BTO RESEARCH REPORT 796

Evaluating aerial LiDAR and other approaches to avian flight height measurement – ReSCUE Project Validation Study Report

Rhoades, J., Feather, A., Harwood, A., Banks, A. & Boersch-Supan, P.H.

BTO Birds Science People

ACKNOWLEDGEMENTS: This work was commissioned by Natural England as part of the ReSCUE (Reducing Seabird Collisions Using Evidence) project, with funding from The Crown Estate and Natural England's Species Recovery Programme. The ReSCUE project forms part of the Offshore Wind Evidence and Change (OWEC) programme, led by The Crown Estate in partnership with the Department for Energy Security and Net Zero and Department for Environment, Food & Rural Affairs. The Offshore Wind Evidence and Change programme is an ambitious strategic research and data-led programme. Its aim is to facilitate the sustainable and coordinated expansion of offshore wind to help meet the UK's commitments to low carbon energy transition whilst supporting clean, healthy, productive, and biologically diverse seas. The ReSCUE project is led by Natural England, and delivered as part of the Natural England/BTO Research Partnership providing the evidence needed to support nature's recovery and people's experience of the natural world. We are particularly grateful for the input from Eddie Cole as the project manager. We also thank members of the ReSCUE Project Advisory Group, including the Royal Society for the Protection of Birds, Joint Nature Conservation Committee, The Crown Estate, Defra, Offshore Wind Industry Council and Marine Directorate, and attendees of an Expert Panel workshop in May 2024, who have provided valuable input into the design of the validation study and the resulting report. The contents of this report reflect the author's views, and The Crown Estate is not responsible for any use that may be made of the information it contains.

The validation trial fieldwork was supported by Caroline Brighton, Emma Caulfield, Mark Grantham (BTO); Eddie Cole, Justin Hart, Alasdair Robertson (NE); Bob Ashington, Zoe Hopper, Andrew Marchant, Faye Morgan, Chris Palmer (NE Earth Observation). We thank Andy MacWilliam and Gary Soar (WholeShip Ltd.) for facilitating ground and flight operations at Predannack Airfield and providing liaison with air traffic control and pilots.

Lastly, we thank the aerial survey suppliers, APEM and Bluesky, for their engagement with the trial and delivering the onshore and offshore LiDAR-DAS surveys.

Evaluating aerial LiDAR and other approaches to avian flight height measurement – ReSCUE Project Validation Study Report

Rhoades, J.[1], Feather, A.[1], Harwood, A.[2], Banks, A.[2] & Boersch-Supan, P.H.[1]

BTO Research Report 796

[1.] British Trust for Ornithology

[2.] Natural England

BTO, The Nunnery, Thetford, Norfolk IP24 2PU
Tel: +44 (0)1842 750050 Email: info@bto.org
Registered Charity Number 216652 (England & Wales), SC039193 (Scotland).

ISBN 978-1-912642-95-3

Contents

Executive summary

1. The ReSCUE project investigates the accuracy and reliability of seabird flight height data to mitigate impacts of offshore wind farms on seabird populations.
2. A set of validation trials was conducted onshore at Predannack airfield in Cornwall (19–30 August 2024) and offshore near the Flamborough and Filey Coast (FFC) Special Protection Area (SPA) in September 2024.
3. The focus of the trials was to validate Light Detection and Ranging (LiDAR)-coupled Digital Aerial Surveys (DAS), and to a lesser extent size-based methods. Additional technologies evaluated in the trials included bird-borne Global Positioning System (GPS) telemetry tags and human-operated laser rangefinders (LRFs).
4. Onshore trials used static and suspended target arrays of artificial birds and moving targets (drones) to explore detection rates, flight height accuracy, and flight height precision under controlled conditions.
5. Offshore trials examined the real-world performance of LiDAR-coupled DAS systems under varying flight altitudes.
6. Two aerial survey providers, here referred to as Supplier 1 and Supplier 2, participated in the trials using different instrumentation and data collection approaches. Both suppliers provided LiDAR-based height measurements, and Supplier 1 provided size-based DAS height estimates.
7. We found that the comparative performance of the two LiDAR-coupled DAS suppliers varied based on system differences and weather conditions during surveys, particularly in terms of detection rates from different sensors. LiDAR flight height measurements, however, proved to be accurate and precise for both suppliers, with measurement uncertainties on the scale of centimetres. In contrast, size-based DAS estimates had uncertainties on the order of tens of metres.
8. Inaccurate target size determination in the imagery was the primary driver of DAS height estimate errors. Heterogeneity in target sizes, intended to mimic natural body size variation, was a secondary driver.
9. Results of GPS telemetry and LRF trials provided insights into precision and accuracy and the ability of tracking data to capture dynamic flight patterns, complementing data from aircraft-based technologies.
10. GPS data showed flight height uncertainties on the scale of tens of metres, whereas LRF measurement achieved accuracy and precision on the scale of metres under optimal conditions, but sampling efficiency and reliability was challenging with these devices.
11. Findings from this study will directly inform the development of best practice guidance for seabird flight height data collection and analysis, supporting impact assessments for offshore wind farms while minimising ecological risks to seabird populations.

Glossary of terms

Accuracy: The difference between a measurement or estimate of a value and a recognised standard value (e.g. the true value of a quantity).

Bias: A systematic deviation of measurements and statistics from the truth due to characteristics of the experimental design (e.g. sampling, measurement, analysis). Bias therefore represents the extent to which a sampling, measurement or analytical method inaccurately represents the population, value or statistic thought to be described.

Collision risk model (CRM): The typical model used to predict the number of bird collisions potentially caused by a wind farm development. The current industry standard in the UK is a stochastic version of the Band Model, which requires input parameters describing species-specific information on flight characteristics (including height), biometrics and the expected amount of flight activity.

Detection probability: The probability that a specific sampling method will detect all individuals present.

Digital aerial survey (DAS): Successive photographs or video taken from an aircraft as it surveys a particular location.

Ellipsoid: A geometrically perfect but simplistic model which approximates the shape of the earth, providing a global reference surface where geodetic coordinates can be located on a curved surface.

Eulerian: A survey that samples at predetermined stations or along continuous transects, often replicated through time, to obtain information about animal distribution and abundance in a predefined space and time.

Field of view (FOV): The aperture angle covered by the LiDAR sensor, i.e. the angle over which LiDAR pulses are emitted and detected.

Flight altitude: In the context of this report, flight altitude refers to the altitude of the aircraft conducting the onshore and offshore surveys.

Flight height: The distance between a bird and the ground or sea surface.

Flight height distribution (FHD): The frequency with which birds are observed to occupy airspace along the vertical axis. Depending on the observation technology, flight height distributions can be derived as continuous or discrete (e.g. 1 m bins) distributions.

Geoid: An irregular model of mean sea level around the globe, assuming only the influence of the local gravitational field and the rotation of the Earth (i.e. no effect of landmass, wind or tide). Ordnance Survey Geoid Model (OSGM15) was used for the onshore data. Earth Gravitational Model (EGM2008) was used for the offshore data.

Ground sampling distance (GSD): The distance between two consecutive pixel centres measured on the ground, i.e. the image resolution in units of the imaged surface.

Instantaneous sea level: The actual elevation of the sea surface at a given moment in space and time.

Laser rangefinder (LRF): Rangefinders refer to methods that require observers to visually identify and track individual birds using optical instruments. Flight height is then formally estimated using laser measurements (e.g. elevation angle, distance, bearing/azimuth) and basic mathematical principles (e.g. trigonometry).

Light detection and ranging (LiDAR): A remote sensing method that uses light in the form of a pulsed laser to measure distances between objects and the sensor.

Mean sea level (MSL): The average height of the ocean's surface over a long period of time. In the UK, mean sea level is typically referenced to Ordnance Datum Newlyn (ODN), which is the average sea level recorded at Newlyn, Cornwall, between 1915 and 1921.

Multipath interference: LiDAR hits that, instead of taking a single direct path to and from the target, bounce between targets and atmospheric bands or other surfaces before reaching the receiver. This can cause an unreliable estimation of position or flight height.

Precision: A measure of the likely spread or consistency of repeated measurements or estimates of a quantity.

Swathe width: The horizontal width of the on-the-ground area surveyed by LiDAR, determined by the FOV and aircraft altitude.

List of acronyms

Acronym	Meaning
AGL	Above ground level
AISL	Above instantaneous sea level
AMSL	Above mean sea level
BTO	British Trust for Ornithology
CI	Credible (for Bayesian applications)/confidence (for frequentist applications) interval
CRM	Collision risk model
DAS	Digital aerial survey
FFC	Flamborough and Filey Coast
FOV	Field of view
GPS	Global positioning system
LiDAR	Light detection and ranging
LRF	Laser rangefinder
ODN	Ordnance Datum Newlyn
OLS	Ordinary least squares
PPM	Points per metre
ReSCUE	Reducing Seabird Collisions Using Evidence
RTK	Real-time kinematic
SPA	Special Protection Area
UHF	Ultra high frequency

1. Background

Expansion of the offshore renewable energy industry (e.g. wind farms) is a key element of the UK's strategy for combatting climate change (Kaldellis & Zafirakis 2011). However, offshore wind farms may cause direct adverse impacts to seabird populations, including through collision mortality, temporary or permanent displacement, barrier effects (e.g. to migration and foraging) and indirect effects (e.g. through impacting productivity and prey resources). Seabirds are one of the most threatened vertebrate groups with many species facing potential extinction (Paleczny *et al.* 2015). Yet, seabirds form a key component of marine ecosystems, providing a wide range of ecological functions and ecosystem services (Mosbech *et al.* 2018, Signa *et al.* 2021, Grant 2022). Accurately predicting, mitigating and compensating for the impacts of offshore development on seabird populations is crucial to ensuring growth in renewable energy can continue without causing significant adverse ecological impacts (Fox *et al.* 2006, Shields *et al.* 2009, Rahman *et al.* 2022).

Collision risk models (CRMs) are used to predict the risk of turbine collision when assessing the impacts of offshore wind farms on seabirds (Masden & Cook 2016). Accurately estimating collision risk relies on information on flight height distributions of seabird species (Largey *et al.* 2021, Masden *et al.* 2021). A number of different sampling methods may be used to collect flight height data, each with their own uncertainties and logistical constraints, including visual surveys, Digital Aerial Surveys (DAS), laser rangefinders (LRFs), animal-borne tracking devices, radar, and Light Detection and Ranging (LiDAR) (Thaxter *et al.* 2015, Largey *et al.* 2021). Characteristics of seabird flight, such as height, location, speed and direction, may also vary in response to individual-based traits (e.g. species), behavioural state (e.g. foraging, commuting, resting), environmental conditions and human activities. These elements lead to considerable spatial and temporal variation in seabird flight that must also be captured within the sampling process (Ainley *et al.* 2015, Lane *et al.* 2020, van Erp *et al.* 2023). Consequently, there is substantial ambiguity over the reliability of estimated flight height distributions used within CRMs and the resulting predicted collision rates.

The ReSCUE (Reducing Seabird Collisions Using Evidence) project aims to provide confidence in flight height data, its use in impact assessments, and development of effective mitigation solutions. As part of the ReSCUE project, Feather *et al.* (2025) conducted a review of available methods to collect data on seabird flight heights and identified key knowledge gaps and research priorities. They identified the requirement for a comprehensive understanding of species-specific detection probabilities associated with Eulerian sampling methods (i.e. recording animals as they cross a fixed observation volume at each moment, without tracking individuals). Species-specific detection probabilities are essential for determining the effectively sampled surveyed volume and accurately estimating flight height frequency distributions. However, these probabilities remain largely unquantified for most (if not all) Eulerian sampling methods (e.g. DAS or LRFs), limiting our ability to define the effectively sampled volume and to model flight height distributions with confidence. A large part of the proposed research under the ReSCUE project therefore aims to focus on improving current understanding of the accuracy, precision, and detection probabilities of observations from airborne platforms, particularly those using LiDAR.

This report details the results of onshore and offshore validation studies conducted in August and September 2024. The studies aimed to compare and investigate sources of uncertainty associated with the use of Global Positioning System (GPS) telemetry tags, LRFs, LiDAR-coupled DAS, and size-based DAS methods for measuring or estimating seabird flight heights. The land-based field study was conducted between 19–30 August at Predannack airfield on the Lizard Peninsula, Cornwall. This survey made use of a range of fixed and moving targets to investigate the accuracy, precision, and associated sources of uncertainty for the different technologies, to provide context for the offshore survey. The offshore surveys were conducted in September off the coast of the Flamborough and Filey Coast (FFC) Special Protection Area (SPA) and aimed to further investigate the real-world performance of LiDAR-coupled DAS systems at different altitudes, alongside providing comparative data using the size-based methods that has been adopted by some DAS providers (Humphries *et al.* 2023, Boersch-Supan *et al.* 2024, Forster *et al.* 2024). Two LiDAR-coupled DAS suppliers, here referred to as Supplier 1 and Supplier 2, were commissioned by Natural England to survey both onshore and offshore areas. Results are not directly comparable between the two suppliers due to different systems, altitudes flown, and because the surveys were flown on different days with different environmental conditions. The findings of this validation study will inform further development of best-practice methods for collecting and analysing seabird flight height data for use in future offshore wind farm impact assessments.

2. Methods

2.1. Field trials

The study design of the field trials was based on the outcomes of the Feather *et al*. (2025) review, combined with the outcomes of an Expert Panel Workshop (May 2024). Four separate field trials took place: onshore and offshore LiDAR trials (including DAS), GPS telemetry trials and LRF trials.

