Values closer to 1 indicate a straighter path) , path length, cumulative turn angle relative to distance travelled and mean speed – between the stereo and mono tracks. NB high values in panel C are a tracking artefact, caused by trajectories with a lot of jitter.

4.3.1 Conclusions and recommendations

The stereo and mono camera systems work in principle for recording the 3D trajectories of seabirds flying in offshore wind farms, for both land-facing and sea-facing camera systems. The stereo camera systems were more accurate than the mono systems – with the stereo-calibration distances being underestimated by around 20% while the mono distances were overestimated by about 50%. The stereo-reconstructions also had much better precision than the mono reconstructions, producing smoothed curving trajectories as observed in the video footage. Whereas the mono trajectories exhibited many wiggles and jerks due to the changing size of the detection bounding box as the birds flapped – and therefore are artefacts of the tracking algorithm. These artefacts made it difficult to ascertain whether movement behaviour inferred from mono trajectories was real, in particular as only one stereo trajectory showed a substantial change in direction (Fig. 4-5A). Information on instantaneous speed estimates and the track geometry may aid the separation of true movement from reconstruction artefacts. In particular it appears likely that artefacts based on bounding box fluctuations alone will result in apparent displacements along ray paths, and tracks that are close to a plane with a near constant elevation angle relative to the cameras, whereas true movement can occur in any plane. However, a much larger sample of stereo trajectories with directional changes will be necessary to develop quantitative assessments of movement artefacts in mono trajectories.

Using a laser range finder to validate camera-derived data is far from straightforward, as when there were many transiting birds it proved extremely difficult to match the two data sources. This was in addition to the fact that the LRF observer could only guess where the edges of the camera field of view were in space, so we cannot be certain that LRF fixes spatially overlapped with the cameras. Future LRF and camera studies require a more systematic approach so that birds may be matched more easily – e.g. having an LRF connected to a computer so that it logs a more precise timestamp or having additional personnel to monitor a live camera image. Additional inquiries with other projects in the offshore monitoring space (including the OWEC ProcBe and ReSCUE projects) revealed that imprecise timestamping and low sampling rates are a common feature of LRF user interfaces, suggesting that this method is fundamentally limited in providing validation data at the scale appropriate for high-throughput monitoring systems such as the Spoor system evaluated here.

As with many other monitoring techniques, the visibility and background contrast are important for successful detection. The latter was a major limitation of the mono approach, as the reconstruction algorithms require further development to enable tracking of birds in front of the sea. Therefore only birds above the horizon were successfully tracked. The biggest limitation of the stereo approach was that the calibration needed suitable landmark points across the capture volume – rather lacking in the offshore environment. The most successful calibrations used points on the turbine and distant landmarks or clouds (the latter being a problem on clear days!). The land-facing stereo cameras were calibrated more easily and more successfully due to being able to use more appropriate landmarks across the capture volume. Changing the value of the focal length in the stereo-reconstructions did not significantly affect the reference measurements (they were quite stable within the ranges tested), which suggests that the source of error came from the geometry of the camera setup (lack of parallax) and/or the absence of calibration points covering the whole capture volume. Calibrations could be improved by using birds as reference objects, but this relies on having an accurate time-synch between cameras and so was not an option here.

The cameras were at a distance of ~900m from the focal turbines and at this distance their image resolution was not sufficient for the human eye to be able to observe any birds or avoidance behaviours close to the turbines. In contrast, the mono automatic tracking system was able to track birds flying behind the turbines (albeit against the sky only), however, its detection ability outperformed the ability of a human observer to identify the species. Thus, while the mono system relies on this manual element species identification remains a significant bottleneck in the approach. Overall, the scalability and practicality of mono systems far outweigh stereo systems, making mono systems the obvious choice for long-term monitoring of meso- and macro- avoidance behaviours. However, the increased accuracy of stereo systems makes them the ideal tool for short-term studies on micro-avoidance behaviours (ideally <500m away, or with larger baseline distances). In addition, with some straightforward modifications the mono and/or stereo systems could have improved accuracy and reduced reliance on manual techniques:

- Using the latest camera technologies – crucially, with higher resolutions, higher frame rates and electronic time-synching (being only 1 or 2 frames out with the time shift resulted in very poor reconstructions).
- Using the horizon for the vertical reference (i.e. perpendicular to it), or camera accelerometer, to improve the real-world transformation (the taper of the turbine tower required some approximation).
- Improve the camera geometry by having a larger baseline distance between the cameras to enable more motion parallax between the camera images.
- Using the horizon and/or the turbine nacelle, or camera accelerometer, to identify if/when a camera has shifted and therefore the capture volume needs re-calibrating.
- More rigid housing and attachments for the cameras so that they do not move – then only a single comprehensive stereo-calibration would need to be performed on installation (which would be expensive/difficult to repeat, such as moving a drone and/or boat through the capture volume as in our onshore experiments).

In summary, we have shown that camera-based systems could be a valuable tool for autonomously monitoring the 3D airspace around turbines. Their ability to operate continuously, including in inclement weather when human observers would struggle, demonstrates some of their principal benefits. However, the downstream post-processing routine needs further development to maximise their potential in these conditions.

4.4 References

Cook ASC, Thaxter CB, Davies J, Green RMW, Wischnewski S, Boersch-Supan P. 2023. Understanding seabird behaviour at sea. Part 2: improved estimates of collision risk model parameters. Scottish Government Report.

Pennycuick CJ. 1987. Flight of Auks (Alcidae) and Other Northern Seabirds Compared with Southern Procellarhformes: Ornithodolite Observations. Journal of Experimental Biology. 128: 335-347. https://doi.org/10.1242/jeb.128.1.335

Walker SM, Thomas ALR, Taylor GK. 2009. Photogrammetric reconstruction of high-resolution surface topographies and deformable wing kinematics of tethered locusts and free-flying hoverflies. Journal of the Royal Society Interface. 6: 351-366. https://doi.org/10.1098/rsif.2008.0245

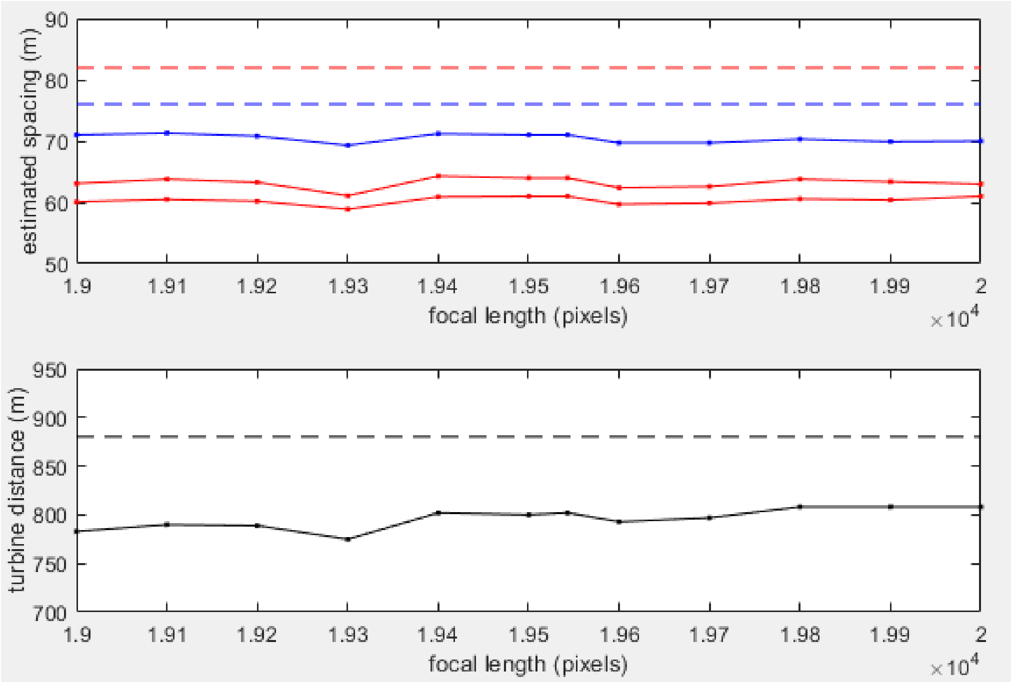

Figure S4-1. Testing different focal lengths on the performance of the 3D calibration (solid lines) for the landward-facing cameras, in terms of how well they estimate the true distances of our scale reference measurements (dashed lines). Red = turbine blades (82m); blue = tower height (76m); black = distance to turbine (880m).

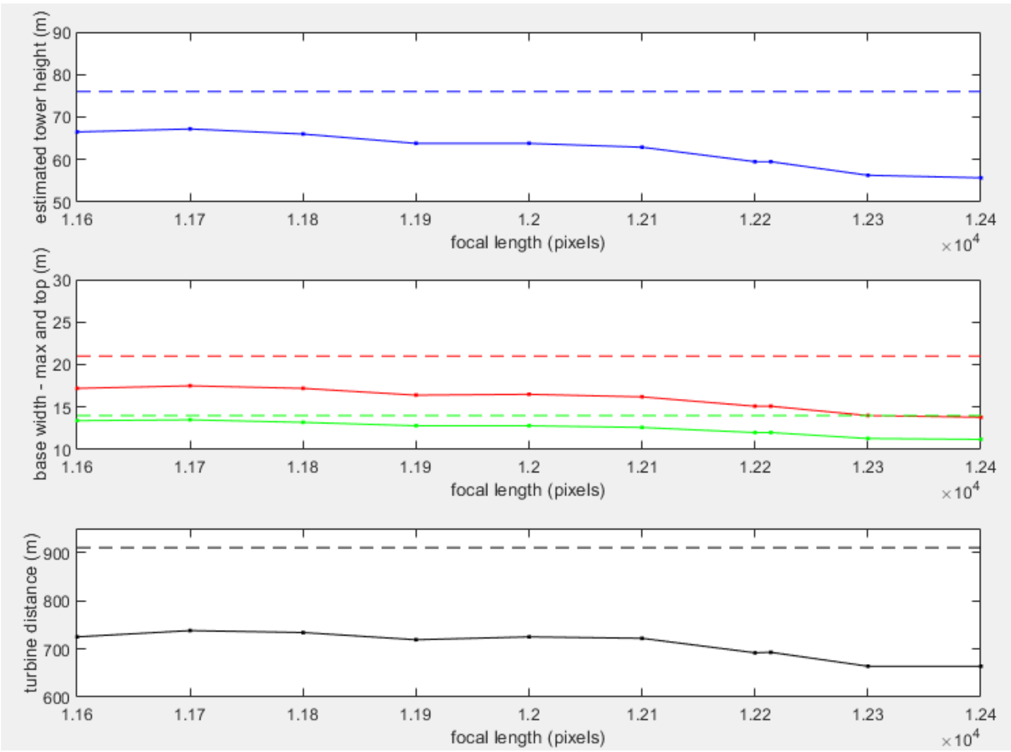

Figure S4-2. Testing different focal lengths on the performance of the 3D calibration (solid lines) for the seaward-facing cameras, in terms of how well they estimate the true distances of our scale reference measurements (dashed lines). Blue = tower height (76m); red = maximum platform width (21m); green = minimum platform width (14m); black = distance to turbine (910m).

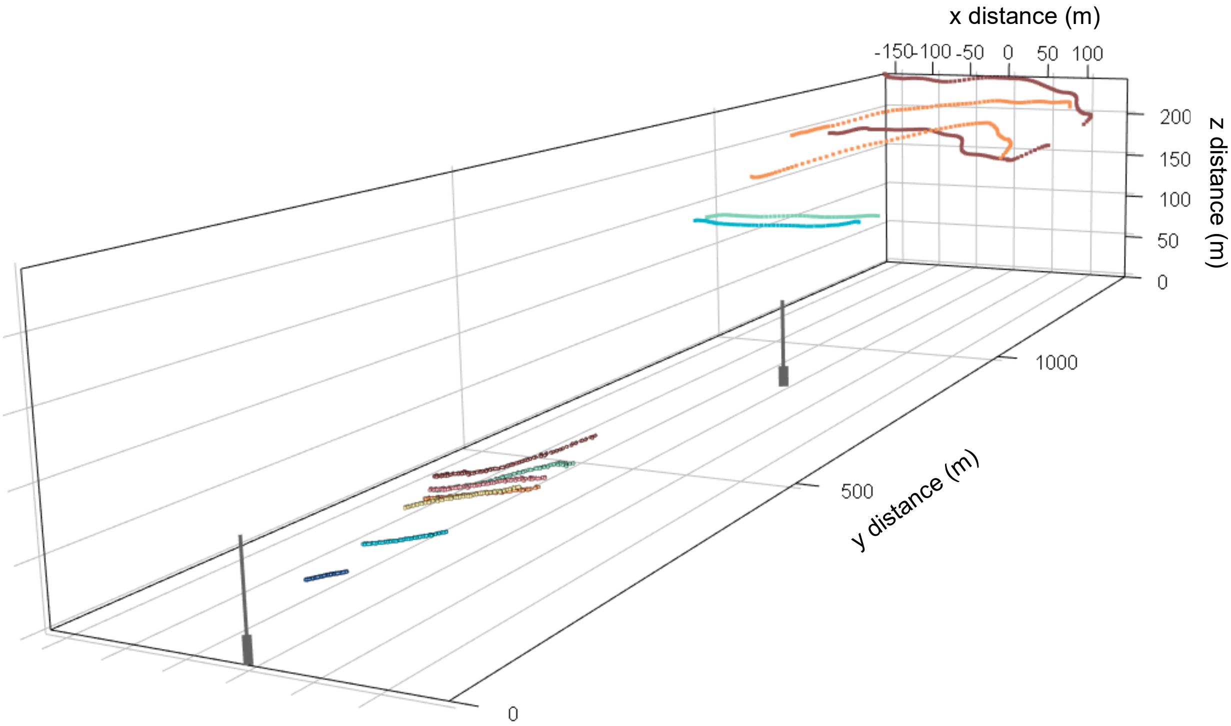

Figure S4-3. 3D trajectories of the seven birds from Figure 4 with 6 matched Spoor mono birds overlain. Tracks with colour duplication are the same bird tracked in the left-hand and right-hand cameras. The grey lines represent the reconstructed turbine positions.