2.1.1. Onshore LiDAR-coupled DAS trials

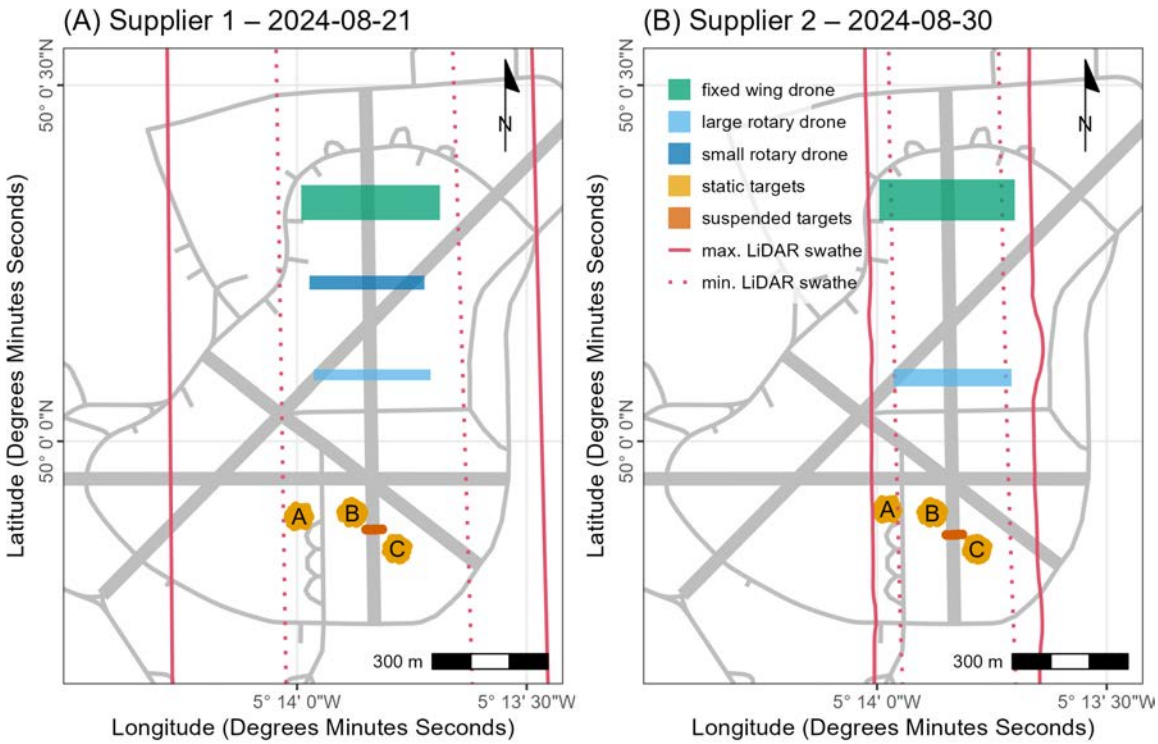
The onshore LiDAR trials used a range of artificial static, suspended, and moving targets with the aim of investigating detection rate and the accuracy, precision, and associated sources of uncertainty for LiDAR and size-based DAS estimates of flight height. The onshore survey also aimed to provide context for the offshore surveys under 'best-case' conditions.

The field trials took place on the 21 August 2024 (Supplier 1) and 30 August 2024 (Supplier 2) at the Predannack airfield on the Lizard Peninsula, Cornwall. For both suppliers, an aircraft equipped with LiDAR and DAS-scanning sensors flew repeatedly along a single transect at multiple predefined altitudes. Static, suspended and moving (drone) targets were detected and matched between LiDAR and DAS detections. Figure 1 shows the layout of the static and suspended targets (seabird-shaped plywood silhouettes and 3D prints), the operation zones of the moving targets (drones), and the LiDAR swathe widths achieved by each supplier.

Both suppliers conducted aircraft flights at four different altitudes within a 300 m range to test detection probability and height accuracy/precision at varying distances between the aircraft and targets. Aircraft flight altitudes differed between the suppliers due to being optimised for different systems, including swathe width. Supplier 1 flew six passes at 400 m altitude above ground level (AGL) and three passes each at 300 m, 500 m and 600 m AGL, noting that 400 m AGL was considered optimal for data collection. A system setting was erroneously configured for the 300 m AGL flight, resulting in reduced LiDAR returns for this flight, which was corrected for the offshore survey. Supplier 2 completed seven flight passes at 625 m AGL altitude and three passes each at 725 m, 825 m and 925 m AGL, noting that 625 m was considered optimum for data collection as the quality of LiDAR measurements and images would have been high without sacrificing data coverage.

Results are not directly comparable between the two suppliers due to different systems and configurations: the system used by Supplier 1 optimised coverage (a wider swathe width with less dense points), whereas the system used by Supplier 2 prioritised point density (a narrower swathe width with denser points). In addition, for Supplier 2, the same area was scanned by LiDAR twice: once each in the forward and reverse directions. The estimated height of targets was averaged between the forward and reverse scans. The temporal offset between the scans was dependent on aircraft speed and swathe size, but was typically around four seconds. The fact that the same area was scanned twice by Supplier 2 may have resulted in greater sampling effort compared to Supplier 1. This was especially pertinent for the moving targets, for which the same drone produced two separate groups of LiDAR returns in different locations which aided identification.

Figure 1: Minimum and maximum LiDAR swathe coverage for each supplier relative to target zones at Predannack airfield during the LiDAR-coupled DAS surveys in August 2024. Static target arrays were designated A–C from west to east.



Results are also not directly comparable between the suppliers due to differing environmental conditions during the surveys. The survey for Supplier 1 was conducted in suboptimal weather conditions: recorded ground wind speed ranged between 14–22 knots during the survey, with wind speeds in excess of 20 knots potentially adversely affecting data quality. High winds may have impacted results as some targets were damaged or blown over (although damaged/fallen targets have been removed from analysis). Wind gusts (which may impact data quality and aircraft operation) were also considered a concern during the survey, although wind gust data was not recorded. Grass length at the survey site also differed between the Supplier 1 and 2 surveys: the grass was approximately 30–50 cm in height during the Supplier 1 survey, but was mown prior to the Supplier 2 survey. Grass length may potentially have impacted detectability of targets and therefore further limits comparability between the results of the two suppliers.

Following the field trials, both suppliers analysed the aerial images to identify the detected targets and classify them by species and colour (for the static and suspended targets; see details of target design below) or drone type (for the moving targets). For Supplier 1, target height was measured directly using LiDAR and also estimated from the aerial imagery using a size-based DAS methodology. The size-based DAS methodology employed in this study is a customised photogrammetric method that uses trigonometric calculations developed by Supplier 1 based on species-specific reference lengths, image ground sampling distance (GSD; the on-the-ground distance between pixel centres), the known height of the aircraft as the image was taken, and the pitch, roll, and yaw of the aircraft. Reference lengths for the artificial targets were provided to Supplier 1 by Natural England. For Supplier 2, target height was measured directly using LiDAR only.

During LiDAR surveys, multipath inference may occur when, instead of taking a single direct path to and from the target, LiDAR hits bounce between targets and atmospheric bands or other surfaces before reaching the receiver. Multipath inference causes an unreliable estimation of position or flight height, and is likely to be more common in offshore environments or in the presence of atmospheric particles. Both suppliers had measures in place to mitigate against multipath effects, including in-flight and post-flight processing. Further, Supplier 1 assigned a low confidence level to measured bird heights where there was noise in the vicinity and only a single LiDAR noise was received, and Supplier 2 flagged potential multipath hits.

Static and suspended targets

One hundred and thirty-one static targets resembling three seabird species of varying sizes were constructed: 30 Northern Gannet (*Morus bassanus*; hereafter 'Gannet'), 35 Black-legged Kittiwake (*Rissa tridactyla*; hereafter 'Kittiwake') and 66 European Storm Petrel (*Hydrobates pelagicus*; hereafter 'Storm Petrel'). The majority of targets were constructed out of plywood, but a small number of Storm Petrel targets were 3D printed. Both dark and light-coloured targets were made for Kittiwake and Storm Petrel, whereas only light Gannet targets were made. The targets were constructed to reflect the typical size of the species while accounting for species-specific natural variation, i.e. mean wing length ± standard deviation as determined from data from the British and Irish ringing scheme, existing literature and museum specimens (Boersch-Supan *et al.* 2024, Brighton *et al.* 2025). Targets were also constructed with varying wing configurations (angled or straight). All possible combinations of target species, colour and wing configurations were made, with the exception of dark-coloured Gannet targets.

Of the 131 targets, 120 were positioned in three circular plots, designated arrays A, B, and C from west to east. Some were affixed to fence posts approximately 1 m above the ground and others were positioned on tripods at heights varying between approximately 2–7 m. The remaining 11 targets were positioned in a linear array (here referred to as 'washing line'), with nine suspended from two strands of braided fishing line at heights of approximately 3–3.7 m and two affixed to winch stands at either end of the array at a height of approximately 5.2 m. The LiDAR and size-based DAS estimated heights were then compared to the known heights of the targets to determine vertical error during analysis. The actual heights of the nine targets suspended from the washing line was uncertain due to swaying in the wind during the field trial. Examples of the plywood static targets and field trial configuration are shown in Figures 2 to 5.

Figure 2: Examples of static plywood targets used in the onshore LiDAR trials. Top: Gannet (light, angled wing); Middle: Storm Petrel (dark, angled wing), Storm Petrel (light, straight wing); Bottom: Kittiwake (light, angled wing); Kittiwake (dark, straight wing). All combinations of target species, colour and wing configuration were tested, except dark-coloured Gannet targets. Image credit: Philipp Boersch-Supan/BTO

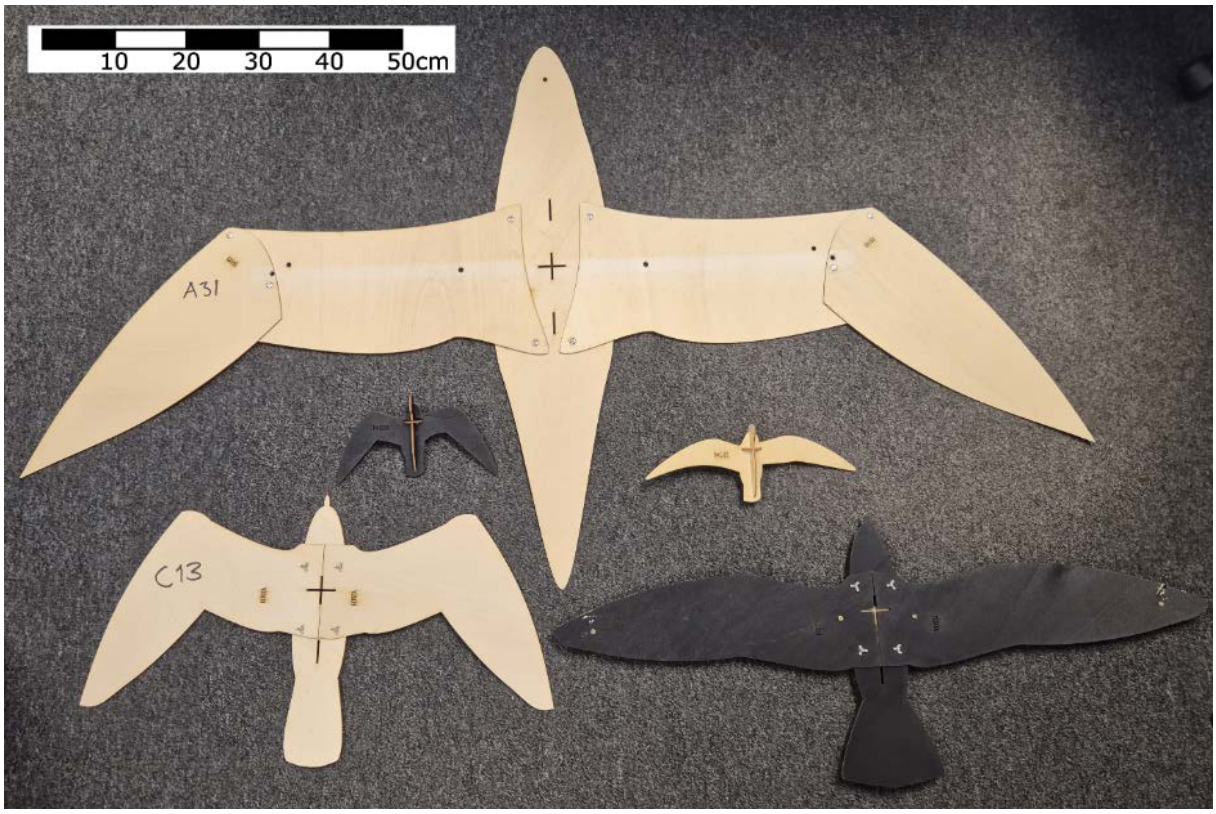


Figure 3: Set-up of static target array C. Image credit: Philipp Boersch-Supan/BTO

Figure 4: Low-altitude aerial view of a static Gannet target. Image credit: NE Earth Observation

Figure 5: Survey altitude aerial view of static arrays B (top left) and C (bottom right), and the suspended targets on the 'washing line' (centre). Image credit: Supplier 2.

Moving targets

Airborne drones operated by the Natural England Earth Observation Team were used as moving targets for the onshore field surveys (Figure 1). Three drones were flown during the Supplier 1 survey: a fixed wing-Sensefly eBee X (wingspan 116 cm), large rotary DJI Matrice M300 (length 81 cm; width 67 cm) and a small rotary DJI Mavic 3E (length 35 cm; width 28 cm). All drones were equipped with high-precision real-time kinematic (RTK) GPS units. The small rotary drone (DJI Mavic 3) was unavailable for the Supplier 2 survey, and only the eBee and M300 drones were flown.

Drones were programmed to fly repeated transects at fixed heights (typically 60 m, 80 m and 100 m AGL) perpendicular to the runway orientation (Figure 1). For the rotary drones, realised transect heights were expected to be within centimetres of nominal transects heights, and mean standard deviation of drone-recorded altitude was 0.03 m for both rotary drones. The flight dynamics of the fixed-wing drone meant that realised transects undulated by c. ±10 m around nominal transect heights, particularly when the drone was turning between transects or transitioning between height levels. Flight speeds were not provided, but calculated ground speeds ranged from 1.80–18.8 m/s, 0.56–3.90 m/s and 0.71–2.23 m/s for the eBee, M300 and Mavic drones respectively.

For the M300 and Mavic flights, the Natural England Earth Observation Team provided the precise timestamps, RTK status, longitude, latitude and flight altitude of drone-captured images. The temporal resolution of these images varied across individual flights but typically ranged between approximately 2–9.5 seconds. For the eBee flights, we reconstructed drone tracks from metadata fields in georeferenced imagery provided by the Natural England Earth Observation Team to produce known locations and flight altitudes of the drones. The temporal resolution of the reconstructed eBee drone tracks typically ranged from approximately 1–2 seconds. The LiDAR and size-based DAS estimated heights were then compared to the known flight altitude of the drones to determine vertical accuracy.

2.1.2. Offshore LiDAR-coupled DAS trials

The offshore LiDAR trials aimed to provide context for the possible sources of uncertainty for detection probability and flight height estimation for LiDAR and DAS under 'real-world' conditions at sea.

The Supplier 1 offshore survey was located 10–24 km off the coast at the FFC SPA on 14 September 2024. The same aircraft used in the onshore trial was flown across five transects of 20 km in length parallel to the coast. Each transect was surveyed at altitudes of 300 m, 400 m, 500 m and 600 m above mean sea level (AMSL). The 400 m AMSL altitude was flown twice for each transect and the remaining altitudes flown once, noting again that the 400 m AMSL altitude was considered optimal given the configuration of the LiDAR-coupled DAS system. Species identification and behaviour (e.g. flying or sitting on the surface of the water) of detected birds, alongside LiDAR-derived measurements and size-based DAS estimates of flight height were provided. For the size-based DAS estimates of flight height, the same photogrammetric methodology used in the onshore survey was applied (see Section 2.1.1), with species-specific reference lengths derived by Supplier 1 from the literature. Supplier 1 also provided information for other detected marine species (not included in this analysis).