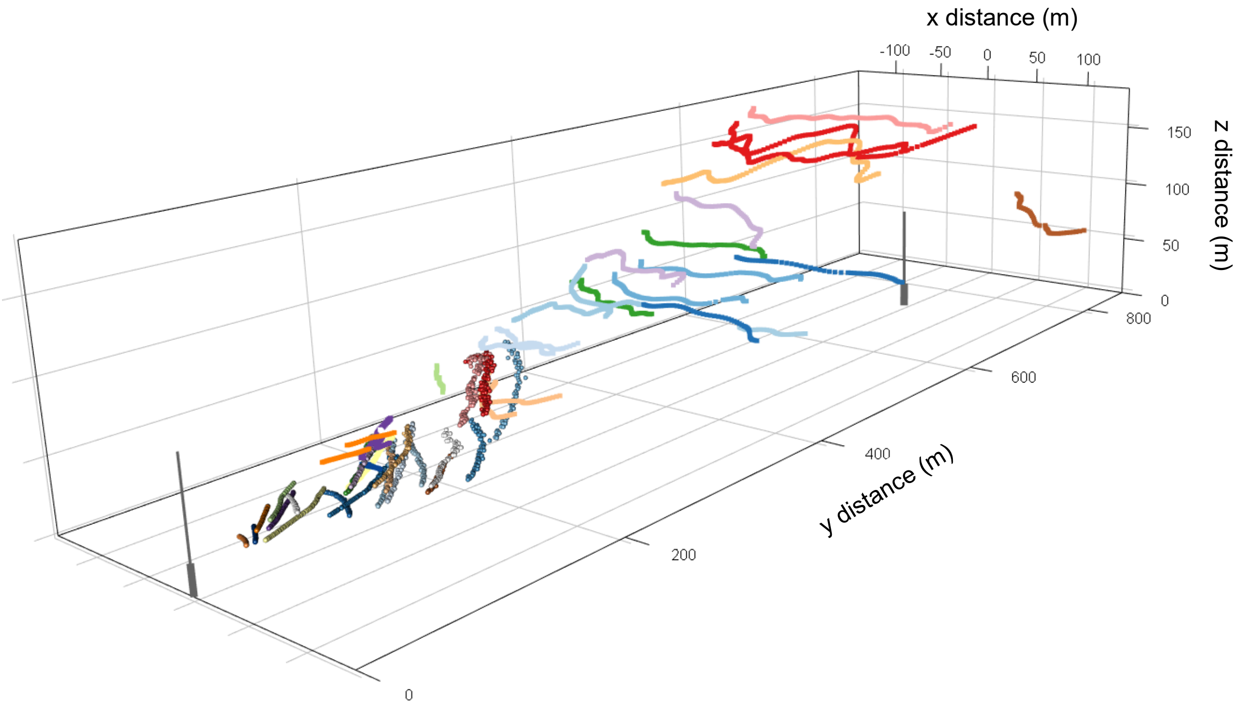

Figure S4-4. 3D trajectories of 22 birds from Figure 4 with 30 matched Spoor mono birds overlain. Tracks with colour duplication are the same bird tracked in the left-hand and right-hand cameras. The grey lines represent the reconstructed turbine positions.

bird ID	turbine	date	UTC real time	LRF time	UTC time diff	species cameras	species LRF	LRF 01 (m)	LRF 02 (m)	LRF 03 (m)	LRF 04 (m)	LRF 05 (m)	LRF 06 (m)	LRF 07 (m)	LRF 08 (m)	LRF 09 (m)	LRF 10 (m)	LRF 11 (m)
01	5	20230810	09:01:48	n/a	n/a	HG	no match											
02	5	20230810	09:01:49	n/a	n/a	Gull	no match											
03	5	20230810	09:02:10	n/a	n/a	Gull	no match											
04	5	20230810	09:02:11	n/a	n/a	Gull	no match											
05	5	20230810	09:02:56	n/a	n/a	HG	no match											
06	5	20230810	09:03:44	10:04:00	00:00:16	HG	HG	187	205									
07	5	20230810	09:03:50	10:04:00	00:00:10	Gull	HG	187	205									
DA	5	20230810	11:11:28	12:11:00	00:00:28	KI juv	KI juv	115	171	188	236	291						
DB	5	20230810	11:20:36	12:21:00	00:00:24	KI	KI	207	236	315	236	389	422					
DC	5	20230810	11:31:42	12:31:00	00:00:42	KI	KI	248	264	289								
DD	5	20230810	11:34:16	12:34:00	00:00:16	KI	KI	214	214	127	221	203	208	246	294	319	294	319
DE	5	20230810	11:35:01	12:34:00	00:01:01	KI?	KI	214	214	127	221	203	208	246	294	319	294	319
DF	5	20230810	11:35:15	12:35:00	00:00:15	HG	HG	206	327									
DG	5	20230810	11:36:36	12:37:00	00:00:24	HG	HG	64	47									
DH	5	20230810	12:21:16	13:21:00	00:00:16	KI	KI	170	262									
DI	5	20230810	12:24:15	13:23:00	00:01:15	GX	GX	428	546	646								
DJ	5	20230810	12:25:48	13:26:00	00:00:12	GX	GX	459	594									
DK	5	20230810	12:36:47	13:36:00	00:00:47	gull	KI	166	176	293	315	410	540					
DL	5	20230810	12:45:45	13:45:00	00:00:45	GX	GX	333										
DM	5	20230810	13:13:47	14:14:00	00:00:13	KI	KI	136	182	123	186							
DN	5	20230810	13:29:11	14:27:00	00:02:11	GX	GX	269										
DO	5	20230810	13:57:26	14:57:00	00:00:26	gull	KI	186										
DP	5	20230810	13:58:26	14:58:00	00:00:26	GX	GX	264										
001	11	20230810	09:12:47	10:12:00	00:00:47	gull	KI	108	108	128								
002	11	20230810	09:12:47	10:12:00	00:00:47	gull	KI	108	108	128								
003	11	20230810	09:12:47	10:12:00	00:00:47	gull	KI	251	330	391								
A	11	20230810	09:26:34	10:29:00	00:02:26	KI	KI	132	158	205								
B	11	20230810	09:26:38	n/a	n/a	HG juv	no match											
C	11	20230810	09:28:07	10:29:00	00:00:53	gull	HG	77	113									
C1	11	20230810	09:26:49	10:29:00	00:02:11	GX	GX	525	570	602	645	937	1003	1062	1214			
C2	11	20230810	09:26:52	n/a	n/a	UU	no match											
D	11	20230810	09:28:19	n/a	n/a	gull	no match											
E	11	20230810	09:31:16	n/a	n/a	KI	no match											
F	11	20230810	09:32:49	n/a	n/a	gull	no match											
G	11	20230810	09:33:27	n/a	n/a	gull	no match											
H	11	20230810	09:33:29	n/a	n/a	KI	no match											
I	11	20230810	09:33:31	n/a	n/a	gull	no match											
J	11	20230810	09:33:31	n/a	n/a	gull	no match											
K	11	20230810	09:33:33	n/a	n/a	gull	no match											
L	11	20230810	09:33:41	n/a	n/a	KI	no match											
M	11	20230810	09:33:44	n/a	n/a	KI	no match											
N	11	20230810	09:34:15	10:35:00	00:00:45	KI	KI12	255	251	259	204							
O	11	20230810	09:35:41	10:35:00	00:00:41	HG	KI13	302	366	419								
P	11	20230810	09:35:45	10:35:00	00:00:45	UU	KI14	299	217	320	347	371	403	453				
Q	11	20230810	09:36:11	10:36:00	00:00:11	Gull	HG	88	165	200	251	406	356	514				
R	11	20230810	09:37:15	10:37:00	00:00:15	KI	KI	245	239	207	206							
S	11	20230810	09:37:28	10:38:00	00:00:32	KI	KI	360	388	375	404	377	547					
T	11	20230810	09:09:28	10:06:00	00:03:28	UU	F.	397	287	259	237	242						
U	11	20230810	09:17:03	10:19:00	00:01:57	GX?	GX	615	1022									
V	11	20230810	09:46:00	10:46:00	00:00:00	GX	GX	102	312	308	312	201	355	408				
W	11	20230810	09:57:02	10:58:00	00:00:58	GX	GX	294	240	244	261	338	406	494	625			
SP1	11	20230810	10:48:13	11:48:00	00:00:13	Skua	AC	128	124	140	179							
SP2	11	20230810	10:25:13	11:24:00	00:01:13	KI	KI	127	116	130	174							
SP3	11	20230810	10:23:06	11:24:00	00:00:54	HG	HG	134	92	89	120	164						
SP4	11	20230810	11:04:53	12:05:00	00:00:07	GX	GX	655	643	619	623	631	646					
SP5	11	20230810	11:02:34	12:02:00	00:00:34	KI	KI	136	146	158	225	318	389					
AB	11	20230809	15:02:56	16:02:00	00:00:56	GX	GX	280	265	308	366	418	477	476	723			
AD	11	20230811	10:42:26	11:42:00	00:00:26	GX	GX	541	434									
AE	11	20230811	10:47:41	11:47:00	00:00:41	GX	GX	398	543									
AF	11	20230811	12:09:13	13:09:00	00:00:13	GX	GX	533	627	711								
AG	11	20230811	08:28:36	09:30:00	00:01:24	gull	KI	162	217									
AH	11	20230811	08:52:53	n/a	n/a	gull	no match											
AI	11	20230811	08:53:07	09:55:00	00:01:53	KI juv?	KI	170	121	99								
AJ	11	20230811	09:34:54	10:35:00	00:00:06	GX	GX	245	256	333	386	467	621	776				
AK	11	20230811	09:36:29	10:36:00	00:00:29	GX	GX	268	308	340	525							
AL	11	20230811	10:09:45	n/a	n/a	GX	no match											
AM	11	20230811	10:09:59	11:11:00	00:01:01	GX	GX	163	131	117	140	156	167					
AN	11	20230811	10:13:11	11:12:00	00:01:11	GX?	GX	720										
AO	11	20230811	11:21:15	12:21:00	00:00:15	GX	GX	581	478	399	459							
AP	11	20230811	11:21:25	12:21:00	00:00:25	GX	GX	581	478	399	459							
AQ	11	20230811	11:25:12	12:25:00	00:00:12	GX	GX	632	660	745								
AR	11	20230811	11:26:43	12:26:00	00:00:43	GX	GX	714	675	671								
AS	11	20230811	11:27:25	12:27:00	00:00:25	GX	GX	615	550	497								
BB	11	20230809	14:05:24	15:05:00	00:00:24	HG	HG	388	422	486	589	641	714					
BC	11	20230809	14:21:35	15:22:00	00:00:25	HG	HG	295	477									
BD	11	20230810	09:30:40	10:29:00	00:01:40	HG	HG	77	113									
BE	11	20230810	10:07:12	11:08:00	00:00:48	HG	HG	100	187	236	280	374	437	485	587	582	670	
BF	11	20230811	09:01:49	10:03:00	00:01:11	KI	KI	132	177	267	545							
BG	11	20230810	09:17:03	10:17:00	00:00:03	KI	KI	227	246	269	287	320	264					
BH	11	20230810	09:45:33	10:44:00	00:01:33	KI	KI	386	337	324	336	337	344	371	362	471		
BJ	11	20230810	09:45:50	10:44:00	00:01:50	KI	KI	386	337	324	336	337	344	371	362	471		
BK	11	20230810	09:55:54	10:55:00	00:00:54	KI	KI	354	340	359	693							
CA	11	20230811	12:45:31	n/a	n/a	gull	no match											
CB1	11	20230811	12:49:26	13:50:00	00:00:34	gull	KI	522	479	519								
CB2	11	20230811	12:49:26	13:50:00	00:00:34	gull	KI	522	479	519								
CC	11	20230811	12:50:12	13:50:00	00:00:12	GX	GX	532	602	666	388							
CD1	11	20230811	12:55:08	13:55:00	00:00:08	GX juv?	GX	405	418	497	518	680						
CD2	11	20230811	12:55:22	13:55:00	00:00:22	GX juv?	GX	405	418	497	518	680						
CE	11	20230811	12:57:05	13:57:00	00:00:05	gull	KI	561	502	507	517	539						
CF	11	20230811	12:57:49	13:58:00	00:00:11	gull	KI	571	546	512								

Table S4-1. All stereo and mono tracked birds identified as being matches to the LRF field data, based on the LRF start times. The camera data is UTC time whereas the field data time is BST (UTC+1hr). NB Large jumps in distance since the first LRF reading may be suggestive of the laser beam not hitting the correct target or the bird has flown beyond the camera field of view.

5 Simultaneously estimating flux and flight height distributions using 3D distance sampling methods

5.1 Introduction

In this report the process of determining the flux of birds around the Aberdeen Bay wind farm using camera data is described and the results summarised. Within the wind farm the camera setup captures bird tracks travelling through the area. A model is developed using a distance sampling framework which utilises these camera data to determine the flux of birds within the area. Distance sampling is a means of estimating densities within an area by accounting for detectability by utilising the fact that individuals become less detectable at greater distances from the observer (Buckland et al., 2001). Accounting for detectability in this way allows the estimation of true abundance by accounting for the individuals that were not detected. However, distance sampling requires assumptions that are expected to be broken within this study. An assumption of concern within this setup is that individuals are spread uniformly relative to a point or a line. This is known not to be the case for birds as they tend to fly at a preferred range of flight heights and may horizontally avoid turbines. The non-uniform distribution of birds must be accounted for to ensure that the distance sampling framework can be used. This means that the data should be analysed in three dimensions to account for both this vertical distribution of flight heights as well as the horizontal distribution of distance from the camera setup. The model that has been developed is described in this report before it is then used to perform exploratory analysis using simulated data, which can determine any limitations in the analytical methodology. If the model is deemed to be suitable, raw data can then be analysed to determine the flux of birds within the wind farm.

5.2 Model description

The methods developed were inspired by other work in which distance sampling methodology was used when there was a non-uniform distribution of individuals from the observer (Cox et al., 2011, Marques et al., 2010).

Initially the Cartesian coordinates relative to the position of the camera in space needed to be transformed to spherical coordinates in the form (ρ, ϕ, θ). This consists of the range from the camera ρ, the vertical angle from the camera position ϕ and the horizontal angle θ relative to the plane as shown in Fig. 5-1.

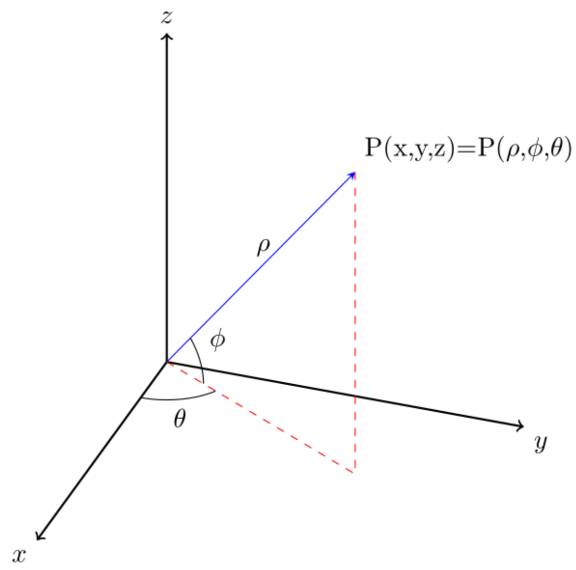

Figure 5-1: Diagram indicating how spherical coordinates are generated from Cartesian coordinates.

The likelihood used in these papers was akin to that given by Buckland et al. (2016), however for the 3D distance sampling setup a correction term was needed to analyse volumes rather than areas.

The likelihood can be summarised as:

$$L_{\rho,\phi} = \prod_{i=1}^{n} \frac{g(\rho_i)\pi_\rho(\rho_i \sin(\phi_i))\rho_i^2 \cos(\phi_i)}{P_a},$$

with $g(\rho)$ the function which describes the detectability of points based upon distance from the camera, ρ, the vertical distribution of flight heights is described by $\pi(\rho)$, and $\rho^2\cos(\phi)$ is the volume correction term.

The likelihood function P_a is:

$$P_a = \int_{min(\phi)}^{max(\phi)} \int_{min(\theta)}^{max(\theta)} \int_{min(\rho)}^{max(\rho)} g\,(\rho)\pi_\rho\big(\rho\sin(\phi)\big)\rho^2\cos(\phi)\; d\rho d\theta d\phi.$$

Dividing by P_a ensures that this is a valid probability density function by integrating across all possible values within the field of view.

The function $g(\rho)$ is the detectability function and in conventional distance sampling a range of functions can be used. It was decided to either use a half-normal or a hazard rate function. The half-normal has one less parameter than the hazard rate but the hazard rate provides greater control of the shape of the detectability function and allows for scenarios in which the probability of detection does not decline rapidly with distance from the camera. The half-normal is specified as:

$$g(\rho) = \exp\left(\frac{-\rho^2}{2\nu^2}\right),$$

whereas the hazard rate function is:

$$g(\rho) = 1 - \exp\left[-\left(\frac{\rho}{\nu}\right)^{-\beta}\right].$$

For both functions ν controls the rate of decline in the detection function and the hazard rate function has a parameter β which allows for the decline in detectability to be flexible in the distance from camera this decline begins. The most appropriate detection function to use will be dictated by the data, therefore model comparison can be used to determine the most appropriate to use.

The flight height distribution, $\pi(\rho\sin(\phi))$, is described either using a truncated normal or a truncated Cauchy distribution. These are both truncated at zero to ensure positive flight heights. The Cauchy will be more appropriate to use when there are thicker tails in the data, i.e. more data at the extremes of the distribution. Once again model comparison can be used to determine the most appropriate to use for each subset of data.

The model was fitted using the *optim* function in R using the Nelder-Mead optimisation method. Volume integrals were calculated numerically using the *integral3d* function from the *pracma* package (Borchers, 2023). Model code is available from https://doi.org/10.5281/zenodo.14975574 .

5.2.1 Density and abundance estimation

The density and abundance can be derived after fitting the model by using the maximum likelihood estimates for the parameters of interest.

5.2.1.1 Density

To estimate the density $\widehat{D}$ within the field of view we use the maximum likelihood estimates to calculate the probability of detection across the full volume V,

$$\widehat{D} = \frac{n\hat{P}_{fh}}{V\hat{P}_a},$$

where $\hat{P}_{fh} = \int_{min(\phi)}^{max(\phi)} \int_{min(\theta)}^{max(\theta)} \int_{min(\rho)}^{max(\rho)} \pi_\rho\left(\rho\sin(\phi)\right)\rho^2\cos(\phi)\, d\rho d\theta d\phi$, which is the integral of the flight height distribution.

The volume of total space covered by the field of view, V, can be calculated using the following equation:

$$V = \int_{min(\phi)}^{max(\phi)} \int_{min(\theta)}^{max(\theta)} \int_{min(\rho)}^{max(\rho)} \rho^2 \cos(\phi)\, d\rho d\theta d\phi,$$

which uses an integral to estimate the volume of a section of a sphere.

To estimate the density of birds within a given height band then the density equation can be used with integrals that assign probabilities of zero for points outside of the chosen height band.

5.2.1.2 Abundance

The estimated true abundance in the field of view, $\hat{N}$, also uses the derived density estimates within the following equation:

$$\hat{N} = \hat{D}V.$$

5.3 Simulation

To understand the limitations of the model, data was simulated under a series of scenarios and the model was fitted to this simulated data.

Within a cube of dimensions 2000x1000x250m, points were simulated with uniform distribution horizontally and vertical spread determined by the vertical distribution of choice. The points were produced with Cartesian coordinates within the cube but these were transformed to spherical coordinates using the *pracma* package in R (Borchers, 2023) and set relative to the camera. The camera height was set to 21.2m above sea level to match the position of the cameras within the field setup. Each point was set either as observed or unobserved with the probability of detection defined by the range, ρ, of that point from the camera according to the chosen detection function. The points marked as observed were also only chosen if they were within the field of view of the camera. The raw data from the offshore facing cameras were used to determine the field of view used in the simulations. This was used to ensure that the geometric setup of the simulated data matched that of the true data as closely as possible. The 3D distance sampling model was then fitted to the simulated points marked as observed.

To obtain an appropriate range of simulation scenarios, Latin hypercube sampling was undertaken. This was used to sample ranges of values for the input parameters. These included; the number of points produced initially within the cube (and so related to the sample size), the population mean μ and standard deviation σ of the flight height distribution, the scale parameter ν of the detection function (e.g. the standard deviation of the half-normal detection function), as well as the shape parameter β of the detection function, if the hazard rate detection function was used. A total of 500 simulation runs were done using the random values for each simulation scenario.

Table 5-1: Summary of range of values used in the simulations, each value chosen via latin hypercube sampling.

Simulation parameter	Range of values used
ν	50-700
μ	10-100
σ	5-50
number of points within cube	7000-500000
β	2-6

5.3.1 Simulation results- half-normal detection function and truncated normal flight height distribution

The analysis was done initially using a combination of the truncated normal distribution, for the flight height distribution, and the half-normal detection function. The results will be analysed in terms of the accuracy of both the parameters for the detection function and the flight height distribution before attempting other simulation scenarios, such as if the true flight height distribution is a truncated Cauchy and is analysed using a truncated normal, or the impact of using a smaller field of view on the sample sizes required.

5.3.1.1 Half-normal detection

For the model using the half-normal detection function and the truncated normal, the output of the 500 simulation runs, as shown in Fig. 5-2 shows that the model estimates for ν tend to be accurate, except when the sample sizes are low, especially at larger values of ν. Fig. 5-3 demonstrates this in more detail with the sample size required differing based on the size of the ν. The lower the value of ν the fewer the number of samples that are needed to avoid bias, but most samples will have a sufficient sample size between 100 and 200 observations.

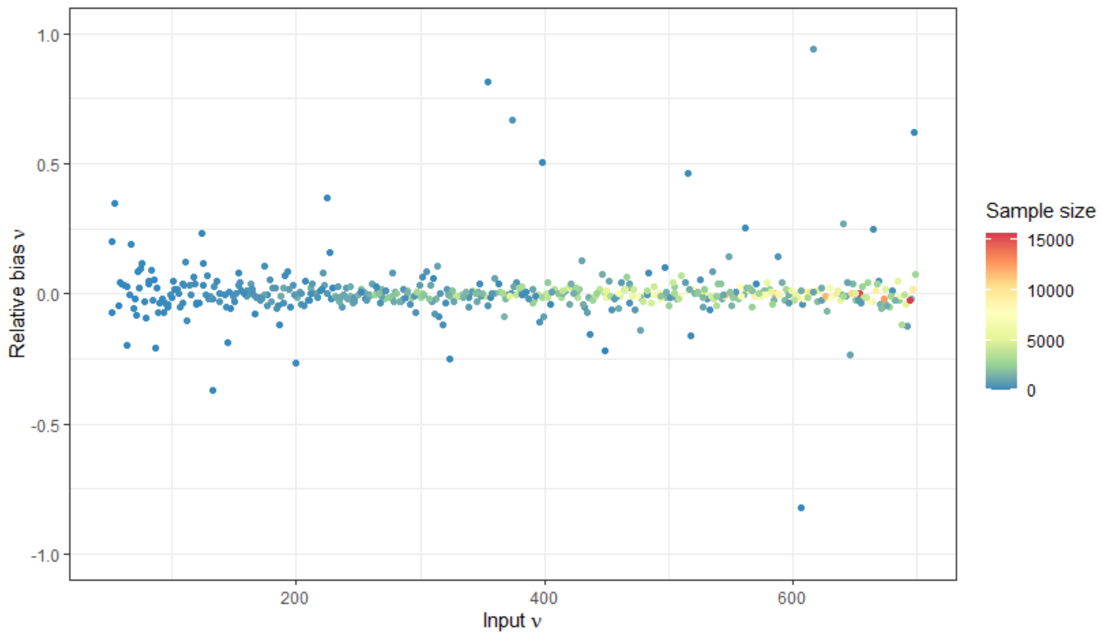

Figure 5-2: Relative bias of the ν of the half-normal detection function against the input value of this parameter used in the simulation. Coloured by sample size of observed points within the field of view.

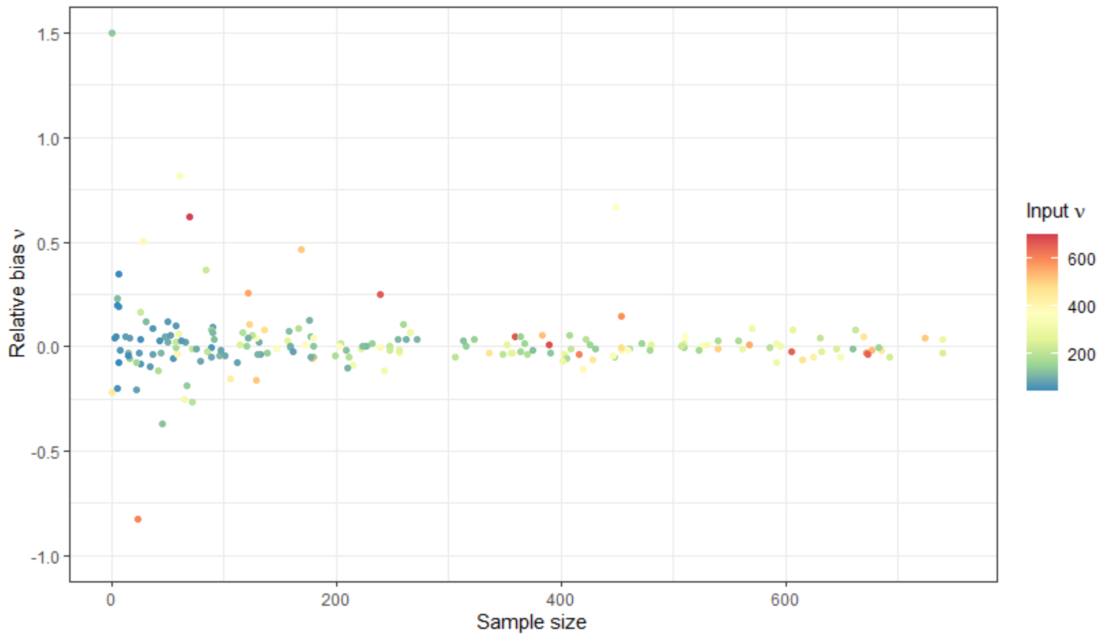

Figure 5-3: Relative bias of the ν of the half-normal detection function against the sample size of observed points in the field of view. Coloured by input ν of the half-normal in the simulation.

5.3.1.2 Flight height distribution

Currently, the field of view given by the data for the inshore camera gives a steadily increasing range of heights above sea level. This means that no point with heights below the camera height (of 21.2m) are being obtained. This is summarised in Fig. 5-4.

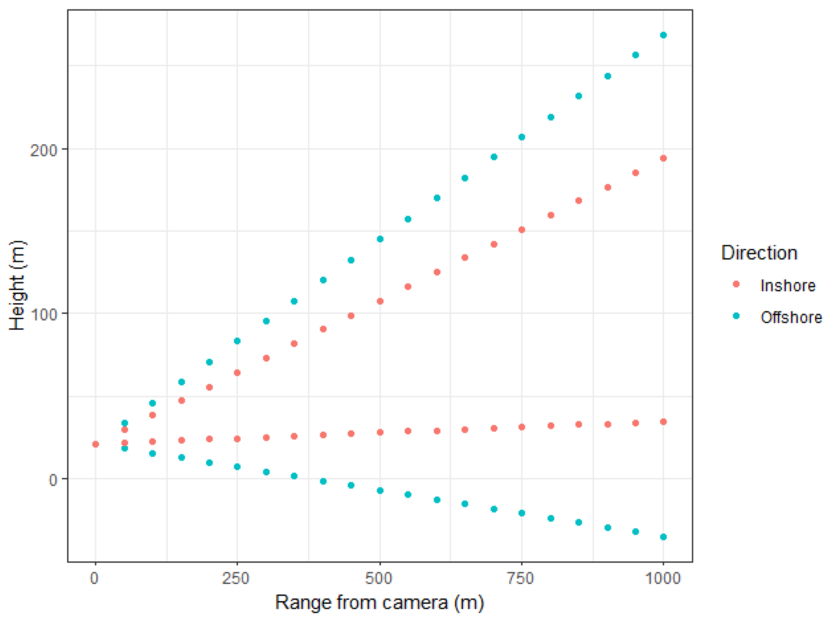

Figure 5-4: Summary of the flight height ranges covered at differing distances from both the inshore and offshore facing cameras. Angles for the vertical field of view were obtained from the raw data (i.e. tracks of birds). The narrower view angle for the inshore camera is due to a combination of a higher focal length, and a higher horizon angle caused by the landmass in the background of the turbine.

This would be expected to impact the flight height distribution estimates obtained from the model in situations when the flight height distribution passes below the camera height. To investigate this within the simulation flight height distributions were created in which the distributions were below the camera height.

From Fig. 5-5, as expected, biases associated with the peak of the distribution are large when the μ values are low. Even larger values of μ have high biases when associated with larger values of σ. Identifiability issues can be seen from this plot by the relative biases approaching -1. This occurs when the estimated value of μ approaches zero. Again these occur mostly when the peak flight height μ is low, or when the associated σ is large.

As can be seen from Fig. 5-6, the sample size needed to reduce bias in μ depends on the input value. Larger values of μ reduced the bias using fewer observations, but the model struggles even at larger sample sizes when the μ is below the camera height. Identifiability issues also tend to occur when the sample sizes are small (< 200).

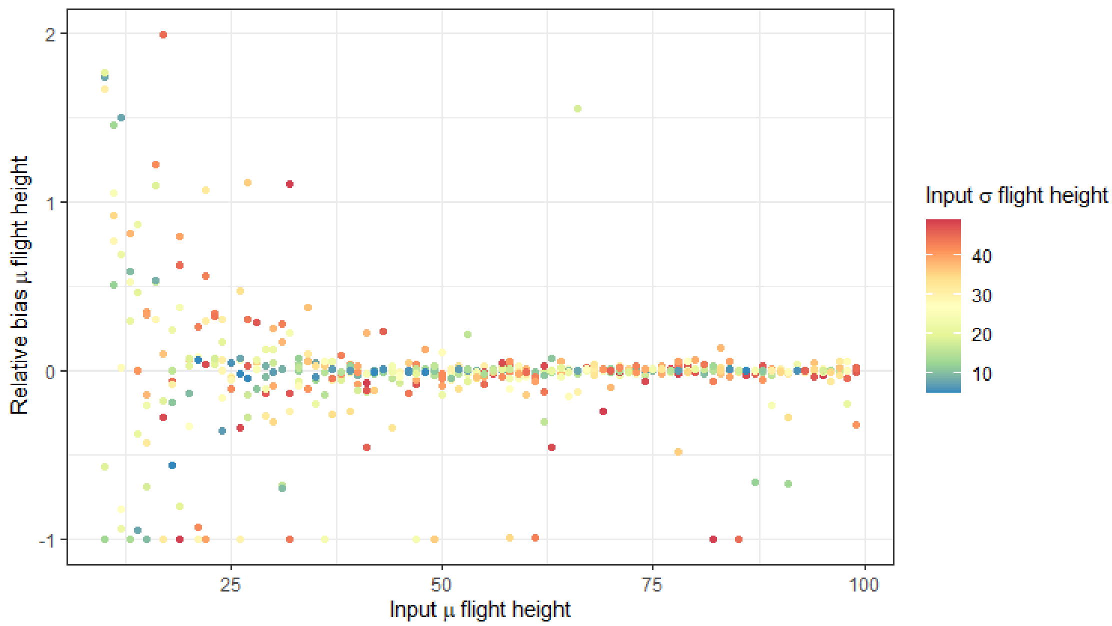

Figure 5-5: Relative bias of the μ of the truncated normal vertical distribution against the input value of this parameter used in the simulation. Coloured by the input σ value of the flight height distribution of the simulation.