The Supplier 2 offshore survey was located 8–20 km off the coast of Flamborough Head, Yorkshire, on 28 September 2024. The same aircraft used in the onshore trial was flown across five transects of 20 km in length, once each at altitudes of 731 m, 762 m and 792 m AMSL. The number of repeated transects flown, and variation in flight altitudes, differed from those of the onshore survey due to aircraft endurance limitations and a lower cloud base during the offshore survey, which prevented flight at higher altitudes. Given these limitations, the supplier made the decision to fly three altitudes that were close to the optimum altitude ASL for the onshore survey (762 m, equating to 625 m AGL in the onshore survey). Species identification and behaviour of detected birds, alongside LiDAR-estimated flight heights were provided. Like the onshore survey (see Section 2.1.1), the same area was scanned twice by Supplier 2, once each in the forward and reverse directions. This enabled estimates of flight speed where two separate LiDAR returns were present for an individual bird.

For the offshore survey, the same mitigation measures for multipath inference were followed as described for the onshore survey (Section 2.1.1).

2.1.3. GPS telemetry trials

The aim of the GPS telemetry trials was to evaluate the accuracy, precision and associated sources of uncertainty of estimated altitude recorded by GPS telemetry tags. The telemetry trials were conducted concurrent with the onshore LiDAR-coupled DAS trials. GPS tags (Nanofix GPS + Ultra High Frequency (UHF) + Solar, Pathtrack Ltd., Otley, UK) were either affixed to the drones used as moving targets for the LiDAR trial or were positioned in static locations. Tags were programmed to record readings with a logging interval between 30 seconds and five minutes, downloading data periodically via UHF to a base station on the airfield.

2.1.4. LRF trials

LRF trials were undertaken to collect flight height readings and ranging information across the two onshore trials using a Vectronics Terrapin X (Safran Vectronix AG, Heerbrugg, Switzerland) and a Nikon Forestry Pro II (Nikon Europe BV, Amstelveen, The Netherlands). The Terrapin X provided three-dimensional information in the form of range, elevation angle and azimuth angle (magnetic compass bearing), and the Forestry Pro II provided two-dimensional information in the form of range, elevation angle and height relative to the device. Both devices are designed for primary use cases other than tracking bird-like mobile targets; the Vectronix device is primarily marketed for ballistic applications, the Nikon device for forestry and land survey

applications, and the primary interfaces in both cases are displays on the instrument, with limited support for automatic and comprehensive logging of data. Time stamps (to the nearest second) of measurements were estimated from photo records of the device display for the Forestry Pro II and from screenshots on the companion app for the Terrapin X.

On 21 August 2024 a LRF trial was conducted during the Supplier 1 onshore survey, where fieldworkers attempted to track each of the three moving targets for bouts of 10 minutes, recording ranges, flight heights, and azimuth angles where possible. Only positive hits were recorded – misses were not.

On 29 August 2024 a user test was conducted to assess the handling time of the LRF devices. Five fieldworkers attempted to target five model Kittiwake of different size and height at distances from 25 m to 200 m, recording time to achieve a fix and number of attempts to get a fix, with the maximum number of attempts limited to 10. Furthermore, an initial assessment of azimuth accuracy for the Terrapin X was conducted, with a single user recording azimuth angles to the static targets at ranges 25 m, 75 m, and 200 m, and a hangar (range c. 400 m). Target and operator positions were referenced using a handheld RTK unit (Leica GS16 GNSS Smart Rover, Leica Geosystems Ltd., Milton Keynes, UK).

On 30 August 2024 an additional assessment of azimuth accuracy was conducted with a fieldworker standing at the centre of static target array B, recording the positions of the static targets (ranges 2–30 m), with a limit of five attempts to obtain a fix. In addition, a LRF trial was conducted during the Supplier 2 survey, tracking positions of the DJI Matrice M300 (large rotary) drone with multiple fieldworkers using both devices from one location. Effort was not consistent across users and time and only positive hits were recorded – misses were not.

For all days, user eye height was recorded in all cases and used to correct derived target height estimates. There was no correction based on variations in terrain across the study area.

2.2. Data analysis

2.2.1. Onshore LiDAR-coupled DAS trials

Static and suspended targets
LiDAR hits were matched to the static and suspended targets during data processing. In many cases, the coordinates of the LiDAR hit did not exactly match the recorded coordinates of the targets, so LiDAR hits were presumed to relate to the closest target. For the Supplier 1 survey, a small amount of LiDAR hits was manually matched to a different neighbouring target based on visual inspection of the hit-cluster patterns. In additional cases for a small number of hits, it was difficult to identify which target was detected so those returns were excluded from subsequent analyses. Targets that were damaged or had fallen were also removed from analysis. Removing these ambiguous hits (i.e. where the detected target could not be identified) and fallen targets lowered the predicted detection probability and hit count for the affected targets; however, the proportion of excluded data (2.06%) was small relative to the total sample, so any bias introduced was unlikely to alter the overall model conclusions.

For two targets (both dark Kittiwake), some LiDAR hits detected the tripod rather than the target itself during the Supplier 1 survey. As the tripod hits could not be distinguished from the target hits, all hits for these targets were removed from the analysis. A subsequent analysis was conducted to determine the impact of these targets being identified as 'detected' or 'not detected' on the overall results of detection rate for dark Kittiwake, the results of which are shown in Appendix 1.

Statistical analysis was conducted in R v.4.4.1 (R Core Team 2024) using the 'brms' package (Bürkner 2017). We modelled the estimated proportion of targets detected as a function of target species, target colour and flight altitude using a binomial model with a logit link function (which is appropriate for proportion data). We applied a Gaussian prior (μ=0, σ=2.5) to all predictor coefficients to allow each predictor to potentially have a positive or negative effect on proportion of targets detected. Four separate models were run for each individual dataset: one each for the LiDAR and DAS imagery detections for the Supplier 1 and Supplier 2 surveys.

We modelled the number of LiDAR hits per target as a function of target species, target colour and flight altitude, using a hurdle model (to simultaneously model target detection and the number of LiDAR hits

conditional on detection) with a Poisson distribution and a random effect for target ID. We applied a Gaussian prior (μ=0, σ=2.5) to all coefficients in both the hurdle and count elements of the model. We ran two separate models for the Supplier 1 and Supplier 2 surveys.

To model the accuracy and precision of target height estimation, we used a Gaussian distribution where both μ (accuracy) and σ (precision) were modelled as a function of target species, target colour, flight altitude and target type (static or suspended), alongside a random effect for target ID. Target type was included as a predictor because the actual height of suspended targets was uncertain due to swaying in the wind, which may have impacted the accuracy/precision of target height estimation. Target type was not considered likely to influence detection probability or the number of LiDAR hits, and so was not included as a predictor for those models. We ran three separate models for Supplier 1 LiDAR heights, Supplier 1 DAS heights and Supplier 2 LiDAR heights. For the Supplier 1 LiDAR model, we also included a factor effect to indicate the influence of the number of LiDAR hits on target height estimation (single or multiple hits). We did not include this effect in the Supplier 2 model due to a strong association between number of hits and target species. A Gaussian prior (μ=0, σ=5) was applied to all predictors for both the accuracy and precision components.

We attempted to decompose the Supplier 1 DAS size-based height estimation errors into errors arising from within-species target size variation and photogrammetric measurement error, based on a simplified theoretical photogrammetry error model presented in Boersch-Supan *et al.* (2024). Error decomposition was based on DAS heights and target size estimates provided by Supplier 1 and true target sizes and heights (which were not known to Supplier 1).

For all models, we included interactions between target species and both flight altitude and target colour. We ran four sampler chains for each model. The sampler was either run for 2,000 or 4,000 iterations depending on the model complexity, with the first 1,000 iterations of each model discarded as warm-up. Model convergence was evaluated by assessing trace plots, Rhat values and effective sample sizes.

Moving targets

To assess the vertical accuracy of the flight height estimates for the moving targets, the drone detections for Suppliers 1 and 2 were linked to the closest recorded fix for the drone tracks. Where the temporal resolution of the drone tracks was coarser than one per second, we interpolated the drone tracks to give estimated latitude, longitude and flight altitude at a resolution of one per second. Turning circles for the eBee drone were excluded from the interpolation procedure due to being large, leading to uncertainty of the actual location of the drone (turning circles for the remaining drones were minimal and were retained in the interpolation). The drone detection was then linked to the drone fix with the nearest timestamp to assess vertical accuracy. Of a total of 46 individual drone detections across both suppliers, five detections were made in periods where the actual location of the drone was unknown (i.e. made in turning circles or periods of time outside of the provided drone tracks). These five detections were removed from the analysis.

Drone altitude was provided as height above the ellipsoid. For comparability with the LiDAR measurements and size-based DAS flight height estimates, the drone heights were converted to height AMSL by subtracting geoid height from the ellipsoid height, as done in Lato *et al.* (2022). We used the OSGM15 National Geoid Model to extract height above the geoid, referenced to Ordnance Datum Newlyn (ODN). The deviation between the known heights of the drones and the flight height measured by LiDAR or estimated by the size-based DAS survey was then summarised.

2.2.2. Offshore LiDAR-coupled DAS trials

As the true number of birds and their associated flight heights were unknown, detection probability and vertical accuracy could not be assessed for the offshore surveys. The observed number and height of birds of different species was summarised for each supplier. For Supplier 1, the deviation between flight height measured by LiDAR and estimated by the size-based DAS method for individual birds was also summarised.

2.2.3. GPS telemetry trials

Vertical accuracy of the GPS tags was estimated by comparing the altitude recorded by the GPS tag with the known altitudes of the drone or the stationary location. For the tags on drones, each GPS tag fix was matched to the drone fix with the closest time stamp. As described in Section 2.2.1, individual drone flight tracks were

interpolated to provide an estimated latitude, longitude and flight altitude per second. Of 6,470 GPS fixes with matched known drone locations, 24 had anomalously high vertical deviations between GPS-recorded altitude and known drone altitude (absolute values > 500 m). Twenty-one of those were consecutive fixes from the same stationary tag, presumably arising from inaccurate readings (possibly due to being placed in the corner of hangars during the field trial). These 21 observations were removed from the analysis.

To model the potential impact of covariates on the accuracy and precision of GPS tag height estimation, we used a Gaussian distribution where both μ (accuracy) and σ (precision) were modelled as a function of the number of visible satellites, battery level of the tag, accuracy indicator (an indicator of accuracy specific to the type of GPS tag used), known altitude, and fix type (stationary or drone), plus a random effect for tag number for μ only. Data analysis was conducted in R v.4.4.1 (R Core Team 2024) using the 'brms' package (Bürkner 2017).

2.2.4. LRF trials

Ranging information from the devices was transcribed from photographs, screenshots and other field records, and where necessary heights and distances were calculated from the relevant distances and inclination angles. Measured heights were adjusted for operator eye height. For the Terrapin X, there was a time lag of approximately four seconds between acquiring the fix and obtaining a screenshot. The timestamps were adjusted accordingly, but mismatches between timestamps may have introduced uncertainty into the results.

Measurements of mobile targets were referenced to drone positions which were reconstructed from RTK georeferenced drone imagery collected at variable intervals (five seconds to five minutes). RTK positions were retained for analysis only while the drones were in level flight at the predetermined experimental heights (60 m, 80 m, 100 m above take-off elevation). Drone elevations were adjusted to the ground level of the LRF observer as determined by a handheld RTK unit. Statistical analysis was conducted in R v.4.4.1 (R Core Team 2024) using the 'brms' package (Bürkner 2017). We modelled flight height deviations as a function of target type (i.e. drone model) and LRF device type using a robust regression model with Student's t-errors. We applied default priors to coefficients.

Azimuth (compass bearing) measurements were referenced against bearings calculated from RTK measurements of the relevant static target. Angular quantities were summarised using circular means, standard deviations, and standard errors (Fisher 1993) as implemented in the 'circular' and 'tectonicr' packages (Stephan *et al.* 2023, Agostinelli & Lund 2024). There were no metal structures at the study site with the potential to influence the performance of the integrated compass, and as such, the results may represent a best-case scenario (as opposed to LRF studies conducted on boats or offshore wind farms).

The handling time user test was analysed using time-to-event models, to simultaneously assess overall success rate and time to first fix as a function of device used and distance to target. Models were estimated using a Cox proportional hazards framework (Therneau & Grambsch 2000) as implemented in the 'survival' R package (Therneau 2024). Separate models were fitted using number of attempts to first fix and time to first fix, respectively, as temporal measures.

3. Results

3.1. Onshore LiDAR-coupled DAS trials

3.1.1. Static targets

Target detection

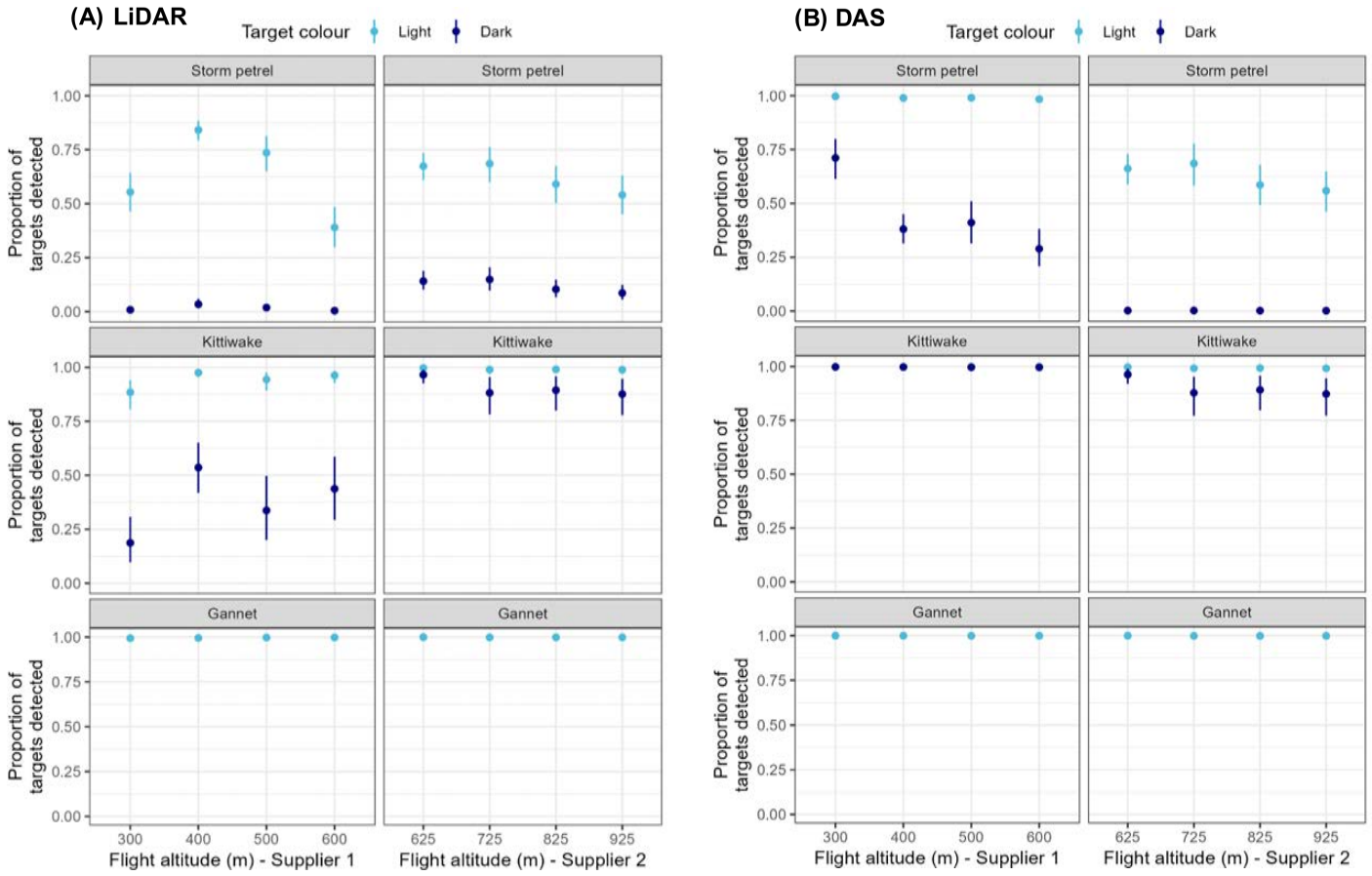
General trends across the onshore LiDAR-coupled DAS survey for both providers showed lower detection rates for smaller and dark targets. The conditional influence of flight altitude was minimal for targets with high detection rates (i.e. larger/light targets), but a slight reduction in detection rate with increasing flight altitude was observed for targets with lower detection (i.e. Storm Petrel targets).