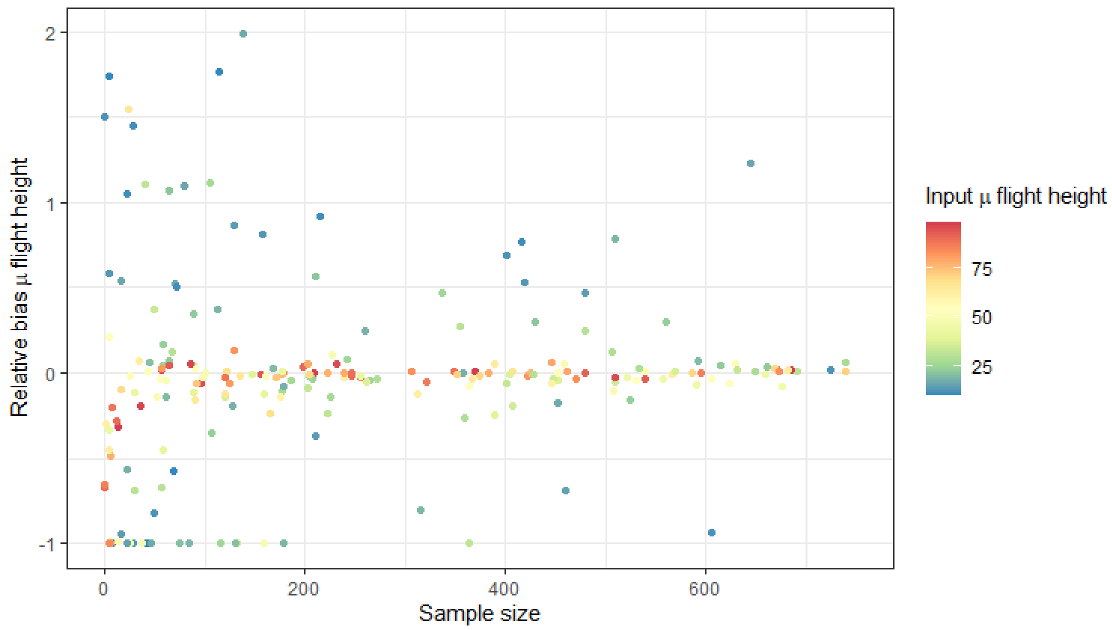

Figure 5-6: Relative bias of the μ of the flight height distribution against the sample size of observed points in the field of view. Coloured by input μ of the flight height distribution in the simulation. Larger sample sizes have been removed.

The bias for σ appears to be relatively small as shown in Fig. 5-7. Fig. 5-8 shows that the sample sizes required to minimise relative bias is smaller for smaller values of σ, however generally samples sizes 200 are enough to minimise bias.

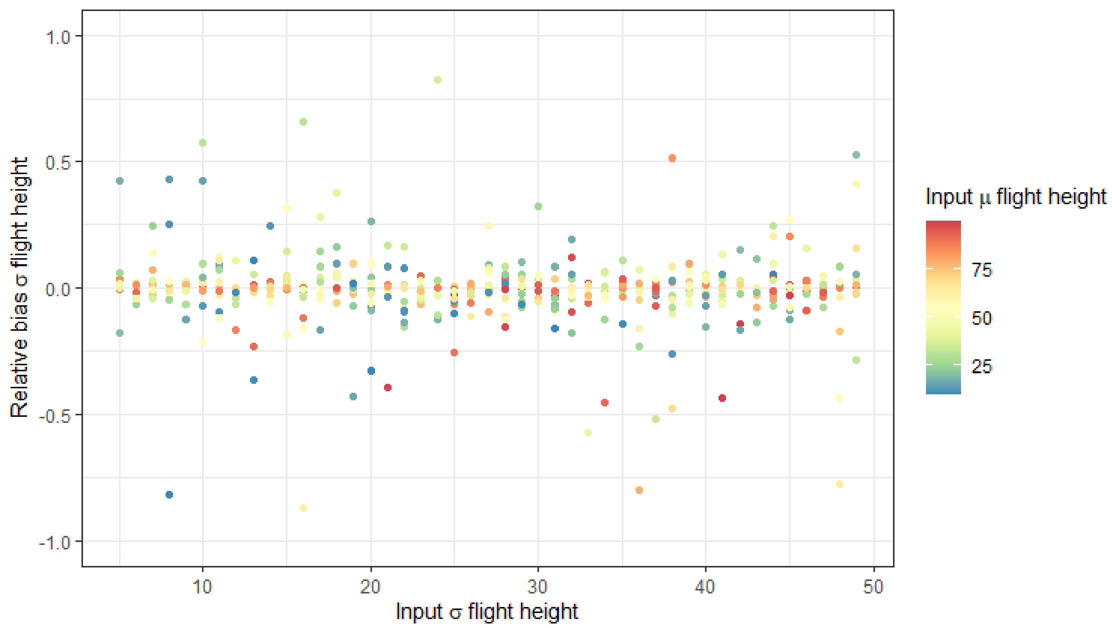

Figure 5-7: Relative bias of the σ of the truncated normal vertical distribution against the input value of this parameter used in the simulation. Coloured by the input μ value of the flight height distribution of the simulation. Larger biases have been cut from the plot.

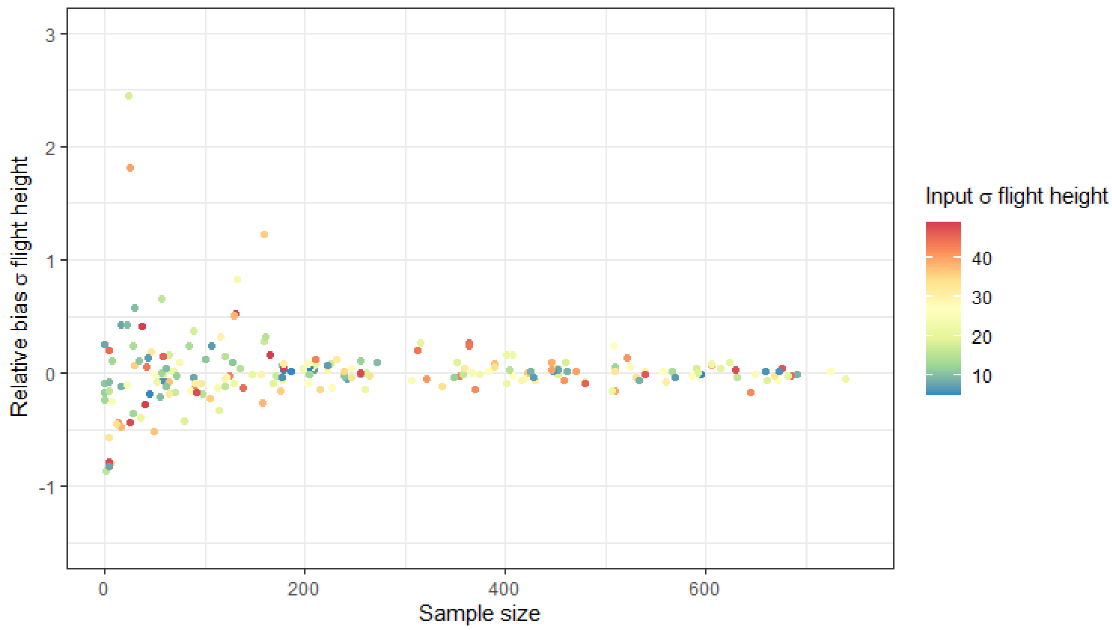

Figure 5-8: Relative bias of the σ of the flight height distribution against the sample size of observed points in the field of view. Coloured by input σ of the vertical distribution in the simulation. Larger relative biases and sample sizes have been removed.

5.3.2 Simulation scenario- modelling using the wrong flight height distribution

To assess the impact of using the wrong flight height distribution data was simulated using a truncated Cauchy, with the range of values shown in Table 1, but then fit to a model with a truncated normal distribution. It was found for the μ parameter that when there were identifiability issues values around -1 were obtained because the μ is estimated at around 0. In Fig. 5-9 it can be seen that there were identifiability issues for a large range of flight height peaks when using the wrong flight height distribution. This occurs even at flight heights above the camera height.

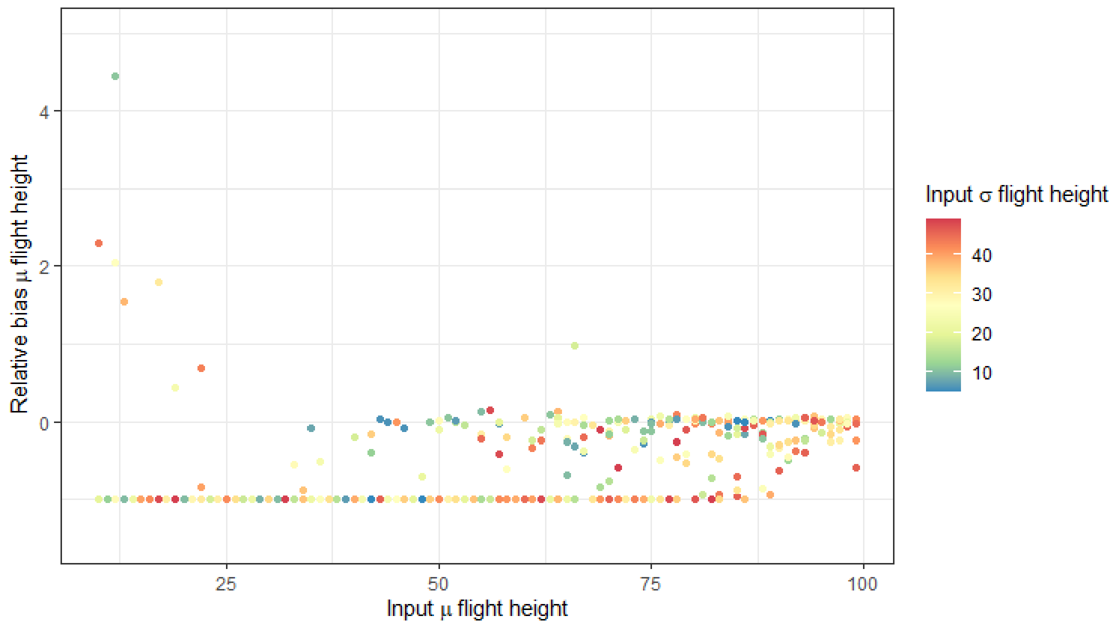

Figure 5-9: Relative bias of the σ of the truncated normal vertical distribution against the input value of this parameter used in the simulation. Coloured by the input μ value of the flight height distribution of the simulation when data is simulated using a truncated Cauchy distribution. Larger biases have been cut from the plot.

When looking at the associated σ for the flight height distribution in Fig. 5-10 it can be seen that there is a significant positive bias especially for lower values of the σ and when the peak flight height is smaller.

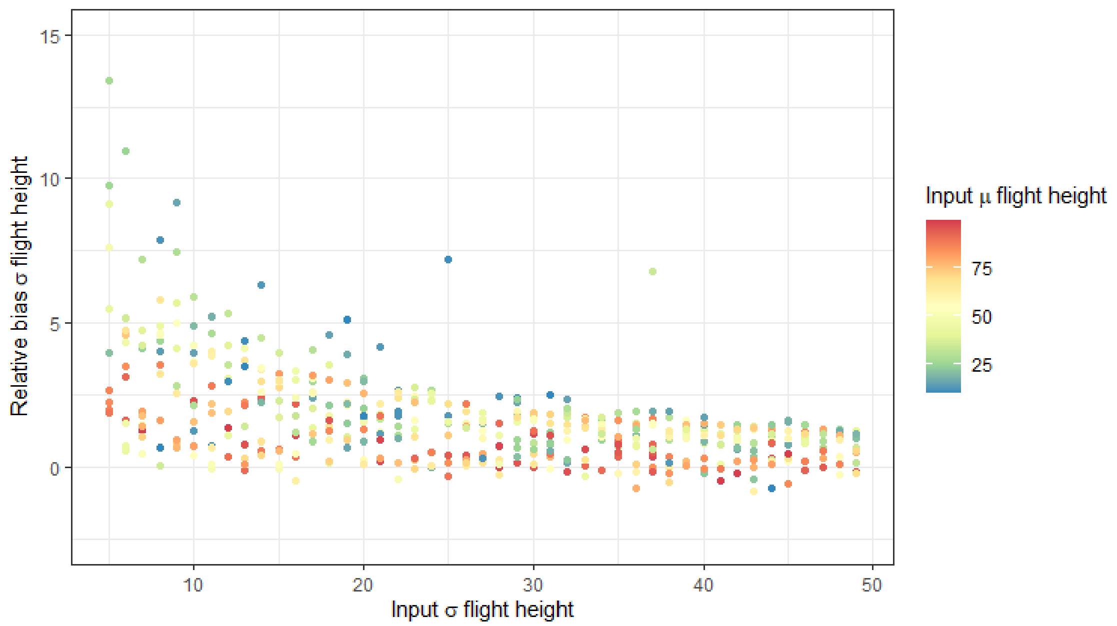

Figure 5-10: Relative bias of the σ of the truncated normal vertical distribution based on the input value of this parameter used in the simulation. Coloured by the input μ value of the flight height distribution of the simulation when the data are simulated using a truncated Cauchy distribution. Larger biases have been cut from the plot.

Overall, these results suggest that the choice of the vertical distribution can have a large impact on the identifiability and biases associated with these parameters. Therefore, it is important to analyse with the true flight height distribution and the identifiability issues that we are obtaining from the raw data suggest that this could be occurring.

5.3.3 Simulation scenario- modelling using the wrong detection function

Similarly to the flight height distribution there could potentially be issues with specifying the wrong detection function. Therefore, in this scenario the data will be simulated using a hazard rate detection function but fit to a model that specifies a half-normal detection function. Potential biases were then explored in the detection function obtained. Fig. 5-11 shows that at low values of ν there are large biases occurring, the value of ν at which the values decrease depends on the β value of the hazard rate function, with smaller values indicating larger biases at higher values of ν. Again this indicates the importance of using the most appropriate detection function within the analysis.

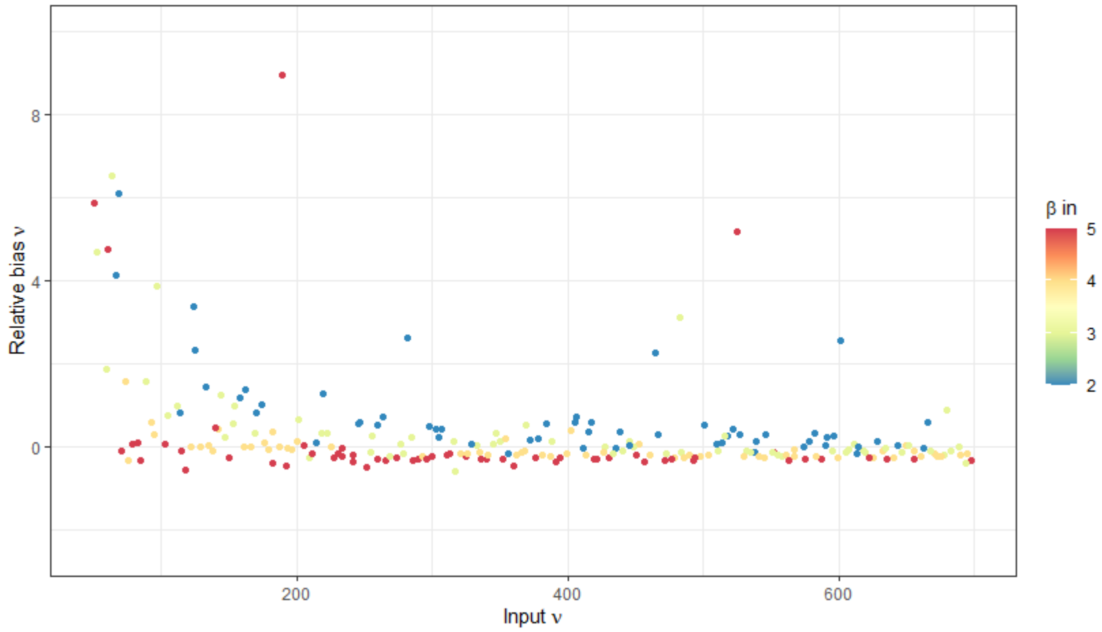

Figure 5-11: Relative bias of the ν of the detection function against the input value of this parameter when the data is simulated using a hazard rate detection function and fitted to a model with a half-normal detection function. Coloured by the input β value of the hazard rate detection function. Larger biases have been cut from the plot.

5.3.4 Simulation scenario- modelling using a narrower field of view

So far, the data has been simulated using the minimum and maximum values from the raw data from the offshore cameras as the range of the vertical field of view. However, when using the raw tracks data will also be collected from the inshore facing camera pair, which have a narrower field of view than the offshore facing cameras. Therefore, a simulation was done using the raw data from the inshore facing cameras as well, to check the sample sizes required for this camera setup. Fig. 5-12, 5-13 and 5-14 indicate that the sample sizes are similar in this field of view compared to the larger offshore field of view.

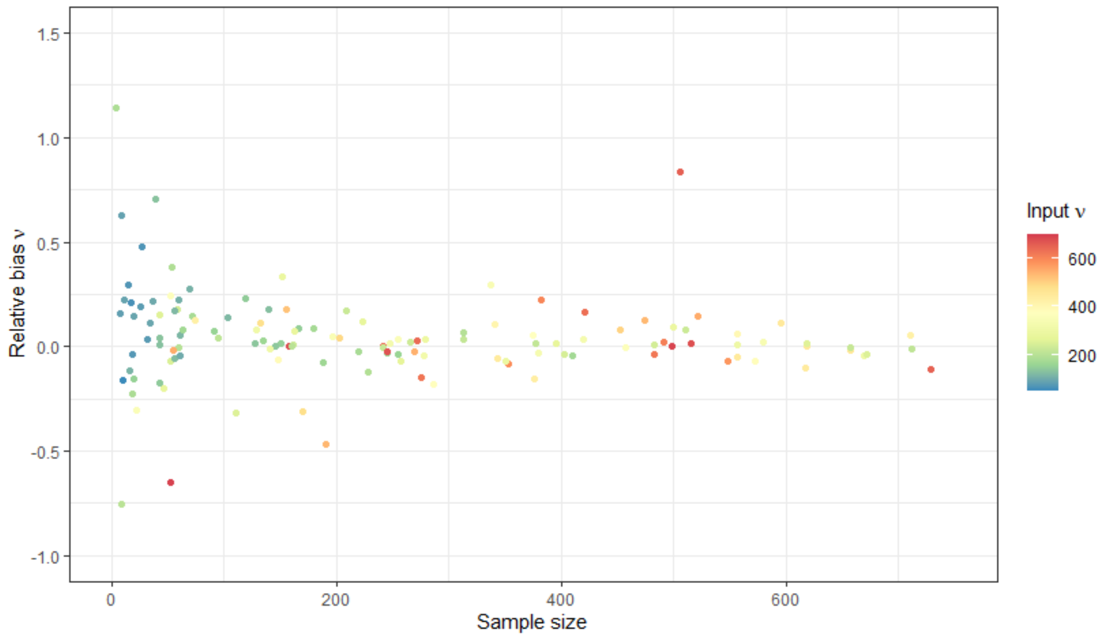

Figure 5-12: Relative bias of the ν of the half-normal detection function against the sample size within the field of view. Coloured by the input ν value used to simulate the data. Larger biases have been cut from the plot.

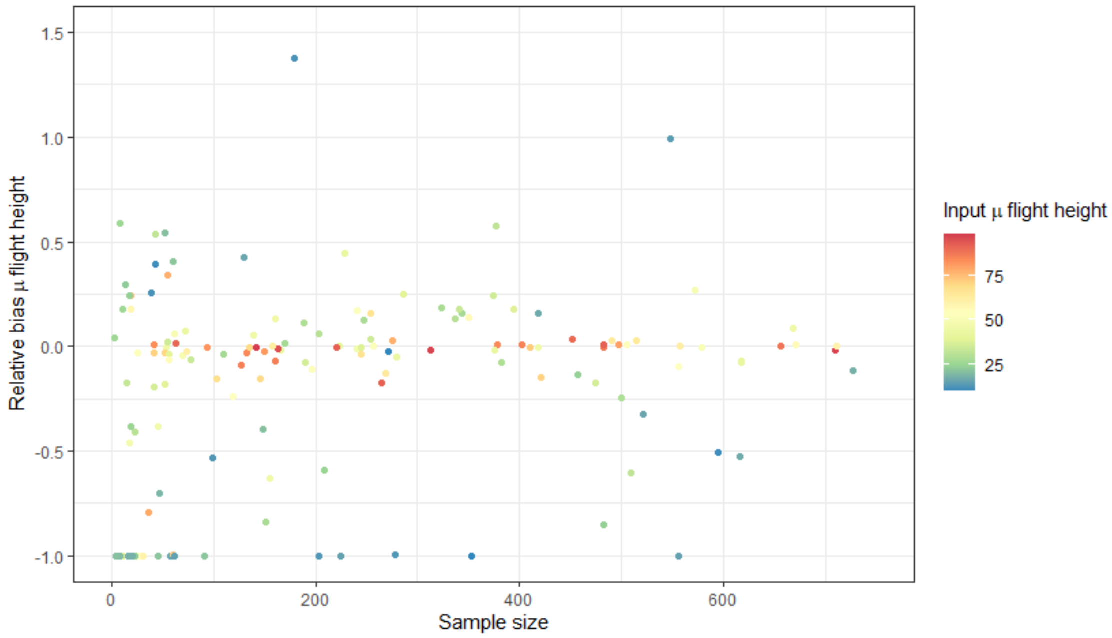

Figure 5-13: Relative bias of the μ of the truncated normal vertical distribution against the sample size within the field of view. Coloured by the input μ value of the flight height distribution of the simulation. Larger biases have been cut from the plot.

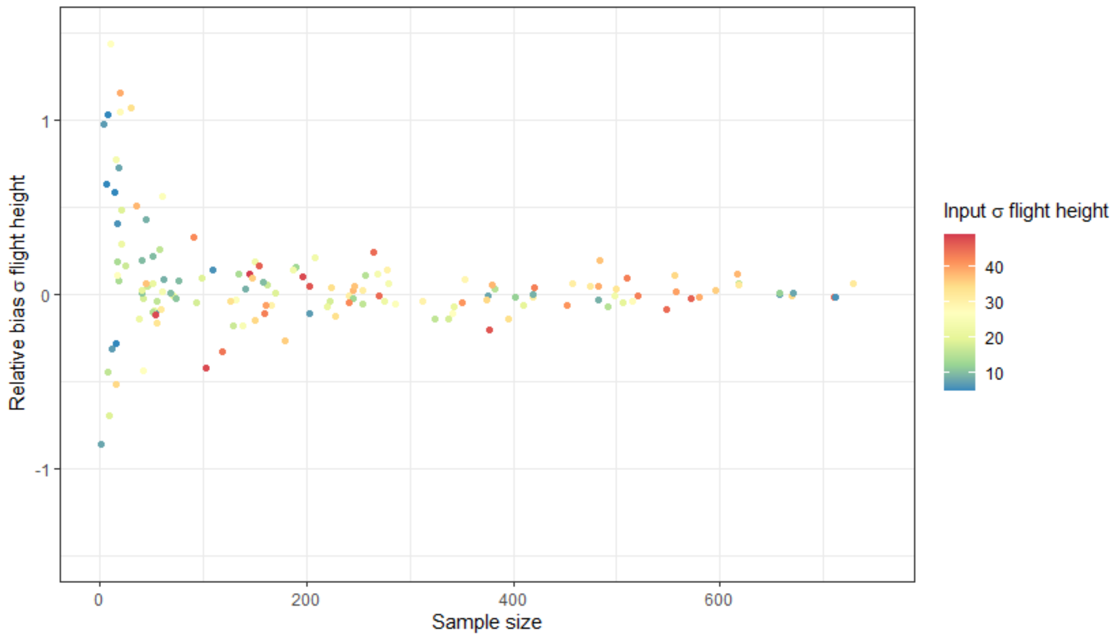

Figure 5-14: Relative bias of the σ of the truncated normal vertical distribution against the sample size within the field of view. Coloured by the input σ value of the flight height distribution of the simulation. Larger biases have been cut from the plot.

5.3.5 Simulation summary

- Generally the 3D distance sampling methodology works well with unbiased estimates of the three parameters used in the model.
- Occasions in which biased results or identifiability issues occur is when the peak flight height is not within the field of view of the camera, or when the associated standard deviation around the peak flight height is large and thus extends beyond the field of view of the camera.
- Greater biases and identifiability issues occur when the data is fitted to a model that has a different detection function or flight height distribution to that found within the data. Several options can be used in the modelling framework and model comparison can be used to determine the most appropriate combination to use.
- Generally a sample size of 150-200 is enough to remove the majority of biases, although this can be dependent on the true values of the parameters, with lower sample sizes needed at higher flight height peaks, at lower standard deviations around this and lower values of the ν in the detection function.

5.4 Raw data analysis

The raw data consists of tracks of birds which have passed by the camera setup in the wind farm. For the 3D distance sampling analysis individual points were used rather than tracks. From each track a single point was chosen randomly to be used in the analysis. Raw data were obtained from four cameras on the wind turbine, two of which face inshore and two offshore. The inshore and offshore cameras also have differing fields of view and therefore it was decided to subset the data based upon camera direction. Tracks were also analysed at the species level to provide species level estimates of flux and flight height. From the simulation studies it was decided that at least 150 tracks will be collected for analysis for each of European Herring Gull, Black-legged Kittiwake and Northern Gannet for both the inshore and offshore facing cameras. To determine if the model is producing robust results identifiability of the parameters can be checked by obtaining the scaled minimum eigenvalues, which, if small, suggest identifiability issues and that the estimates may not be robust (Chis et al., 2016, Cole and Morgan, 2010).

Table 5-2 summarises the amount of data gathered for each species for each camera pair, including all species grouped together and all gull species.

Table 5-2: Summary of number of tracks within the field of view of each camera pair for each species or species group.

Species	Tracks- Offshore facing	Tracks- Inshore facing
All	1675	1675
Gull	1264	1578
European Herring Gull	652	986
Black-legged Kittiwake	252	250
Northern Gannet	416	92

The data was then fitted to the 3D distance sampling model after being subset by species and camera pair. The model was fitted using the half-normal detection function and the truncated normal distribution to describe the flight height distribution. The values of the limits used in the integrals were obtained from the raw data, including the maximum and minimum values of ϕ and θ and the maximum value of ρ. The minimum value of ρ was set as 0. Table 5-3 summarises the results obtained from the model for each species and camera pair.

Table 5-3: Estimates for species from the 3D distance sampling model for each camera pair.

Species	Camera Pair	Mean flight height	σ flight height	ν detection	Density
HG	Offshore	79.01 (70.66, 87.37)	39.77 (30.94, 48.59)	263.29 (229.8, 296.79)	807.34 (687.79, 968.98)
HG	Inshore	74.04 (67.78, 80.31)	47.79 (38.52, 57.06)	479.47 (424.44, 534.5)	571.4 (506.22, 669.12)
KI	Offshore	38.42 (36.2, 40.63)	15.26 (13.35, 17.17)	251.88 (194.61, 309.15)	206.38 (167.84, 262.55)
KI	Inshore	NA (NA, NA)	NA (NA, NA)	NA (NA, NA)	NA (NA, NA)
GX	Offshore	43.73 (41.01, 46.44)	21.91 (19.64, 24.18)	499.35 (408.19, 590.51)	28.97 (25.05, 36.45)
GX	Inshore	0.00032 (-0.12, 0.12)	40.94 (34.38, 47.49)	759.93 (356.17, 1163.69)	42.81 (34.49, 76.86)

Table 3 demonstrates the difficulty that was encountered using the inshore facing camera data for the Kittiwake and Gannet data. The Kittiwake model was not identifiable and the Gannet model provided a flight height result that was similar to those seen in the simulation studies when the model was not identifiable due to the mean flight height being below the field of view of the camera. It was believed that this could be due to potential masking of individuals in the camera field of view due to the land, as well as the smaller field of view of the inshore facing cameras. It was therefore decided to concentrate on the results from the offshore facing cameras. The results from these cameras indicate that the mean flight height is largest for the Herring Gull and lowest for the Kittiwake. This is in agreement with other studies that state that the Herring Gull flies at greater heights than other gull species and the Gannet (Jongbloed, 2016) . It was also found that the Gannet exhibited greater detectability at larger distances which was expected as this species is larger and therefore easier to detect at these distances. Herring Gull was also shown to have the greatest densities in the observed field of view whereas the Gannet occurred in the lowest densities.

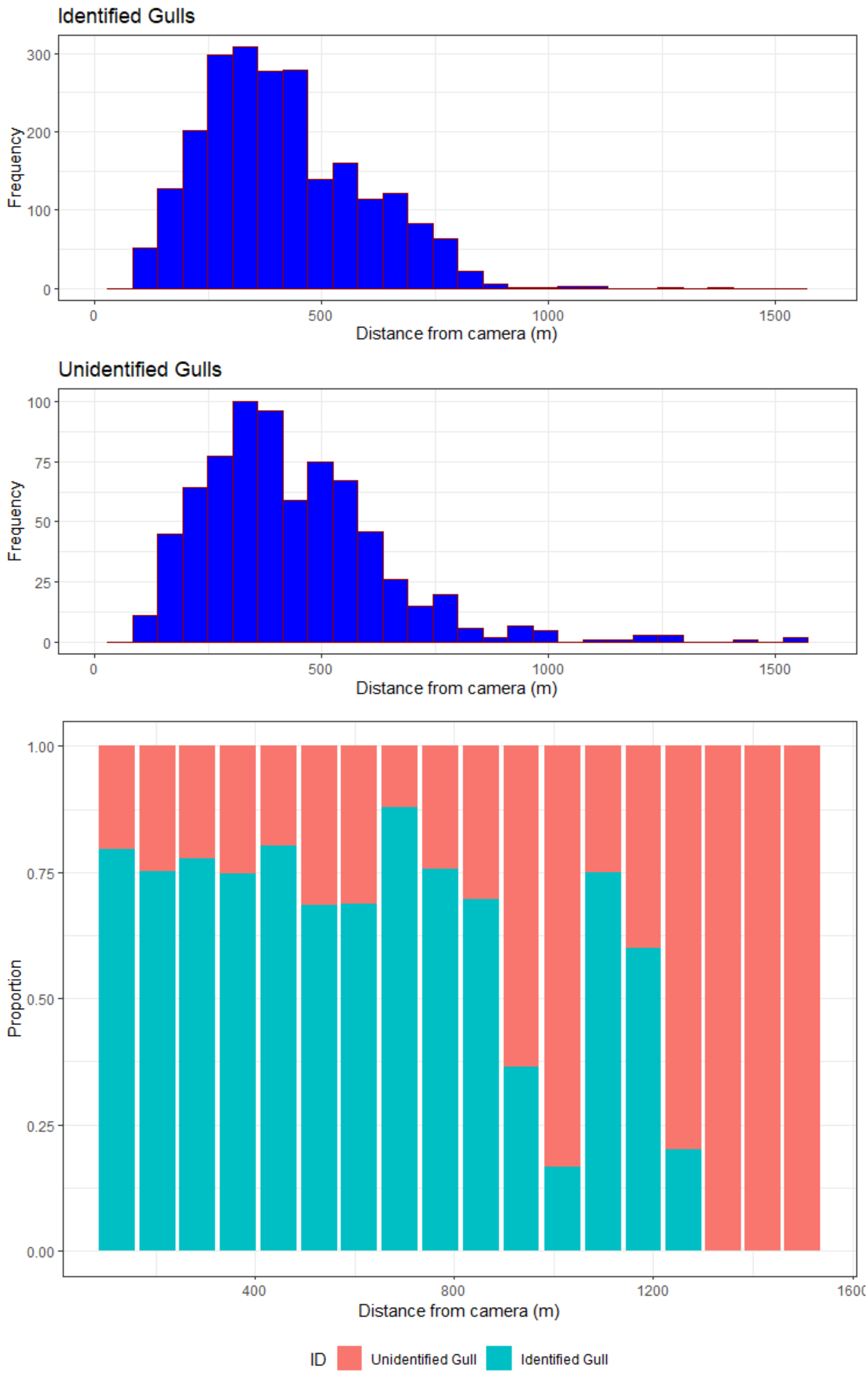

Figure 5-15: Detectability and density estimates derived by the distance sampling model are conditional on birds being tracked and identified to species level. The identification process appears to be the limiting factor in this workflow, as the proportion of unidentified birds increases with nominal distance from the camera.