For the LiDAR data (Figure 6), predicted detection of Gannet targets was almost perfect for both providers. For Kittiwake, predicted detection of light targets was very high for both providers, with reduced detection of dark targets of 89–97% for Supplier 2 and 36–55% for Supplier 1 (discounting the 300 m AGL flight altitude

which operated at suboptimal instrument settings). Both providers also detected substantially fewer dark Storm Petrel targets (c. 10% for Supplier 2, < 5% for Supplier 1) than light ones (30–80% across providers).

For the DAS imagery (Figure 6), detection rates were generally higher for Supplier 1 than Supplier 2. Predicted detection for Supplier 1 was almost perfect for all targets except dark Storm Petrels (30–75% detected). For Supplier 2, light Kittiwake and Storm Petrel targets were detected more frequently (> 95% and c. 60%, respectively) than dark targets of the same species, with c. 90% of dark Kittiwake and no dark Storm Petrels detected in the imagery alone.

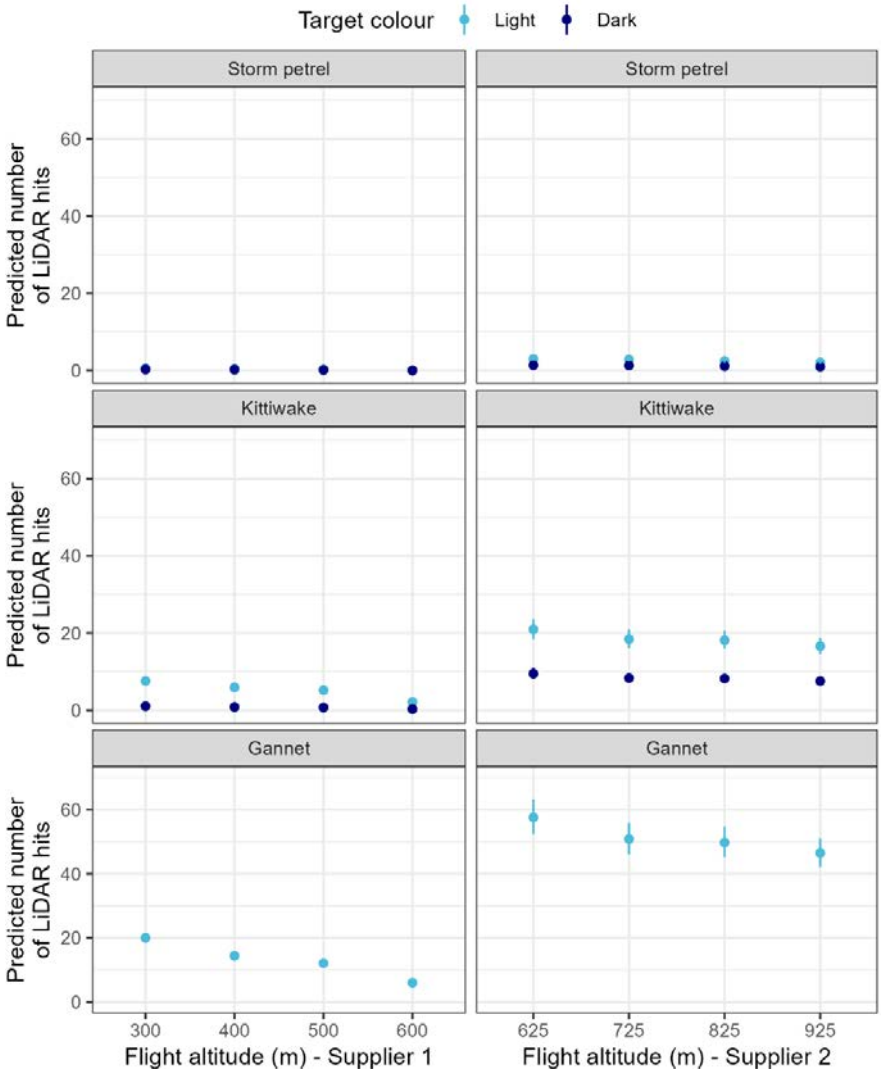
Figure 6: Conditional estimates (including 95% credible intervals) of proportion of targets detected for (A) LiDAR and (B) DAS surveys conducted by Supplier 1 and Supplier 2, according to target species, target colour and flight altitude (m AGL).



Number of LiDAR hits

Generally, Supplier 2's sensors delivered about twice the amount of LiDAR hits per target as Supplier 1's (Figure 7). Larger and light targets had more hits than smaller and dark targets for both suppliers, e.g. Supplier 2 achieved up to 60 LiDAR hits on a Gannet silhouette but fewer than three hits on a light Storm Petrel target. A slight reduction in the number of hits was observed with increasing flight altitude for both suppliers, as would be expected from the geometry of the point spacing. The predicted number of LiDAR hits for dark Storm Petrel targets was close to zero for both Supplier 1 and Supplier 2, reflecting their low detection rate in the LiDAR surveys. The predicted number of hits was also close to zero for dark Kittiwake targets for Supplier 1.

Figure 7: Conditional estimates (including 95% credible intervals) of number of LiDAR targets hits for Supplier 1 and Supplier 2 according to target species, target colour and flight altitude (m AGL).

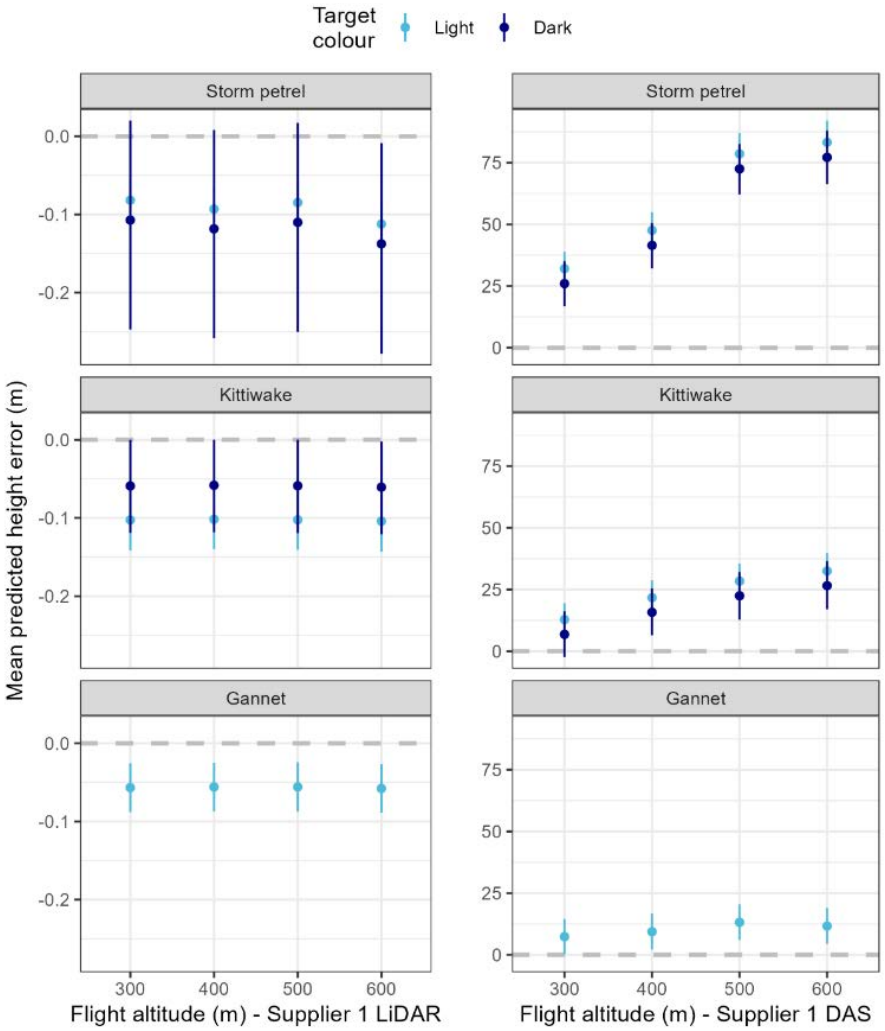


Height estimation

Results for accuracy and precision of target height estimation are shown for static targets only. Suspended targets show similar trends but with wider confidence intervals as actual height may have differed from recorded height due to targets on washing lines swaying in the wind (see Appendix 2).

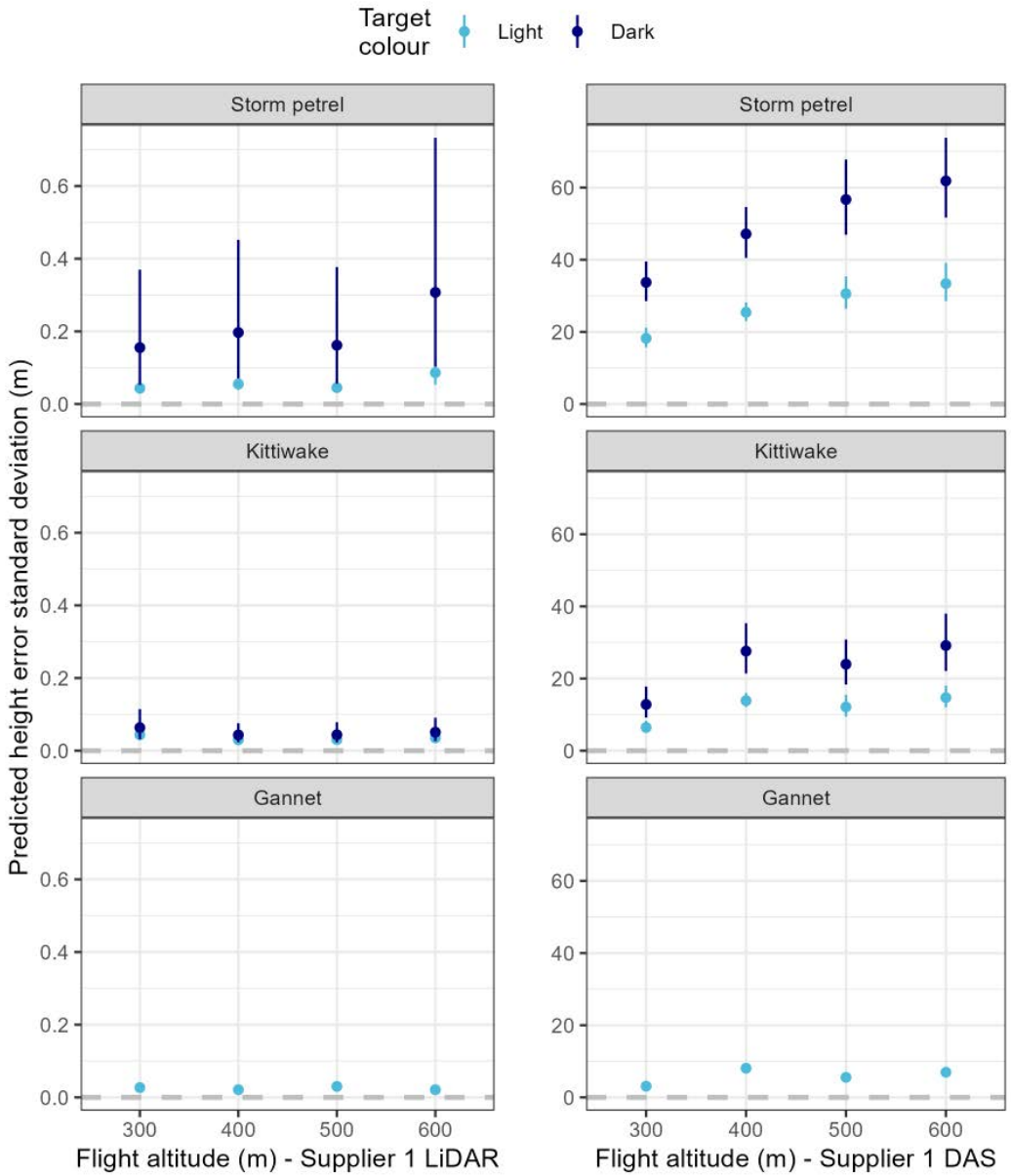
Comparing the accuracy of estimated target height between the Supplier 1 LiDAR survey and size-based DAS (Figure 8), absolute error differed in the range of approximately two to three orders of magnitude between the surveys. Absolute mean error ranged from approximately 0.05–0.1 m for the LiDAR data and approximately 7–80 m for the size-based DAS method. There was little effect of any covariate on accuracy of height estimation for the LiDAR survey. For the size-based method, height estimation was less accurate with increasing flight altitude and for smaller targets.

Figure 8: Conditional estimates (including 95% credible intervals) of the predicted mean discrepancy in estimated and measured height for Supplier 1 LiDAR-derived and size-based flight height methods, according to target species, colour, and flight altitude (m AGL). Results are shown for targets that were not suspended and with multiple LiDAR hits (see Appendix 2 for results of targets with single hits or suspended targets). Note differing y-axis scales for LiDAR and DAS results.