5.4.1 Weather covariates

The bird flight track data were collected on different days across the study. These will have different weather conditions associated with them and could impact the detectability of species as well as the flight height of species (Jongbloed, 2016; Aschwanden et al., 2024). Therefore, it was decided to attempt various models that combined placing a weather covariate on the ν of the detectability function, the μ of the flight height distribution and the σ of the flight height distribution. This will be done by adding weather condition categorical data to the model. Table 5-4 indicates the constitution of these weather covariates for each species.

Table 5-4: Summary of number of tracks within the field of view of the offshore facing camera for all species, Herring Gull, Kittiwake and Gannet for the weather condition categories.

Weather Condition	Tracks-All Species	Tracks-Herring Gull	Tracks-Kittiwake	Tracks-Gannet
Clear	169	16	40	63
Fair	92	26	9	35
Cloudy	160	63	15	51
Overcast	498	183	82	57
Rain	431	300	18	36
Fog	15	0	0	12
Light Rain	106	37	11	29
Light Snowfall	20	7	0	0
Rain Shower	4	0	1	2
Snow Shower	16	2	0	0
Snowfall	9	4	2	0

It is expected that for weather conditions that restrict visibility of the camera the ν parameter will be lower as detection declines more rapidly with distance from the camera and that birds fly higher during clearer weather conditions (Aschwanden et al., 2024). The weather condition data used was from Aberdeen Bay and was collected at hourly temporal scales. Each track was then matched to the weather condition at the nearest hour. The distance sampling model used the half-normal detection function and the truncated normal flight height distribution. When adding the weather covariate, the integral required calculation separately for each data point. This slows down the computation speed of the distance sampling model. The offshore facing cameras were used as these had fewer identifiability issues compared with the inshore facing cameras and therefore will be more robust to these additions. Furthermore, it was decided to only use the categories which had enough data to fit to the distance sampling model. This included joining categories which we believed were similar, such as the cloudy and overcast, and clear and fair categories into two distinct categories.

Data was subset by species and it was decided that there was enough data for two weather conditions for each species. The Herring Gull was analysed using the rain (300) and cloudy/overcast (246) categories. The Kittiwake and Gannet data were analysed using clear/fair and cloudy/overcast categories for each. For the Gannet the clear/fair category had 98 tracks and the cloudy/overcast category had 108 tracks. For the Kittiwake the clear/fair category had 49 tracks and the cloudy/overcast category had 97 tracks.

Table 5-5 indicates the outputs for each species and the values of the detection function and the flight height under the different weather conditions. A single σ was estimated for the flight height distribution for each species model.

Table 5-5: Estimates for species from the 3D distance sampling model using data from the offshore facing camera pair for the different weather conditions in a model that allows variation in the mean flight height and nu detection based on weather conditions.

Species	Weather	Mean flight height	σ flight height	ν detection	Density
HG	Rain	88.84 (79.86, 97.82)	32.51 (26.56, 39.77)	404.47 (215.54, 593.41)	860.84 (708.14, 2257.19)
HG	Cloudy/overcast	53.37 (40.12, 66.61)	32.51 (26.56, 39.77)	210.93 (19.95, 401.91)	1181.88 (951.49, 1518.28)
GX	Clear/fair	43.76 (38.97, 48.55)	20.18 (17.24, 23.12)	484.2 (321.49, 646.9)	34.45 (27.67, 58.96)
GX	Cloudy/overcast	39.77 (33.2, 46.35)	20.18 (17.24, 23.12)	541.14 (271.61, 810.67)	29.5 (23.85, 54.85)
KI	Clear/fair	38.35 (33.47, 43.24)	15.03 (12.63, 17.43)	208.78 (143.16, 274.41)	203.94 (144.67, 420.82)
KI	Cloudy/overcast	38.42 (32.43, 44.41)	15.03 (12.63, 17.43)	283.94 (147.25, 420.62)	232.51 (182.28, 491.01)

A significantly higher mean flight height was seen for the Herring Gull in rainy conditions compared with cloudy/overcast conditions. Other studies have shown that birds tend to fly at lower altitudes in rainy conditions (Aschwanden et al., 2024; Hüppop et al., 2006). Gannets fly higher in clear/fair which agrees with what has been seen in other species (Aschwanden et al., 2024). The detectability at greater distances was best for the Gannet and when there is better visibility. The detectability of the Herring Gull is greater at larger distances in the Herring Gull in rainy conditions than for overcast conditions, although the 95% confidence intervals overlap to a large extent, similar to what is occurring for the Gannet.

Table 5-6: AIC values for the models which vary by how the weather condition covariate was added, either not added, added to the ν of the detection probability only or added both to the detection probability as well as the μ of the flight height distribution.

	No covariates	Weather on detection	Weather on detection and flight height	Weather on detection and mu and sigma flight height
HG	3565.84	3507.34	3472.34	3475.40
GX	1424.82	1426.74	1425.53	1427.49

	No covariates	Weather on detection	Weather on detection and flight height	Weather on detection and mu and sigma flight height
KI	910.09	912.05	914.01	913.16

Table 5-6 shows that the model with no weather covariate is favoured by both the Kittiwake and Gannet, however a model that contains a weather covariate on both the ν and μ terms is favoured for the Herring Gull data.

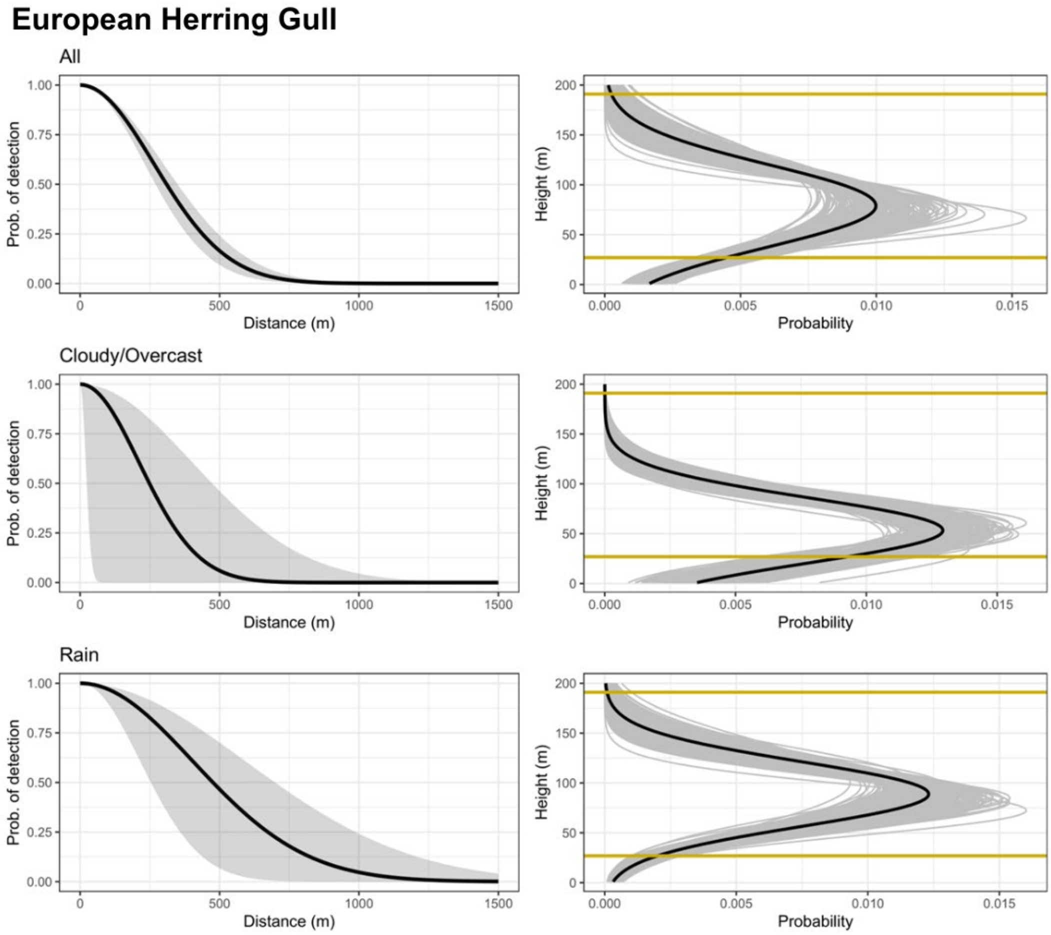

Figure 5-15: The estimated half-normal detection function and truncated normal flight height distribution with associated 95% confidence intervals for the full European Herring Gull dataset as well as separated by the weather condition. For the flight height distribution the extent of the rotor swept zone within Aberdeen Bay is within the yellow lines.

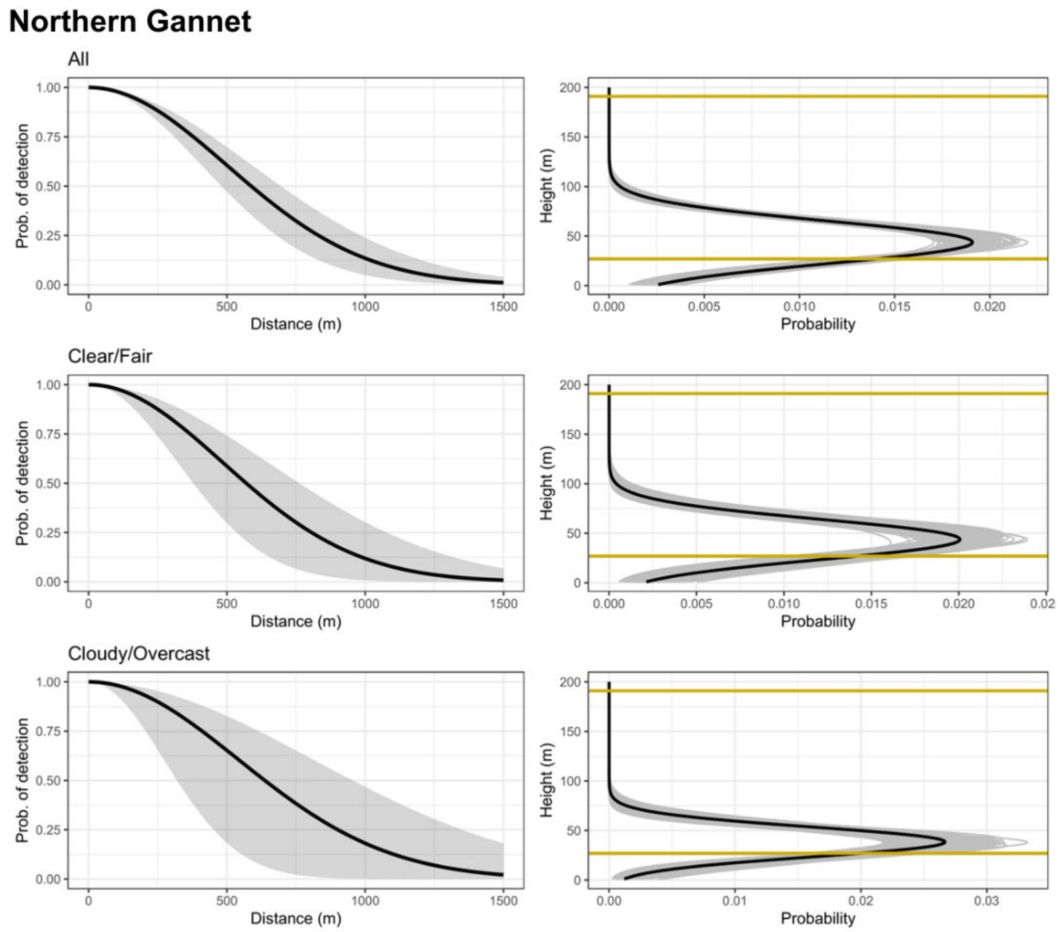

Figure 5-16: The estimated half-normal detection function and truncated normal flight height distribution with associated 95% confidence intervals for the full Northern Gannet dataset as well as separated by the weather condition. For the flight height distribution the extent of the rotor swept zone within Aberdeen Bay is within the yellow lines.

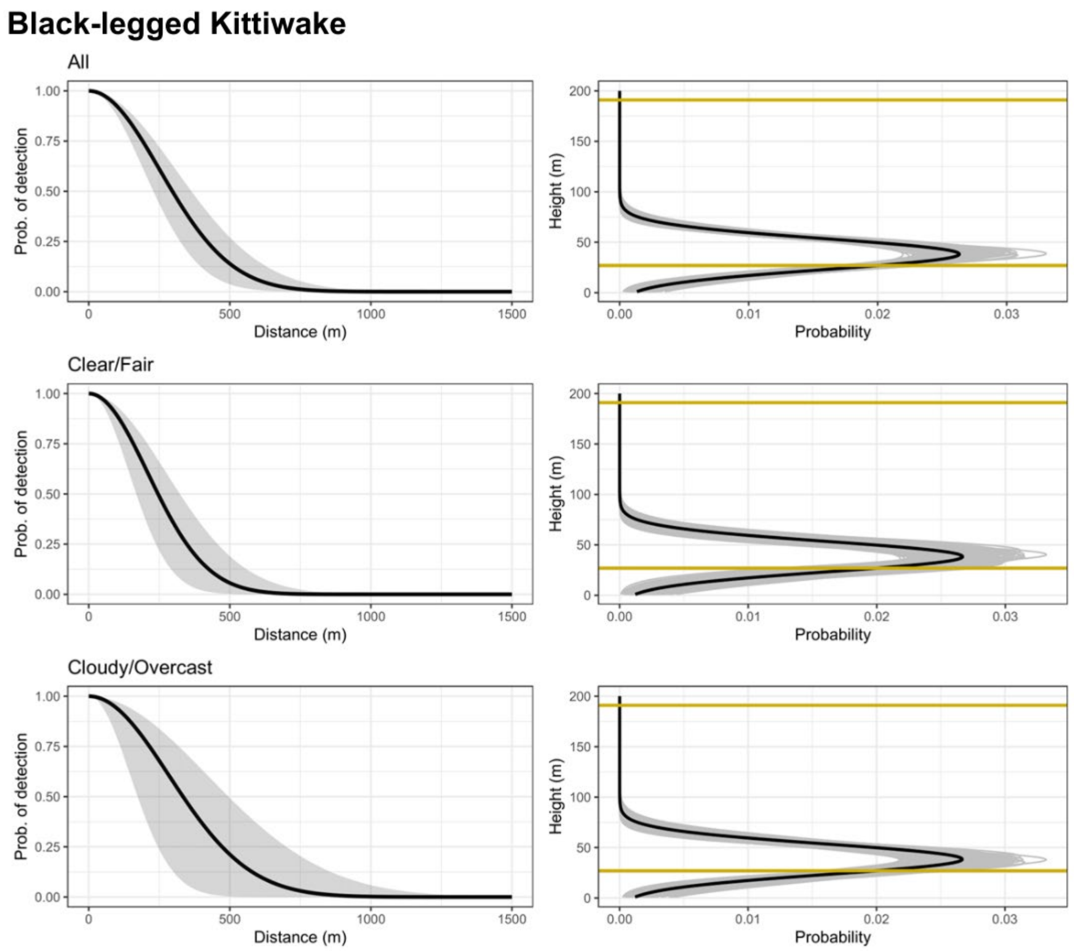

Figure 5-17: The estimated half-normal detection function and truncated normal flight height distribution with associated 95% confidence intervals for the full Black-legged Kittiwake dataset as well as separated by the weather condition. For the flight height distribution the extent of the rotor swept zone within Aberdeen Bay is within the yellow lines.