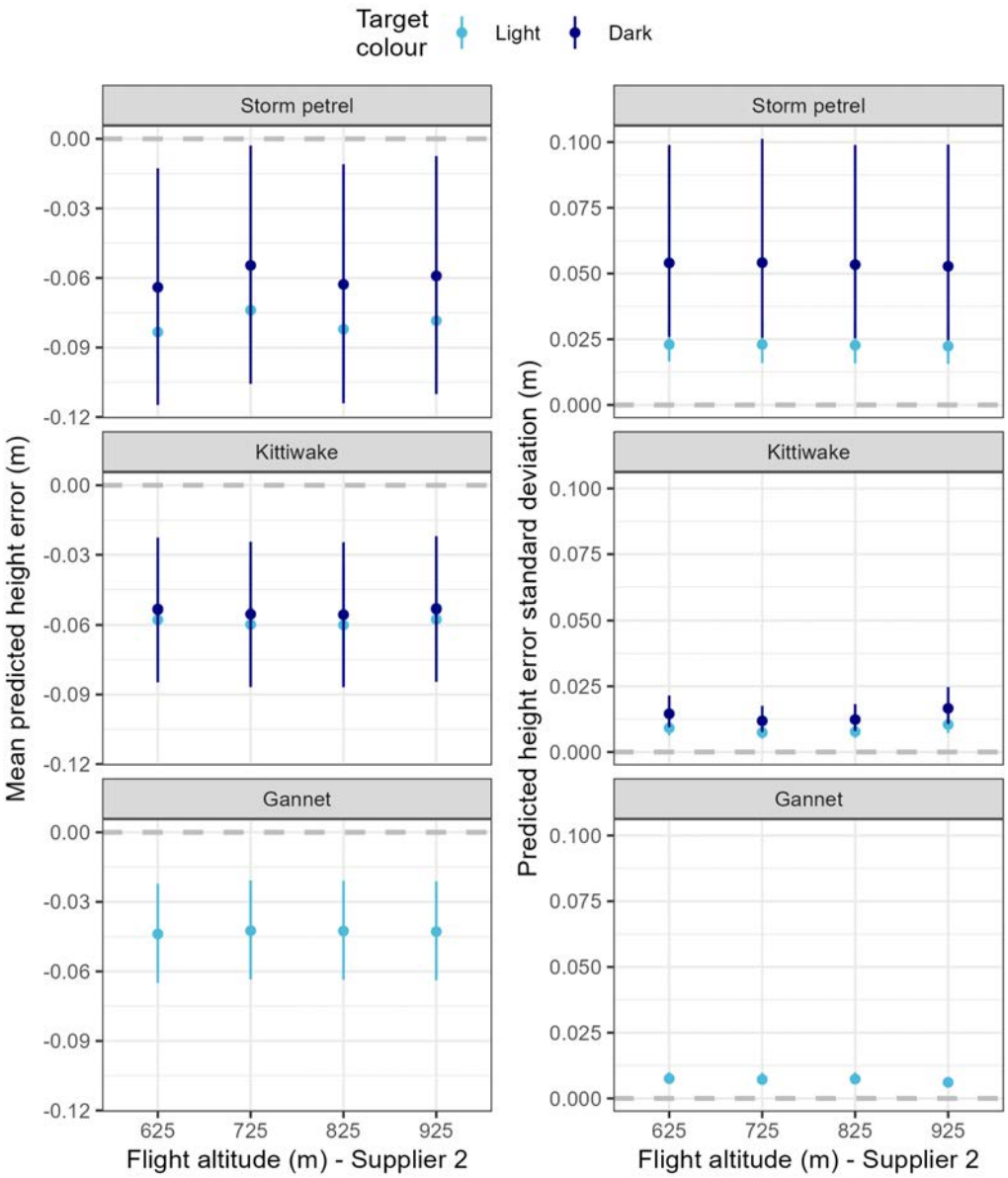
Precision of estimated target height also differed in the range of approximately two to three orders of magnitude between the Supplier 1 LiDAR data and size-based DAS flight heights (Figure 9). Precision of target height estimation was relatively consistent across target types in the LiDAR data, with only dark Storm Petrel targets having a predicted higher standard deviation. For the size-based DAS flight heights, reduced precision (i.e. higher standard deviation of height estimates) was observed for smaller and dark targets. Precision also reduced with increasing flight altitude for both dark and light Storm Petrel targets.

Figure 9: Conditional estimates (including 95% credible intervals) of the predicted precision of estimated target height for Supplier 1 LiDAR-derived and size-based DAS flight height methods, according to target species, colour, and flight altitude (m AGL). Note differing y-axis scale. Results are shown for targets that were not suspended and with multiple LiDAR hits (see Appendix 2 for results of targets with single hits or suspended targets).



For the Supplier 2 LiDAR data (Figure 10), mean error was relatively consistent across target types. Mean predicted estimates of discrepancy in measured and LiDAR-derived heights show that smaller and dark targets had larger absolute errors, although credible intervals overlapped for different target species and colours, showing these differences were not significant. Precision of height estimates were lower (i.e. higher standard deviation) for smaller and dark targets. Flight altitude had little influence on either mean or precision of Supplier 2 height estimates. Size-based height was not estimated for the Supplier 2 DAS survey.

Figure 10: Conditional estimates (including 95% credible intervals) of the predicted mean discrepancy and precision of estimated target height for the Supplier 2 LiDAR data, according to target species, colour, and flight altitude (m AGL). Results are shown for targets that were not suspended (see Appendix 2 for results of suspended targets).



3.1.2. Decomposition of DAS height errors

Photogrammetric measurement error, i.e. the error in determining the apparent size of a target due to image imperfections such as finite image resolution ('pixelation') or chromatic aberrations, is an inherent source of uncertainty in size-based flight height determination (Scherz 1974, Humphries *et al.* 2023). Based on theoretical arguments, biometric measurements, and numerical simulations, Boersch-Supan *et al.* (2024) predicted that intraspecific body size variation in medium and large seabirds provides an additional fundamental limitation to the accuracy and precision of DAS size-based estimates of flight height. This validation study incorporated within-species target size variation to empirically test these predictions.

Supplier 1 provided target size estimates and size-based flight height estimates for the majority of detected targets (N=1,482). Supplier 2 did not provide size-based DAS height estimates, but provided DAS target length estimates for a subset of detections (i.e. one measurement per detected target across all passes, N=86 in total). For both suppliers the target length estimates clustered around true target lengths (Figure 11, Figure 12). Target measurements were on average biased positive by c. 1 cm for Supplier 1 (Table 1) and negative by

c. 0.5 cm by Supplier 2 (Table 2). Mean absolute errors for target size estimates were similar for both providers at c. 2 cm. The image resolution (i.e. ground sampling distance) was estimated between 1.05 cm and 2.08 cm by Supplier 1, and 2.00 cm and 3.00 cm by Supplier 2. The target measurement errors therefore equated on average to 1.20–1.93 pixels for Supplier 1 and 0.73–1.03 pixels for Supplier 2 (Table 1, Table 2).

Table 1: Photogrammetric errors for target measurements provided by supplier 1 (N=1,482)

Flight altitude (m AGL)	Mean (95% CI) photogrammetric error (cm)	Mean absolute photogrammetric error (cm)	Mean (95% CI) photogrammetric error (pixel)	Mean absolute photogrammetric error (pixel)
300	0.71 (-5.62, 5.46)	2.03	0.68 (-5.35, 5.2)	1.93
400	1.1 (-5.81, 5.82)	2.35	0.79 (-4.15, 4.15)	1.68
500	1.78 (-4.07, 5.94)	2.56	1.02 (-2.34, 3.42)	1.47
600	1.39 (-6.05, 6.7)	2.50	0.67 (-2.91, 3.22)	1.20

Table 2: Photogrammetric errors for target measurements provided by supplier 2 (N=86)

Flight altitude (m AGL)	Mean (95% CI) photogrammetric error (cm)	Mean absolute photogrammetric error (cm)	Mean (95% CI) photogrammetric error (pixel)	Mean absolute photogrammetric error (pixel)
625	-0.45 (-3.93, 6.58)	2.06	-0.23 (-1.97, 3.29)	1.03
725	-0.53 (-7.41, 9.91)	2.33	-0.21 (-2.97, 3.96)	0.93
825	-0.45 (-7.19, 8.35)	2.22	-0.16 (-2.61, 3.03)	0.81
925	-0.66 (-7.18, 8.44)	2.18	-0.22 (-2.39, 2.81)	0.73

The data collected on DAS height errors and photogrammetric errors followed the predictions of Boersch-Supan *et al.* (2024), with size-based DAS height errors increasing for targets deviating from the mean target size and with greater photogrammetric measurement error (Figure 13).

Variation in DAS height errors was well explained by a regression against photogrammetric measurement error and the heterogeneity of the target sizes (the exact nature of which was not known to the survey providers; Ordinary Least Squares regression, R^2Gannet=0.92, R^2Kittiwake=0.93, R^2Storm Petrel=0.85). Inaccurate target size determination in the imagery was the primary driver of DAS height estimate errors. Heterogeneity in target sizes, intended to mimic natural body size variation, was a secondary driver (Figure 14).

Figure 11: DAS target sizes reported from Supplier 1 were consistent with expectations and scattered around the 1:1 line. Top: scatter plot of theoretical apparent target lengths (based on known target size target height, and aircraft height) versus DAS estimates of apparent target lengths reported by Supplier 1. Black line is the 1:1 line. Bottom: Histogram of DAS estimates of apparent length reported by Supplier 1. Coloured vertical lines give the species-specific mean target length that was provided to Supplier 1 to derive DAS height estimates. Gannet (GX), Kittiwake (KI), Storm Petrel (TM).

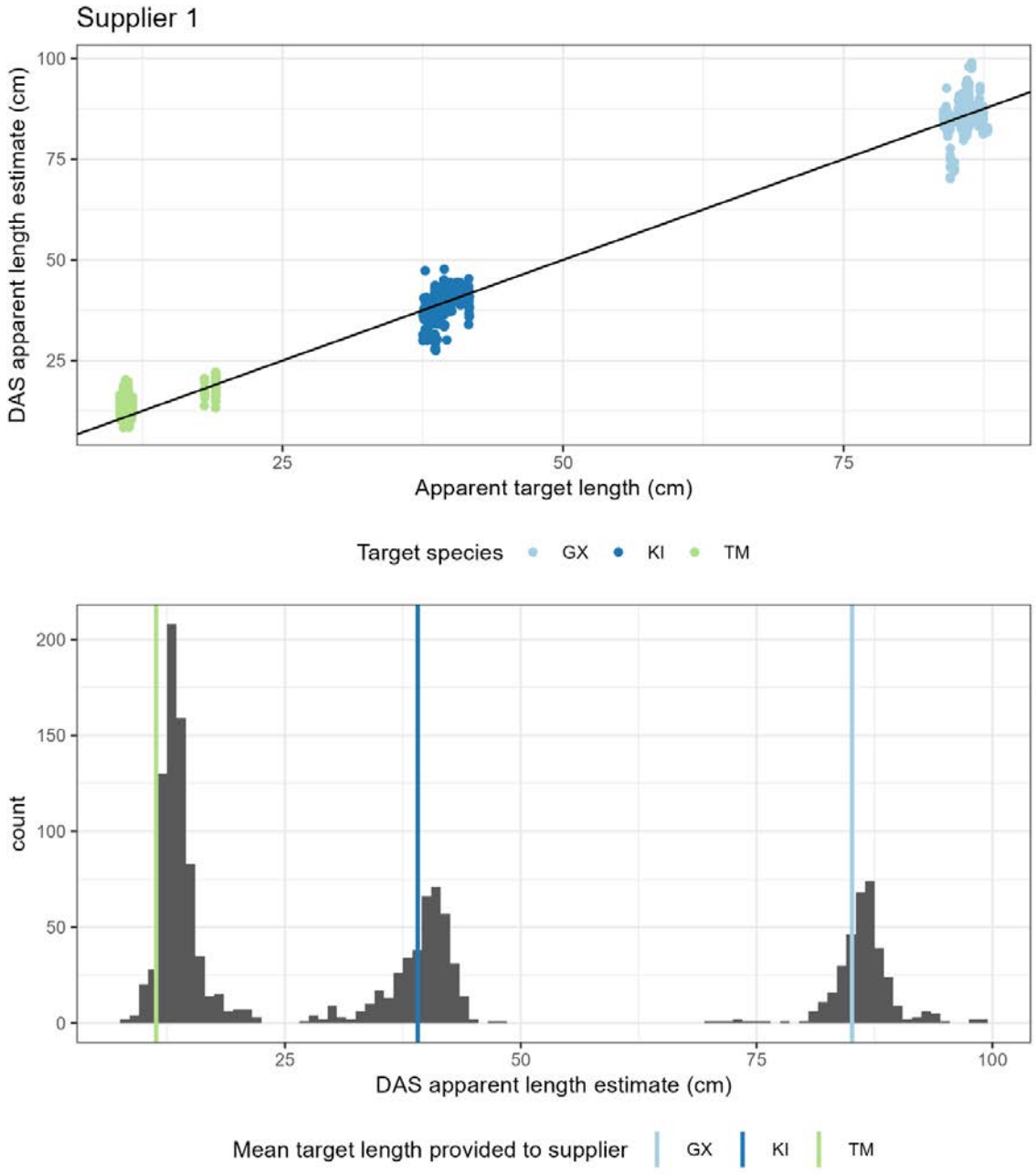


Figure 12: DAS target sizes reported from Supplier 2 were consistent with expectations and scattered around the 1:1 line. Top: scatter plot of theoretical apparent target lengths (based on known target size target height, and aircraft height) versus DAS estimates of apparent target lengths reported by Supplier 2. Black line is 1:1 line. Bottom: Histogram of DAS estimates of apparent length reported by Supplier 2. Coloured vertical lines give the species-specific mean target lengths. Gannet (GX), Kittiwake (KI), Storm Petrel (TM).