5.4.1.1 Flux within rotor swept zone

The flux within the observed field of view has been calculated already but using the flight height distribution we are able to determine flux within different height bands. Therefore, it is possible to calculate the amount of flux within and outside the rotor swept zone. The rotor swept zone for these turbines is between 27-191m above sea level. Therefore, the flux below this height range for each species would be the flux outside the rotor swept zone.

The flux within and under the rotor swept zone can only be calculated using the data from the offshore facing cameras because as can be seen from Fig. 5-4 the field of view of the inshore facing cameras do not contain much data beneath the rotor swept zone.

Fig. 5-18 shows the estimated flight height distribution for the three species analysed with the area of the rotor swept zone indicated between the yellow lines to indicate the extent of the birds flying at these heights. The flux is also shown for the full field of view as well as within and under the rotor swept zone for the full dataset, as well as in the different analysed weather conditions.

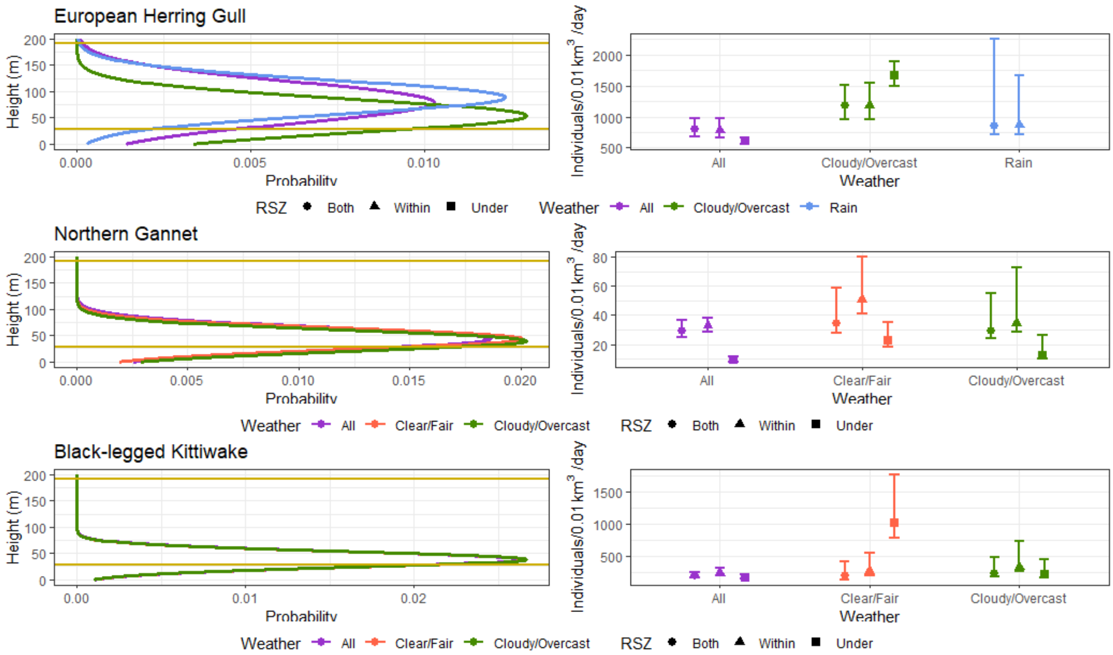

Figure 5-18: The outputs of the 3D distance sampling model for the Herring Gull, Gannet and Kittiwake with the truncated normal flight height distribution for the separate weather conditions as well as the full dataset. Additionally, there is the density estimates for the separate weather conditions as well as the full dataset as well as the density estimates within and under the rotor swept zone of the turbines in Aberdeen Bay.

Although the data is used for below the rotor swept zone for the offshore facing cameras the area within the field of view is relatively small in this area. This means that the density estimates can be sensitive to small increases in the number of birds within this area. Therefore, we would recommend that the estimates be used with caution and if the flux within and outside the rotor swept zone is of particular interest then a setup in which the field of view takes in the area below the rotor swept zone to a greater extent.

5.5 Simulation - accounting for observation error

The analysis has thus far been undertaken with the assumption that there is no observation error in the position of points used in the distance sampling analysis. However, it is likely that this is not the case. When comparing tracks produced from the mono-vision cameras to the same tracks from stereo-vision cameras an error has been discovered in the estimation of the points. The stereo-vision is better able to estimate positions than the mono-vision setup and therefore when comparing the two the stereo-vision was taken as being the true position of the bird in space. Currently it is not possible to carry out the distance sampling analysis on the stereo-vision data because there has not been enough data collected from these cameras. To determine the size of the observation error when using the mono-vision cameras the relationship between the stereo-vision and the mono-vision cameras can be determined using linear regression. This was carried out in a Bayesian framework and predictions from the posterior were made to produce a set of 1,000 different estimates of the true distances for the raw Kittiwake data for the offshore facing camera. These 1,000 datasets were used along with the distance sampling model to produce new maximum likelihood estimates for the three parameters within the model. The maximum likelihood estimate for each was compared to the maximum likelihood estimate from the uncorrected mono-vision data to determine biases in the three parameters, as well as biases in the derived density estimates.

Although we are explicitly accounting for range error it is also understood that there will be an error in the flight height estimates as well. Fig. 5-19 shows the extent of the flight height error in relation to the range error for the range of the vertical field of view from the Kittiwake data for the offshore facing cameras, plus when the angle is 0, which is when the bird is spotted directly ahead of the camera.

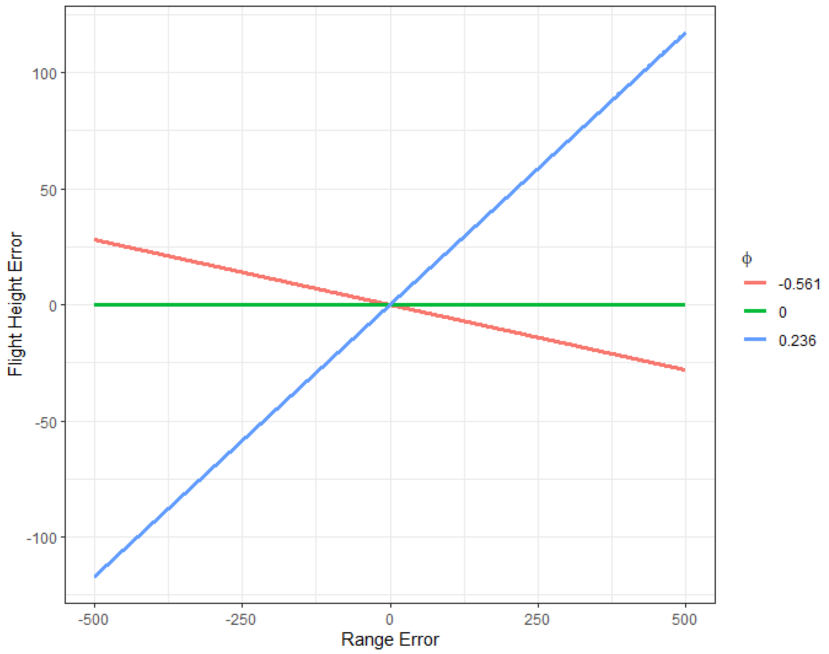

Figure 5-19: The change in absolute flight height error for a range of absolute range error values dependent on the input vertical angle relative to the central axis of the camera.

This shows that as the range error goes from negative to positive for points above the central axis that the flight height will go from being underestimated to overestimated, whereas the opposite occurs for birds seen below the central axis. This is accounted for in the simulation as the updated range values are used to recalculate the flight height while assuming that the vertical angle remains constant.

The relative biases for the three parameters as well as the derived density estimates were calculated from the 1,000 simulation runs. Fig. 5-20 indicates these biases.

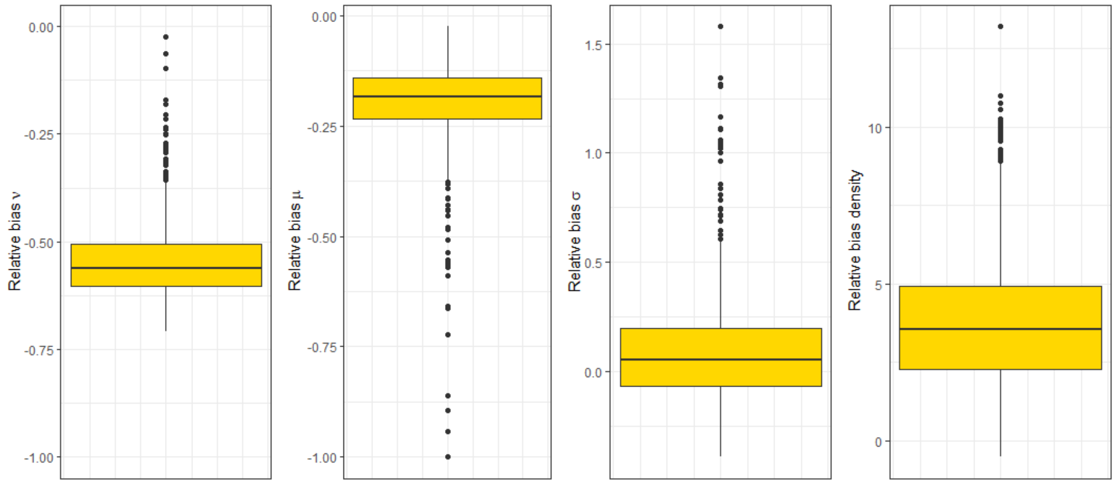

Figure 5-20: Relative biases calculated across 1,000 simulation runs for the ν, μ and σ parameters as well as the derived density estimates. The true values were obtained from uncorrected mono-vision data for the Black-legged Kittiwake for the offshore facing camera, these were compared to the corrected estimates for each data point for the same dataset but the correction was obtained from a posterior prediction from a Bayesian linear regression of the mono-vision against the stereo-vision ranges for tracks from the raw data. This was done using 1,000 different posterior predictions.

The relative biases indicate that without accounting for observation error we would be overestimating both ν and μ, whereas we would be underestimating the density estimates produced. For σ there is also a slight bias which leads to underestimation.

5.6 Simulation - accounting for avoidance

The distance sampling model also makes the assumption that there is no avoidance of the wind turbines of the birds. There are different spatial scales of avoidance. These are described as macro, meso and micro-avoidance. Macro-avoidance is when birds will avoid the wind farm altogether. This cannot be accounted for as we only view birds within the wind farm environment. Micro-avoidance are small scale movements of the bird as they avoid a turbine at short distances. This should not impact our distance sampling model as these are just small scale movements at distances that will not be picked up by the cameras. The scale in which we are interested is meso-avoidance. This is between macro and micro-avoidance where birds change behaviour within the wind farm and avoid turbines at distances of tens to hundreds of metres. Currently the models assume that this avoidance does not occur and it is likely that identifiability issues would occur if this avoidance was added to the model, due to the difficulty in separating this from the detection probability which also relies on distance from the turbine. However, simulations can be created using the levels of meso-avoidance that are expected from the three species under analysis and then fit the model that does not account for avoidance and determine if there are any biases or inaccuracies in the outputs based on this assumption.

To simulate the avoidance a variogram function will be used to determine the probability of presence of an individual based upon the distance from the camera. The variogram is specified as being a function of the distance from the camera,

$$h(\rho) = n + (1-n)\left(1 - \exp\left(\frac{-3\rho}{r}\right)\right),$$

with r being the 'range' which is the distance at which there is no avoidance of the turbine, and n which is the 'nugget', which is the probability of presence at a range of 0m from the turbine. These values will be obtained from Tjørnløv et al. (2023) who investigated levels of avoidance of bird species within Aberdeen Bay.

For this simulation a level of avoidance will be produced that matches each of the three species and other values within the simulation will be set as of Table 5-7 with a range of values for ν as the importance of avoidance is expected to depend on these detectability distances. The half-normal detection function was used for these simulations, as well as the truncated normal flight height distribution.

Table 5-7: Summary of values and range of values used in the simulations. For a range of values each value was chosen via latin hypercube sampling.

Simulation component	Range of values used
ν	50-700
μ	50
σ	10
number of points within cube	50000

The values for the 'nugget' and the 'range' for the variogram will be obtained for each species from Tjørnløv et al. (2023). For the Herring Gull this was set to be 0.3 and 100m for the 'nugget' and 'range' respectively. For the Gannet this was 0.5 and 40m, whereas for the Kittiwake this was 0.5 and 150m. This translates into the probability of presence of a bird for each species depending on the distance from the turbine and is summarised in Fig. 5-21.

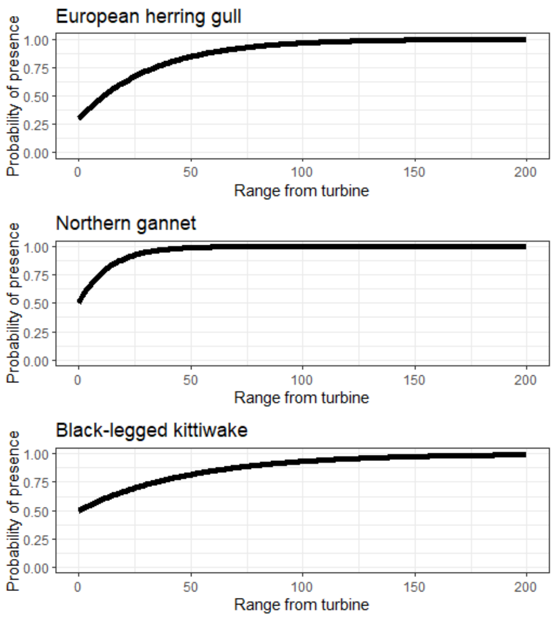

Figure 5-21: The variogram functions to describe how the probability of presence changes with each species as distance from the turbine increases. Values for these were obtained from Tjørnløv et al. (2023)

To account for avoidance a set of data was simulated using values in Table 7. Once these values were simulated then each point was either retained or discarded randomly using a binomial distribution with probability values based upon the variogram for each species. This removed some points that were closer to the turbine from the data. The simulation then, as with the simulations in earlier sections, uses the detection function to determine if individuals were detected or not using these probabilities within another binomial distribution.