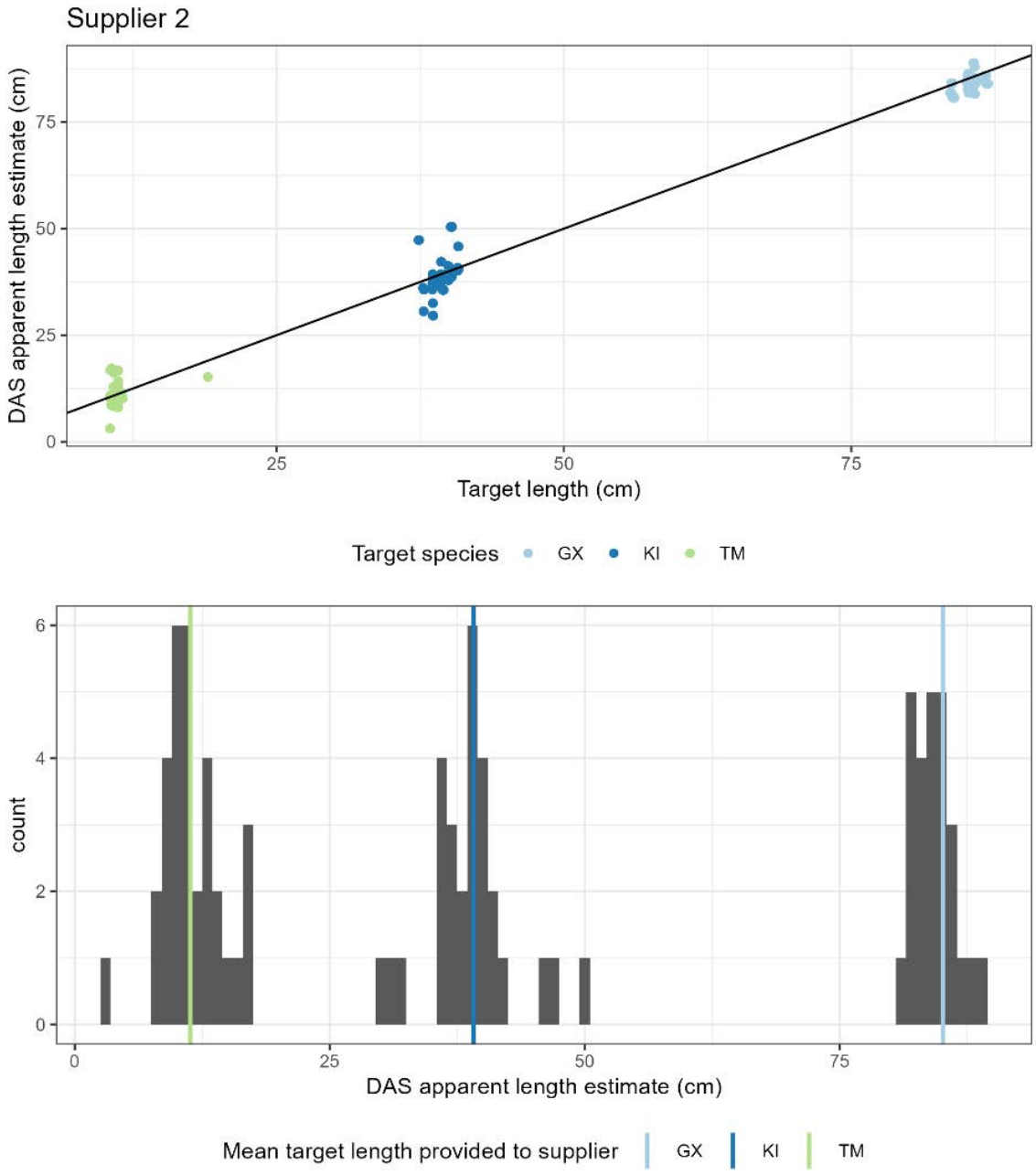


Figure 13: DAS height errors increased when photogrammetric measurements deviated from the apparent length of a target, and when the target size deviated from the mean target size for a given species. The observed DAS height errors closely followed theoretical predictions from the DAS error model in Boersch-Supan *et al.* (2024). Lines show theoretical predictions of DAS errors for a given photogrammetric measurement error (x-axis) and within-species target size deviation (encoded in colour). Points show experimental data from the validation trial. Gannet (GX), Kittiwake (KI), Storm Petrel (TM). Numerical facet labels indicate aircraft altitude in m AGL. Note differing y-axis scales between target types.

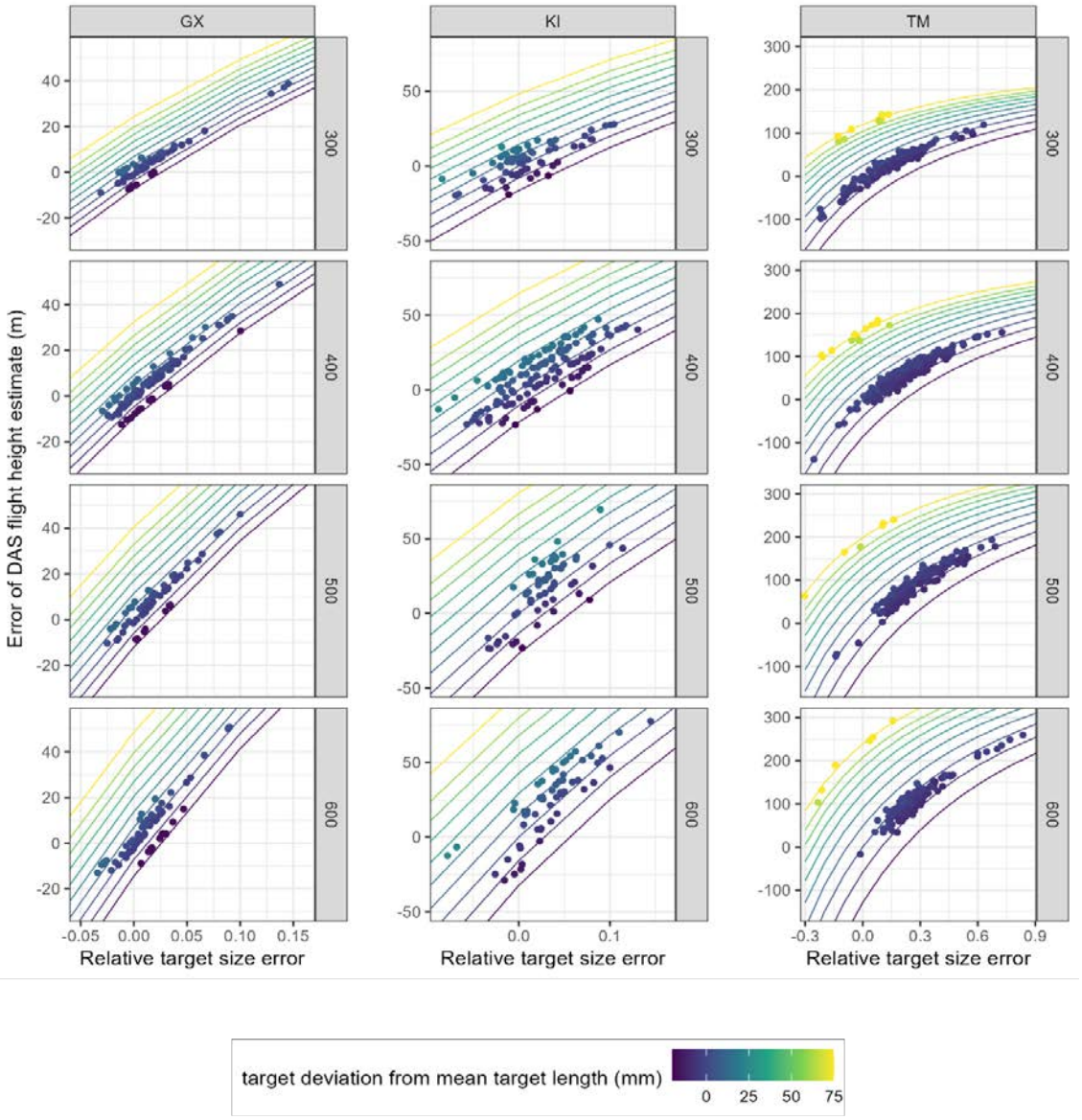
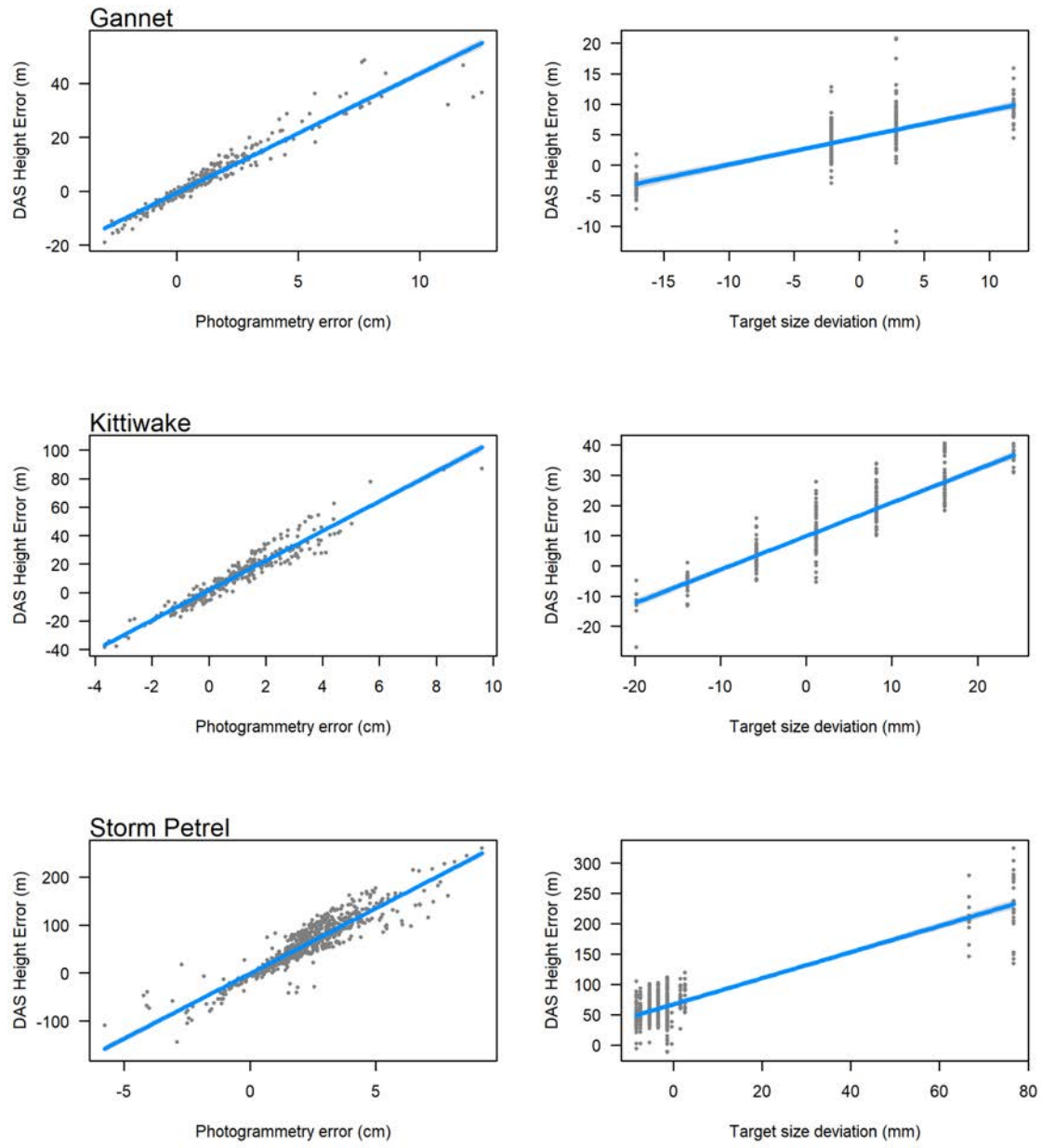


Figure 14: Photogrammetric measurement errors dominated the DAS flight height estimate errors, as the range of measurement errors was considerably larger than the range of target size variation. Figures show conditional effect plot for an additive linear regression model of DAS height error against photogrammetric measurement errors and target size deviation. Top row: Gannet targets. Middle row: Kittiwake targets. Bottom row: Storm Petrel targets. Note differing y-axis scales.



3.1.3. Moving targets

Detection rates

Figure 15 shows detection rates for each drone across all flight altitudes for the Supplier 1 survey. A total of 33 individual drone detections were made. Drone image detection rates were generally greater for the two larger drones (eBee and M300), which were both detected in 100% of passes at 500 m and 400 m altitude. Detection was reduced at flight altitudes of 300 m and 600 m, especially for the eBee drone which was detected in 25% of passes at 300 m and no passes at 600 m. The smaller Mavic drone had reduced detection at 300 m (25%), with detection consistent across the remaining altitudes (66.7%).

Figure 15: Detection rates of drones flown during the Supplier 1 survey (% detected across all passes) according to survey flight altitude (m AGL). Note that there were seven passes at 400 m flight altitude and three passes each for all other flight altitudes.

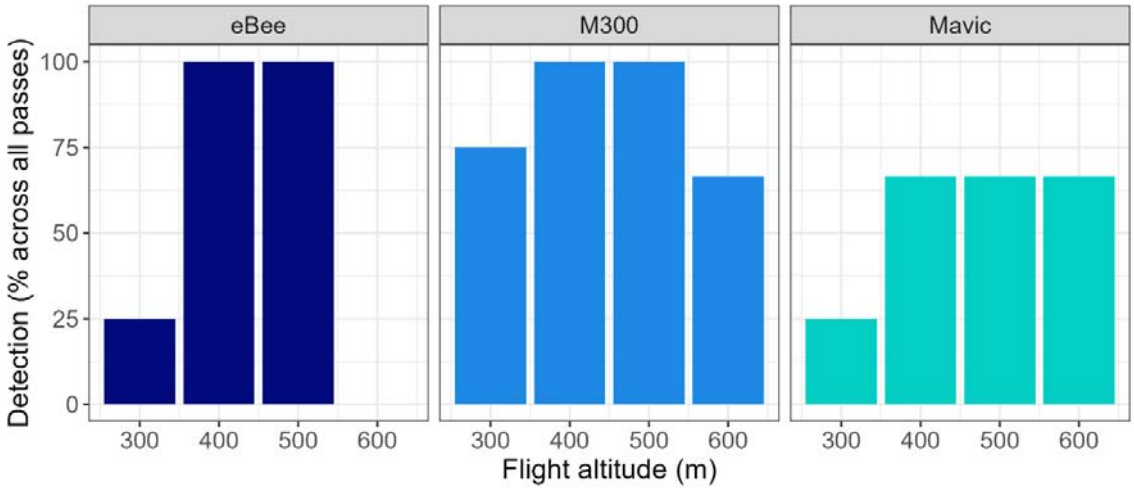
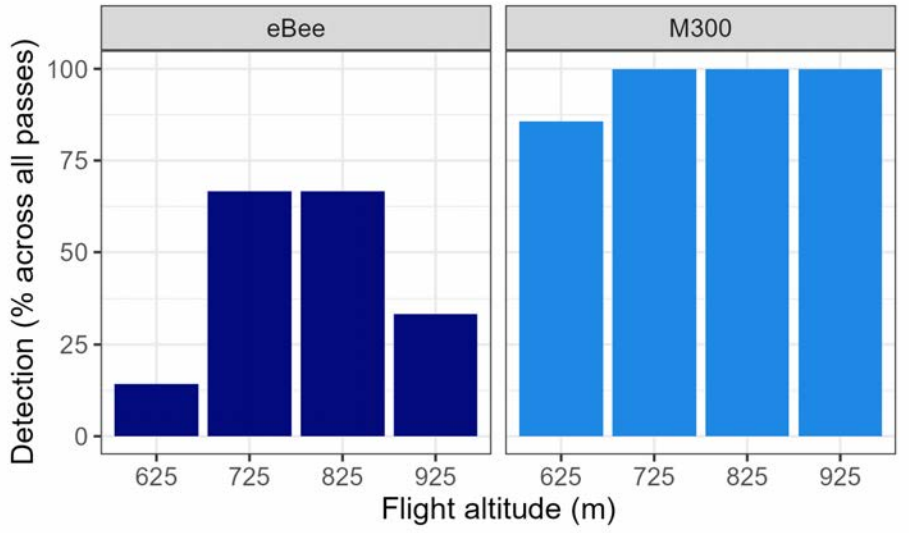


Figure 16 shows detection rates for each drone across all flight altitudes for the Supplier 2 survey (note only the eBee and M300 drones were flown). Where detected, the drones were typically identified in two or three overlapping image frames, amounting to 21 drone detections and 44 individual imagery detections across both drones. Detection of the M300 drone was higher than the eBee drone, with 85.7% detection at 625 m and 100% detection at the remaining flight altitudes. For the eBee drone, detection was highest at 725 m and 825 m at 66.7% each, reducing to 33.3% at 925 m and 14.3% at 625 m.

Figure 16: Detection rates of drones flown during the Supplier 2 survey (% detected across all passes) according to survey flight altitude (m AGL). Note that there were seven passes at 625 m flight altitude and three passes each for all other flight altitudes.



As the drone locations were not known to either supplier, it is possible that a drone may have been outside of the image field of view (FOV) for a given pass.

Positional accuracy

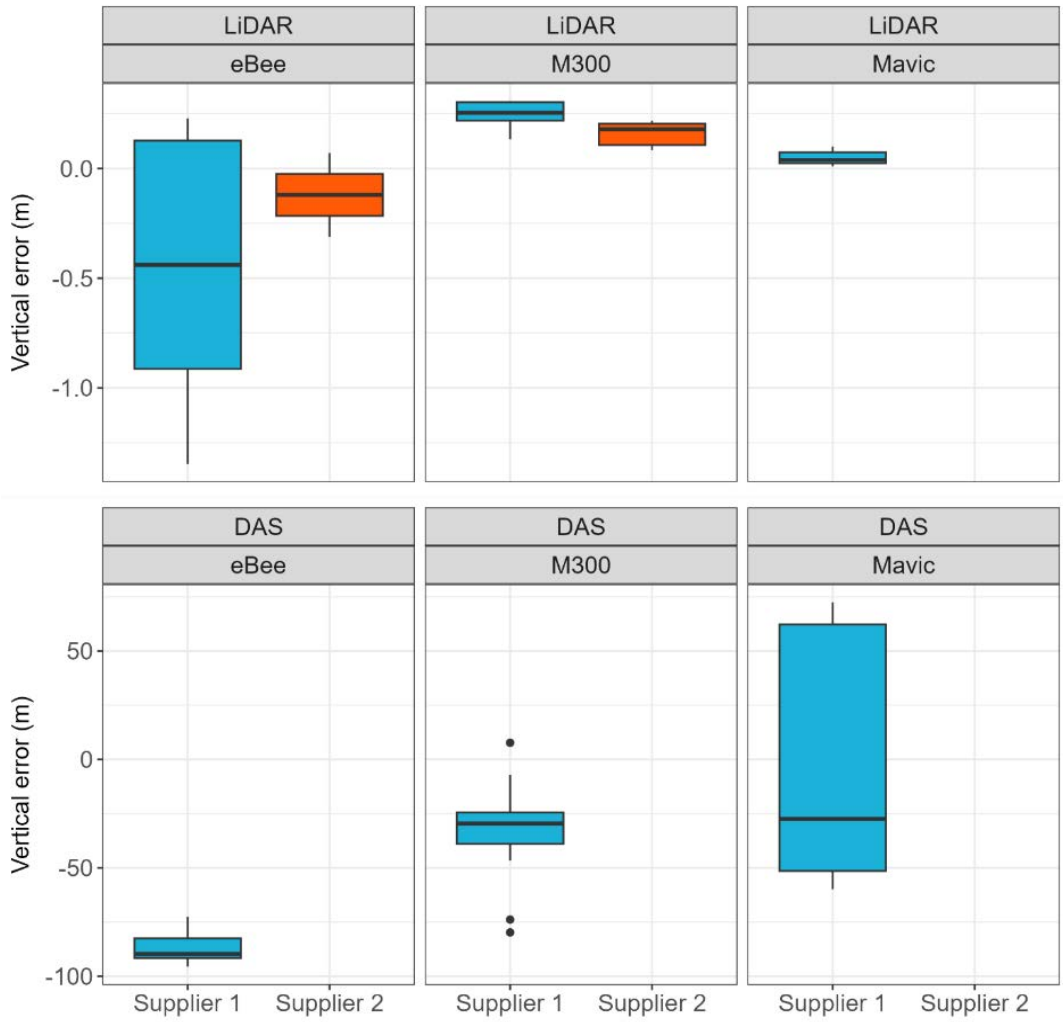
Of the 33 drone imagery detections for Supplier 1, 25 were matched in the LiDAR data. This amounted to eight LiDAR detections for eBee, 11 for M300 and six for Mavic out of a possible 16 detections for each drone (one per pass, assuming the drone was within the LiDAR FOV for the pass). For the Supplier 2 LiDAR data, the eBee and M300 drones were detected by LiDAR for five and 13 of the 16 passes respectively, either on one scan,

both scans, or producing multipath hits (hits bouncing between targets and atmospheric bands, causing an extended range value and unreliable position). As the drone locations were not known to either supplier, it is possible that a drone may have been outside of the LiDAR FOV for a given pass.

Supplier 2 noted an unusually high amount of multipath hits recorded for the drones. Supplier 2 stated that although the reason for this high amount of multipath hits was unclear, it was plausible that cloudy conditions surrounding the study site may have contributed to higher levels of atmospheric noise. Targets with multipath hits were classed as 'detected', but information regarding position or height was inaccurate and not included in the analysis.

Figure 17 shows the vertical deviation between LiDAR/size-based DAS-estimated heights and the known height of the drones. For ease of visualisation, one extreme deviation of -67 m for the Supplier 1 LiDAR estimated height versus the known height of the M300 drone has been removed. This observation was a result of the erroneously configured 300 m flight (see Appendix 2 – Figure A2–5 for inclusion of this outlier). As shown, size-based heights estimated by the Supplier 1 DAS were negatively biased. Absolute estimates of vertical error were 1.5–2 orders of magnitude greater for the size-based DAS method compared to LiDAR measurements. For the smaller Mavic drone, median vertical error in the Supplier 1 LiDAR data was very close to zero (0.05 m) and the standard deviation was also very small (0.04 m). However, in the Supplier 1 size-based DAS data, the standard deviation of vertical error was highest for the Mavic drone (57.0 m).

Figure 17: Deviation in flight heights (m) between heights estimated by LiDAR or size-based DAS heights and the known height of the drones, with extreme observation removed for ease of visualisation (see Appendix 2 for inclusion of this outlier). Note the y-axes differ between the rows and that size-based DAS-estimated heights were not provided by Supplier 2.



3.1.4. Sampling effort

Sampling effort is greater at the ground or sea surface than at high elevations due to the shape of the sampled airspace. Table 3 and Table 4 summarise the data coverage for the Supplier 1 and 2 onshore surveys respectively. As shown, point density decreased with altitude whereas point spacing and swathe width increased with altitude. For Supplier 1, data coverage for the offshore survey is presumed to be broadly similar since the same altitudes were flown.

Table 3: Realised data coverage (mean and standard deviation across individual flight passes) for the Supplier 1 onshore survey.

Flight altitude (m AGL)	Point density (points/m²)		Point spacing (m)		Imagery ground sampling distance (m)	LiDAR swathe width (m)	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Mean	Standard deviation
300	51.2	0.85	0.14	0.00	0.0105	459	7.94
400	24.4	1.23	0.20	0.01	0.0140	614	5.05
500	20.7	0.10	0.22	0.00	0.0174	751	4.00
600	8.48	0.16	0.34	0.01	0.0208	909	18.3

Table 4: Expected data coverage for the Supplier 2 onshore survey.

Flight altitude (m AGL)	Expected LiDAR point density (PPM)	Expected imagery ground sampling distance (m)	Expected LiDAR Swathe (m)
625	44.3	0.0200	298
725	38.8	0.0250	345
825	34.4	0.0275	393
925	30.8	0.0300	441

3.2. Offshore LiDAR-coupled DAS trials

3.2.1. Number of detections

Table 5 presents the number of species detected by each supplier during the offshore trial. In total, 1,435 and 520 bird detections were made by Supplier 1 and Supplier 2 respectively. Although both suppliers stated that imagery was searched for both flying and sitting birds, Supplier 2 only recorded one bird sitting on the water (0.2%), whereas sitting birds made up 51.7% of detections for Supplier 1. For both suppliers, Gannet and Kittiwake made up the majority of detected flying bird species. For Supplier 1, the majority of birds detected sitting on the surface of the water included Gannet, Puffin, Razorbill, Guillemots, auk species and Fulmar.

Absolute detection probabilities are not available for free flying birds, but match rates between imagery and LiDAR for flying birds were highly variable, which may be a result of variations in analysis methodologies of suppliers and detectability in both LiDAR and imagery. Supplier 1 achieved LiDAR match rates of 83.3% for Gannet (N=599) and 72.2% for Kittiwake (N=36), but low match rates for dark-plumaged and low-flying species (i.e. auks 40%; N=10). The majority (79%) of unmatched birds (N=135) were outside or at the edge of the LiDAR swathe (but within the area covered by the imagery). Otherwise, most other unmatched birds were displaying behaviour (e.g. banking or taking off) that prevented a sufficient profile for the LiDAR points to hit (5%), or believed to be flying close to the sea surface (8%), causing LiDAR hits to be obscured by the sea (especially at higher aircraft altitudes). Supplier 1 found that although noise was generally high for the offshore survey, only one bird was unable to be matched due to nearby noise points.

Supplier 2 achieved match rates of 90.7% for Gannet (N=107), 95.2% for Kittiwake (N=392), and 100% for Guillemot (N=6). Like Supplier 1, the majority of birds detected in imagery that that were unmatched in LiDAR were located at the edge of the LiDAR swathe width. For birds detected in LiDAR that were unmatched in imagery, detection in imagery was sometimes hampered by background sea froth or by birds flying in groups (but still producing LiDAR signatures).

Table 5: Number of offshore trial bird detections by species and behaviour for each supplier.

Species	Behaviour	Supplier 1	Supplier 2
Arctic Skua	Flying	1	0
Auk species	Sitting	36	0
Common/Arctic Tern	Flying	2	0
Curlew	Flying	2	0
Fulmar	Flying	21	1
	Sitting	30	0
Fulmar/gull species	Sitting	3	0
Gannet	Deceased	0	1
	Flying	599	107
	Sitting	281	1
	Taking off	4	0
Great Black-backed Gull	Flying	1	0
	Sitting	4	0
Great Skua	Flying	1	1
	Sitting	1	0
Guillemot	Deceased	1	0
	Flying	0	6
	Sitting	75	0
Guillemot/Razorbill	Flying	4	0
	Sitting	52	0
Herring Gull	Sitting	1	0
Kittiwake	Flying	36	392
	Sitting	6	0
Large gull species	Flying	1	0
	Sitting	2	0
Little Gull	Flying	0	2
Puffin	Sitting	149	0
Razorbill	Flying	6	0
	Sitting	93	0
Snipe	Flying	11	0
Sooty Shearwater	Flying	1	0
Unidentified bird species	Flying	4	9
	Sitting	7	0

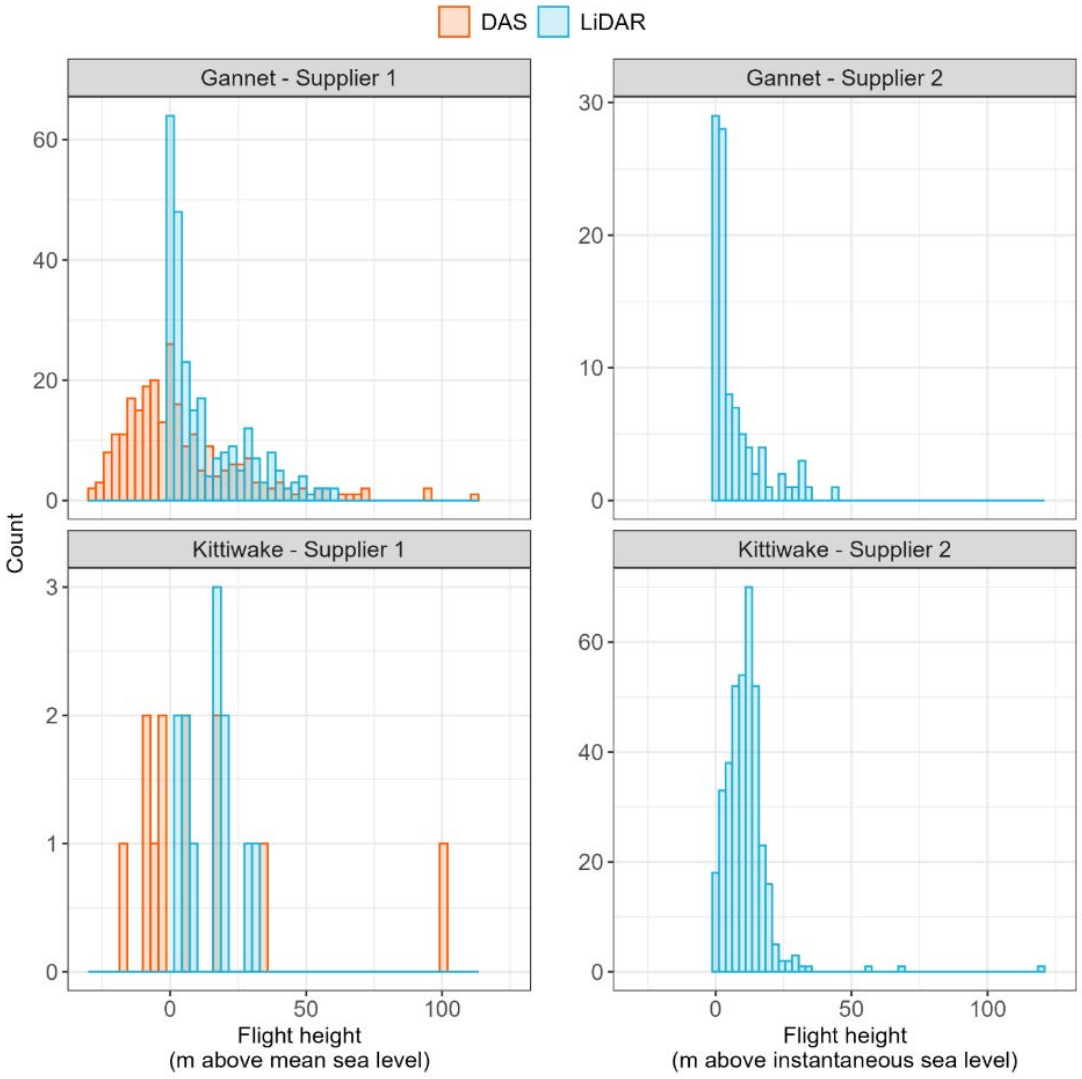
3.2.2. Height estimation

Figure 18 displays the estimated flight heights of flying Gannet and Kittiwake for both suppliers. Supplier 1 provided LiDAR flight heights both in metres AMSL and in metres above instantaneous sea level (AISL), whereas flight heights for Supplier 2 were provided in AISL (averaged across the forward and backward scans). However, flight heights for Supplier 1 are presented as metres AMSL for comparability between the LiDAR and size-based DAS method. Consequently, LiDAR heights are not directly comparable between Supplier 1 and 2.