Data was simulated 250 times for each species match and each set of simulated data were then fit to the 3D distance sampling model, which doesn't account for avoidance in the modelling. This will show us whether not accounting for avoidance causes any biases in the results obtained from the model. Any biases in the three parameters are shown in Fig. 5-22.

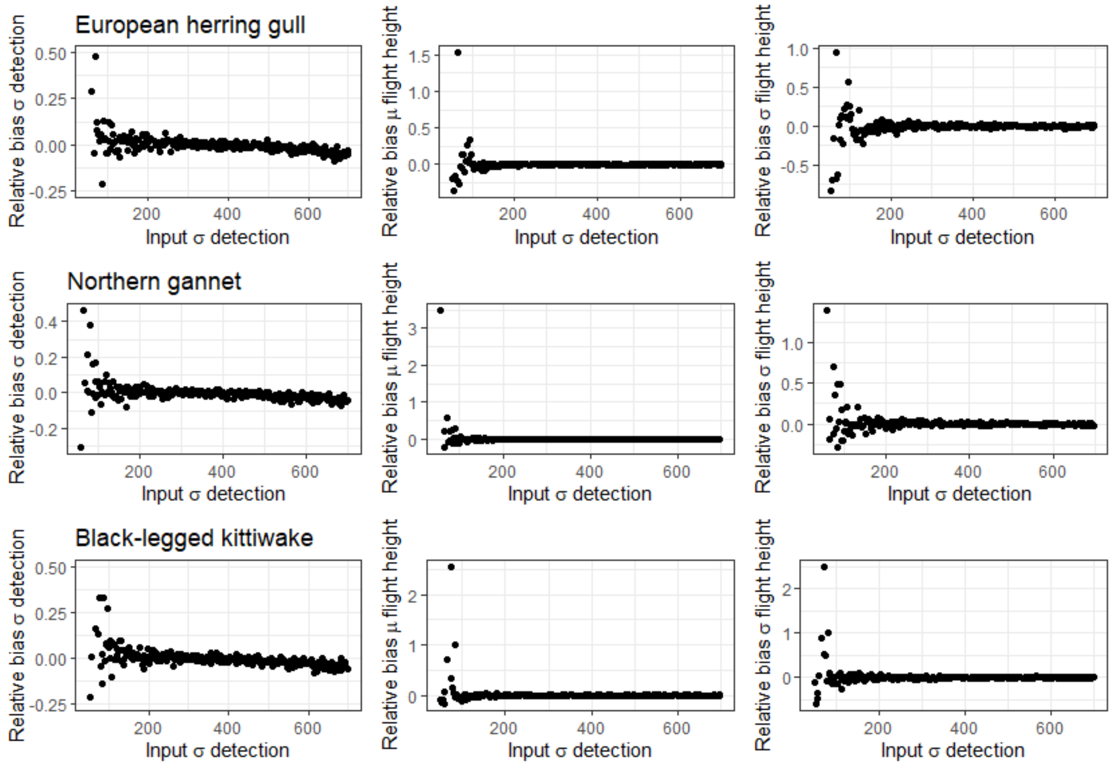

Figure 5-22: Simulation results for each of the matches to the three species showing the relative biases in the estimated parameters based on input values of ν.

From these plots we can see that when the detection range ν is shorter than the avoidance distance then there are inaccuracies for estimates of all three parameters. This should not be an issue for the three species we investigated at this site as the detection range values estimated were greater than the avoidance ranges for these species observed by a previous study in the same OWF. There also appears to be a negative bias at greater values of ν for the estimate of this parameter, meaning that this detection distance is underestimated. Again, these are at greater distances than encountered for the three species analysed. Nonetheless, the simulation results highlight that meso-avoidance behaviour needs to be assessed using auxiliary data (e.g. from tracking) or by improving the system capabilities to reliably track birds in the immediate vicinity of focal turbines to ensure that the assumptions of the distance sampling model are not violated in a way that compromises estimated bird densities.

5.7 Conclusions

In this report a 3D distance sampling model has been created and shown to work well in simulation scenarios. Given sample sizes above 100-200 bird detections accurate estimates of flight height can be made at the same time as estimates of bird densities within the observed volume, accounting for imperfect detection. These estimates in principle allow the estimation of flux within and outside of the rotor swept zone, which is of interest for collision risk assessments. However, estimates of bird densities below the rotor zone were less robust in this study, as the cameras were focussed on the rotor swept zone and provided less good coverage of the near-surface stratum. This was an a priori design choice, as the original focus of the study was to capture micro-avoidance behaviour in the rotor swept zone. For the same reason, the object tracking algorithm for this study was optimised to track birds in front of the sky only. The addition of tracking capabilities for birds in front of the sea or land surface to the mono-vision system would likely greatly improve flux and flight height estimation in future deployments, by providing much better coverage of the airspace close to the sea surface. Tracking algorithms with these capabilities have been deployed in other Spoor AI installations, but site-specific training data is required for these and collection thereof was outside the scope of the current study.

Estimates of total bird flux were difficult to obtain due to species-specific variation in both flight height and detectability, we therefore found that it is best to analyse species separately. The modelling framework can accommodate variations in both detectability and behaviour due to weather conditions, although sample size must be large enough.

Estimated detection functions varied by species and with camera focal length, and under different environmental conditions. Half-normal detection function standard deviations ranged from c. 250m for Kittiwake at 48mm focal length lens to c. 750m for Gannet at 70mm focal length (nominal distances from mono-vision reconstructions). At a range of 500m this translates to detections of c. 14% of available Kittiwake in the former case, and 80% of available Gannets in the latter.

We also determined the impact of various assumptions made in the modelling framework on the results that are obtained. Specific model features that would benefit from further development are the assumptions that there is no meso-avoidance occurring and that there are no observation errors in the position of points.

Simulation results showed that violations of the assumption that there is no meso-avoidance can result in the inaccuracy of estimates if the detectability distance is within the avoidance range of a species, or if the detection distance is large. For the species and site studied here neither was the case. However, a parametric avoidance function can in principle be added into future iterations of this modelling framework, as long as information about the strength and scale of avoidance are available (e.g. from independent radar or biologging data; Tjørnløv et al., 2023, Pollock et al., 2024). Estimation of these parameters directly from imagery data in the current setup is difficult if not impossible, given the confounding of the detection and avoidance functions at the camera location, and the very small sampled volume in the vicinity of the camera. However, in principle similar monitoring setups with a sufficiently large detection range to reliably monitor avoidance around the focal turbine and/or using multiple cameras could overcome this limitation and allow for the simultaneous estimation of both horizontal and vertical density gradients in the presence of imperfect detection.

Accounting for the observation error that is present in the data being used is more critical. Current detectability, range, and flight height estimates are conditional on the nominal ranges returned from the mono-vision system, which may contain substantial measurement error (cf. section 4-3) and as a consequence we show the model likely overestimates mean flight height and the detection range, and hence underestimates the density of birds within the observed field of view. Additional validation based on a large number of stereo-vision tracks, and/or improved calibration of the mono-vision distance estimation procedures are desirable, as model-based correction of large measurement errors is challenging (Marques, 2004).

Lastly, the current modelling approach removes potential issues with autocorrelated detections by randomly subsampling reconstructed tracks. Methods for the use of continuous video footage in distance sampling models exist (Howe et al., 2017), as do methods that more formally account for movements of birds through the sampled volume (Glennie et al., 2021) and could be integrated with the framework developed here.

5.8 References

Aschwanden J, Stark H & Liechti F. 2024. Flight Behaviour of Red Kites Within Their Breeding Area in Relation to Local Weather Variables: Conclusions with Regard to Wind Turbine Collision Mitigation. Journal of Applied Ecology. https://doi.org/10.1111/1365-2664.14739.

Borchers HW. 2023. Pracma: Practical Numerical Math Functions. https://cran.r-project.org/web/packages/pracma/index.html.

Buckland ST, Oedekoven CS, and Borchers DL. 2016. Model-Based Distance Sampling. Journal of Agricultural, Biological, and Environmental Statistics 21 (1): 58–75. https://doi.org/10.1007/s13253-015-0220-7.

Buckland S, Anderson D, Burnham K, Laake J, Borchers DL,and Thomas L. 2001. Introduction to Distance Sampling: Estimating Abundance of Biological Populations. Vol. xv. https://doi.org/10.1093/oso/9780198506492.001.0001.

Chis OT, Villaverde AF, Banga JR, and Balsa-Canto E. (2016). On the relationship between sloppiness and identifiability. Mathematical Biosciences, 282, 147-161. https://doi.org/10.1016/j.mbs.2016.10.009.

Cole DJ, and Morgan BJ. (2010). Parameter redundancy with covariates. Biometrika, 97(4), 1002-1005. https://doi.org/10.1093/biomet/asq041.

Cox MJ, Borchers DL, Demer DA, Cutter GR, and Brierley AS. 2011. Estimating the Density of Antarctic Krill (*Euphausia superba*) from Multi-Beam Echo-Sounder Observations Using Distance Sampling Methods. Journal of the Royal Statistical Society Series C: Applied Statistics 60 (2): 301–16. https://doi.org/10.1111/j.1467-9876.2010.00748.x.

Glennie R, Buckland ST, Langrock R, Gerrodette T, Ballance LT, Chivers SJ, and Scott MD. 2020. Incorporating animal movement into distance sampling. Journal of the American Statistical Association. DOI: 10.1080/01621459.2020.1764362

Howe EJ, Buckland ST, Després-Einspenner ML, and Kühl HS. 2017. Distance sampling with camera traps. Methods Ecol Evol, 8: 1558-1565. https://doi.org/10.1111/2041-210X.12790

Hüppop O, Dierschke J, Exo KM, Fredrich E, and Hill R. 2006. Bird Migration Studies and Potential Collision Risk with Offshore Wind Turbines. Ibis 148 (s1): 90–109. https://doi.org/10.1111/j.1474-919X.2006.00536.x.

Jongbloed RH. 2016. Flight Height of Seabirds. A Literature Study. IMARES rapport C024/16

Marques TA. 2004. Predicting and Correcting Bias Caused by Measurement Error in Line Transect Sampling Using Multiplicative Error Models. Biometrics 60 (3): 757–763. https://doi.org/10.1111/j.0006-341X.2004.00226.x.

Marques TA, Buckland ST, Borchers DL, Tosh D, and McDonald RA. 2010. Point Transect Sampling Along Linear Features. Biometrics 66 (4): 1247–55. https://doi.org/10.1111/j.1541-0420.2009.01381.x.

Pollock CJ, Johnston DT, Boersch-Supan PH, et al. 2024. Avoidance and Attraction Responses of Kittiwakes to Three Offshore Wind Farms in the North Sea. Marine Biology 171 (217). https://doi.org/10.1007/s00227-024-04542-y.

Tjørnløv RS, Skov H, Armitage M, Barker M, Jørgensen JB, Mortensen LO, Thomas K, and Uhrenholdt T. 2023. Resolving Key Uncertainties of Seabird Flight and Avoidance Behaviours at Offshore Wind Farms. AOWFL.

6 General conclusions and recommendations

Imagery-based three-dimensional (3D) tracking systems based on relatively simple camera hardware show considerable potential for cost-effective, scalable monitoring of avian interactions with wind farms. Current CRMs generally combine flux and flight height information from areas outside of OWFs with avoidance factors that account for behaviours at micro-, meso- and macro-scale, but which are generally associated with large uncertainties. Empirical estimates of flux and flight height distributions within OWFs describe bird behaviour where it is most relevant for collision risk. In-situ 3D tracking therefore has the potential to greatly reduce uncertainties in collision risk estimates from collision risk models (CRMs). However, the effectiveness of such systems depends on addressing both technological and environmental constraints.

All monitoring methods are imperfect, and therefore appropriate quantification of uncertainties and biases are required to create robust evidence and inferences. The technology market for offshore bird monitoring systems is growing, but transparent evaluations and validations of systems are generally lacking. Our study shows that both mono-vision and stereo-vision systems powered by Spoor AI can reconstruct the trajectories of well-defined targets under controlled conditions at ranges of at least 500m.

We further highlight the challenges posed to single-camera (mono-vision) systems by intra-specific body size variation in seabirds and raptors to optimise monitoring strategies and mitigate the ecological impacts of wind farms. Stereo-vision systems are not affected by this constraint and deliver superior accuracy at ranges up to 500m even with short baseline distances and unsynchronized cameras, but the calibration requirements and analytical workflows for stereo systems pose significant challenges in practical deployment. In contrast, mono-vision systems are more scalable but require further refinement to achieve the precision needed for specific tasks, such as the assessment of micro-avoidance behaviours at large ranges. Integrating the strengths of both approaches could pave the way for reliable and scalable bird monitoring solutions in renewable energy contexts. Finally, we present a 3D distance sampling model that can simultaneously estimate bird densities and flight height distributions, providing valuable insights into bird behaviours within offshore wind farms, albeit with some simplifying assumptions.

Improving the usability and performance of both stereo-vision and mono-vision systems should be the focus of future work and in particular we recommend additional work in the following areas:

1. **Stereo-vision capability:**

    1.1. Improve rigidity of camera housings in the stereo configuration and monitor camera drift automatically using the horizon and observed turbine(s) for horizontal and vertical referencing and recalibration.

    1.2. Conduct in-situ calibration using drones or surface vessels as reference objects following installation.

1.3. Optimise baseline distances, increase image resolution and/or increase frame rate to enhance stereo-vision range and accuracy.

1.4. Develop automated methods for stereo tracking to reduce manual processing time and effort.

2. **Mono-vision capability:**

2.1. Improve range calibration, e.g. by using matched stereo-vision trajectories.

2.2. Extend tracking algorithms to capture birds below the horizon, i.e. in front of the sea or land surface, and optimise camera focal length and orientation to ensure the field of view includes the airspace below the rotor swept zone.

2.3. Improve detection of movement artefacts caused by bounding box fluctuations around flapping birds.

3. **Further validation:**

3.1. Extend data processing of collected footage and collect additional footage to capture bird behaviour across a wider range of environmental conditions and bird species.

4. **Application strategy:**

4.1. Use mono-vision systems for long-term monitoring due to their scalability and cost-effectiveness.

4.2. Apply stereo-vision systems for short-term studies and/or in small numbers to collect validation and ground-truthing data.

5. **Data collection:**

5.1. Increase sample sizes and expand camera fields of view to enhance data reliability, especially below the rotor swept zone.

6. **Flux/density models:**

6.1. Address meso-avoidance effects, observational inaccuracies, repeated measures and movement through improved data calibration and modelling.

Images: Font cover Kittiwake, by Edmund Fellowes / BTO. Back cover Herring Gull, by Edmund Fellowes / BTO

Scientific support to the trial of Spoor AI at the European Offshore Wind Deployment Centre

This report assesses the capability of a Spoor AI camera system with both mono-vision (single-camera) and stereo-vision capabilities for bird monitoring deployed at the European Offshore Wind Deployment Centre in Aberdeen Bay using both theoretical considerations and onshore and offshore field trials

Brighton, C.H., Clarke, J.A. & Boersch-Supan, P.H. 2025 Scientific support to the trial of Spoor AI at the European Offshore Wind Deployment Centre. *BTO Research Report* **781**. British Trust for Ornithology, Thetford, UK.

ISBN 978-1-912642-78-6

BTO works to secure the future for birds and nature, using our science, monitoring and data to inform good environmental decisions, inspiring others with the wonder of birds and empowering them to work with us. Our priority is to make a positive impact for birds, using our strong and objective science in partnership with others. We strive to work sustainably and to welcome everyone. www.bto.org

BTO Birds Science People