For the Supplier 1 size-based DAS method, 50.5% of estimated flight heights were negative with a minimum estimated height of -28.5 m AMSL. Across the LiDAR data for both suppliers, only one estimated flight height was negative (-0.007 m for Supplier 1; an artefact of referencing Supplier 1 heights AMSL and comparing against Supplier 2 heights AISL, otherwise both supplier's LiDAR flight heights AISL were positive).

The proportion of flight heights below 22 m may be used as a proxy for the proportion of flight heights under the rotor swept zone of a typical wind turbine (Feather *et al.* in prep.) Using the LiDAR survey results, 86.0% (N=499) and 90.7% (N=97) of flying Gannets detected by Supplier 1 and Supplier 2 respectively were estimated to be flying at heights below 22 m (AMSL and AISL respectively). Using the Supplier 1 size-based DAS results, 79.5% (N=312) of flying Gannets were estimated to be flying at heights below 22 m AMSL. For Kittiwake, 92.3% (N=26) and 96.5% (N=373) were estimated to be flying at heights below 22 m for Supplier 1 and Supplier 2 respectively (ASML and AISL respectively). Using the Supplier 1 size-based DAS results, 58.8% (N=17) of flying Kittiwake were estimated to be flying at heights below 22 m AMSL.

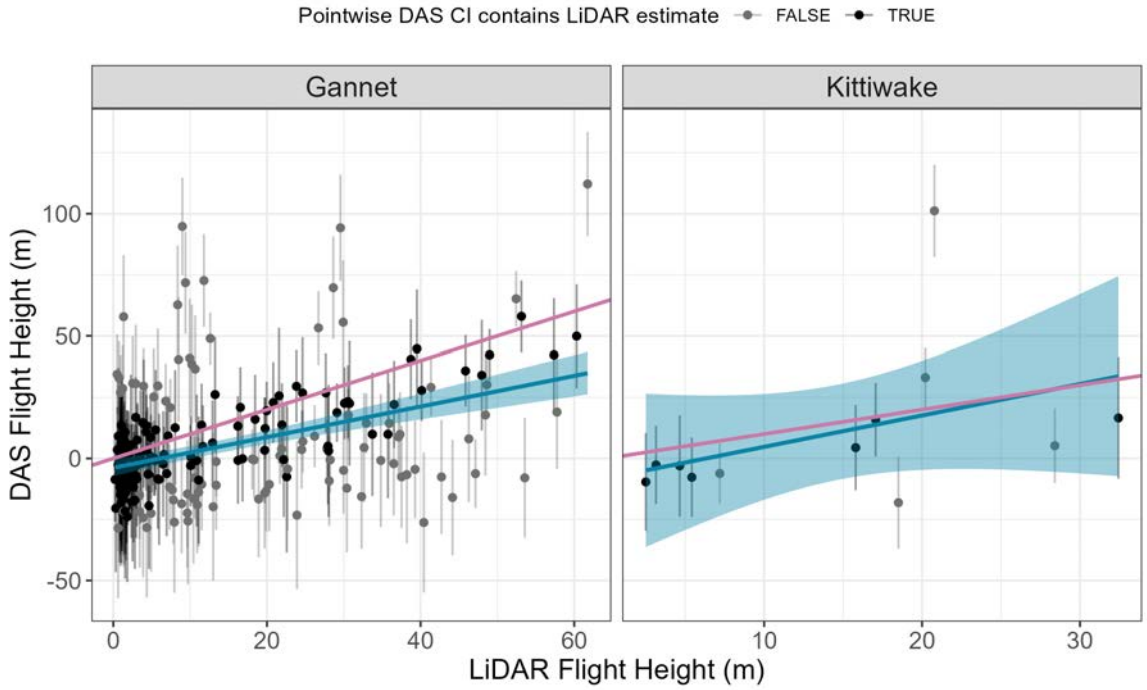
Figure 18: Estimated heights of flying birds detected by each supplier. The Supplier 2 > 100 m Kittiwake heights were presumed to be multipath measurements and therefore unreliable estimations of height.



Supplier 2 estimated the flight heights of two Kittiwake to be greater than 100 m. These observations were presumed to be multipath measurements and therefore unreliable estimations of height. These observations have been retained since they were only identified as probable multipath measurements due to being anomalously high, and so it is possible there may be other unidentified multipath measurements in the dataset. However, atmospheric noise was lower in the offshore survey than in the onshore survey, resulting in fewer multipath hits for the offshore survey. No measurements were noted as being potentially multipath by Supplier 1, although birds were assigned 'low' confidence level where a single LiDAR return is received on a bird where there is noise in the vicinity. Of 559 offshore birds detected with an assigned confidence level, 36 (6.44%) were recorded as low confidence by Supplier 1.

Figure 19 shows the deviation for matched heights for flying Gannet and Kittiwake detected in the LiDAR and DAS imagery for Supplier 1. As shown, size-based DAS-estimated heights are generally lower than LiDAR-measured heights for the same detected bird. The mean estimated LiDAR height was contained within the pointwise 95% confidence interval provided by Supplier 1 for size-based DAS heights for only 60.8% of flying Gannet and Kittiwake detected and measured by both methods.

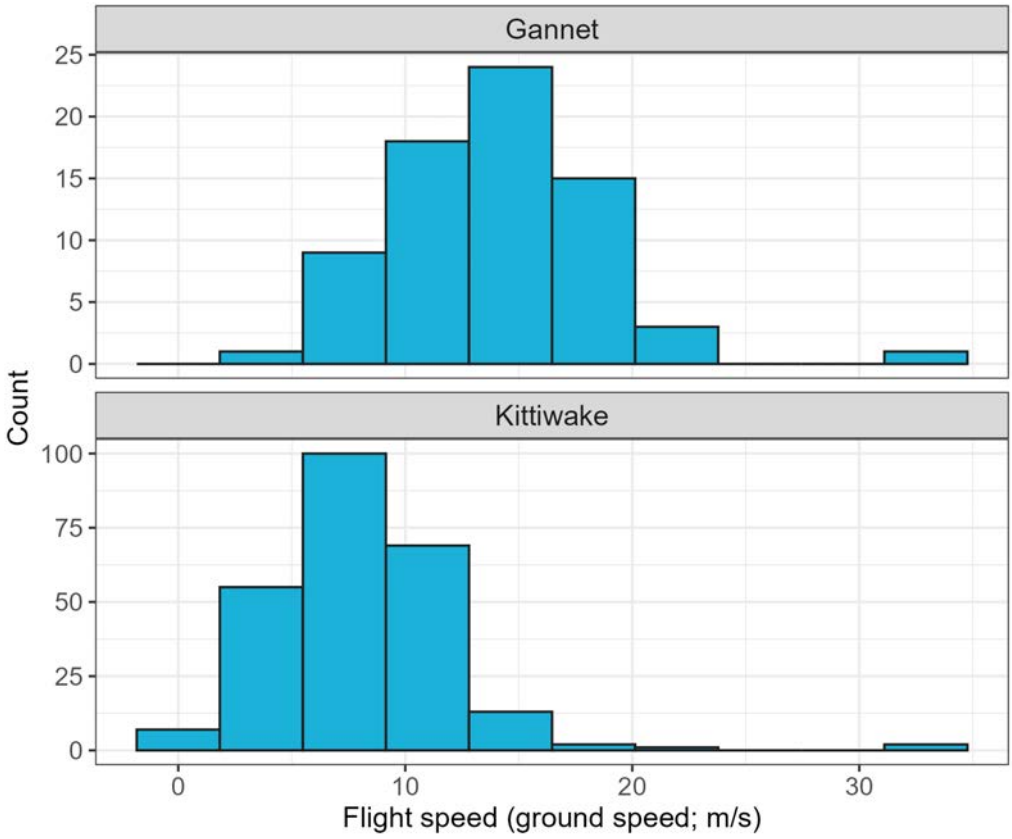
Figure 19: Deviation in flight heights estimated by size-based DAS and LiDAR surveys for flying Gannet and Kittiwake detected by Supplier 1. Purple solid line = 1:1 line. Blue line = linear regression line of best fit between LiDAR and size-based DAS flight heights. Blue shaded area = 95% CI for regression line. Points are black when the pointwise 95% CI for the size-based DAS estimate covers the corresponding LiDAR height and grey otherwise.



3.2.3. Flight speed estimation

Supplier 2 provided estimated flight speeds (ground speed) where two separate LiDAR returns were present for an individual bird, as shown in Figure 20 for Gannet (N=71) and Kittiwake (N=249). Mean estimated flight speed was 14.1 m/s (standard deviation=4.65 m/s) for Gannet and 7.87 m/s (standard deviation=4.04 m/s) for Kittiwake, which is within the range of ground flight speeds observed for these species using GPS data (Cook *et al.* 2023).

Figure 20: Flight speed (ground speed m/s) of offshore Gannet and Kittiwake detected by Supplier 2, for which two separate LiDAR returns were present.



3.3. GPS telemetry trials

Tag performance was poor overall, judged by the average number of satellites tracked compared to previous tag deployments on birds (BTO, unpublished data). For the stationary tags, this may be a result of tags being positioned close to a group of hangars and metal buildings at the airfield perimeter, potentially resulting in horizon shading and/or multipath interference and hence unreliable location fixes which are likely to be of greater uncertainty than typically expected for these tags in at-sea deployments on birds. For the tags on drones, it is possible that interference from the drones themselves may have limited the number of visible satellites and/or signal quality.

Figure 21 compares realised vertical error (i.e. deviation between altitude recorded by the GPS tag and known altitude) for stationary GPS tags and those attached to drones. The median values of vertical error were slightly positive, at 3.69 m for the stationary tags and 4.33 m for the drone tags.

Figure 22 and Figure 23 show the predicted effects of covariates on the accuracy and precision of vertical error estimates. Only battery level and accuracy indicator had a significant effect on accuracy (i.e. credible intervals did not overlap zero), whereas all covariates except known altitude had a significant effect on precision (i.e. credible intervals did not overlap one as effects on precision were modelled on a multiplicative scale).

Increasing the number of satellites was found to have a large positive effect on the precision of height estimates, though little effect on mean accuracy. GPS tag battery levels had little effect on accuracy or precision of height estimates. Coefficient estimates for the Pathtrack 'accuracy indicator' showed little effect on accuracy but a strong effect on precision, where fixes with worse (i.e. more positive) accuracy indicator values were less precise. The effects of known altitude and fix type (stationary or drone) were not independent and had little effect on accuracy, but higher altitude drone fixes were estimated to be of reduced precision than lower altitude stationary fixes.

Figure 21: Realised vertical error (deviation between altitude recorded by the GPS tag and known altitude), for tags that were stationary or attached to drones in flight.

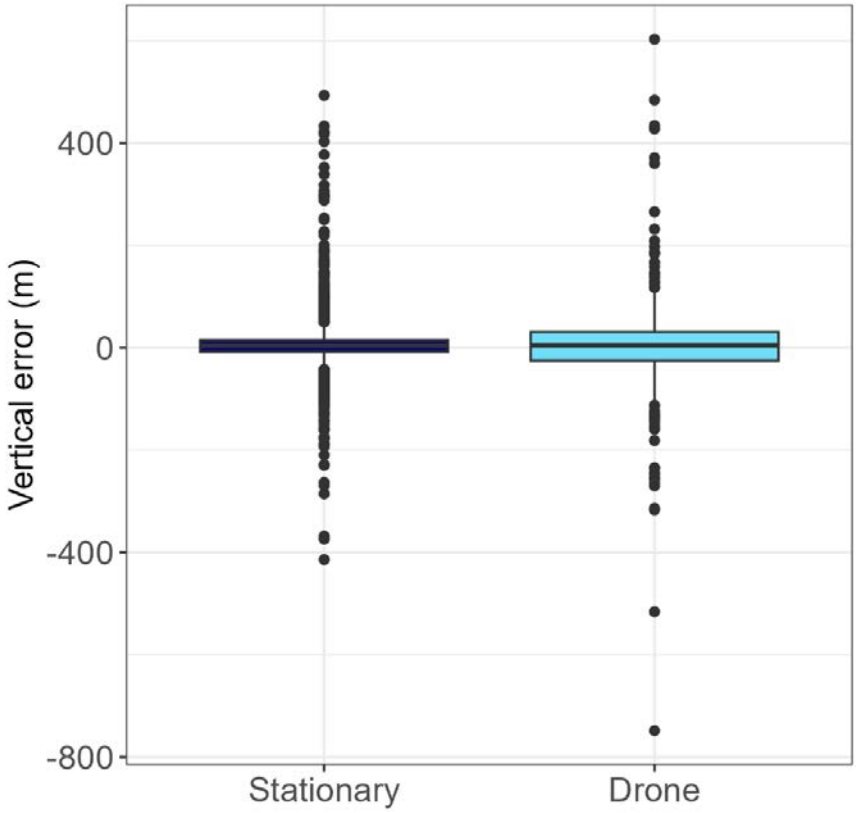


Figure 22: Predicted coefficient estimates (inlcuding 95% credible intervals) for accuracy (μ) and precision (τ, converted from σ) of vertical height error of the GPS tags. Coefficients for accuracy are on an additive scale and so a coefficient of 0 represents no effect, whereas coefficients for precision are on a multiplicative scale where a coefficient of 1 represents no effect. Coefficient estimates are for scaled predictors (mean-centred with a standard deviation of 1) to facilitate comparison of effect sizes between predictors.

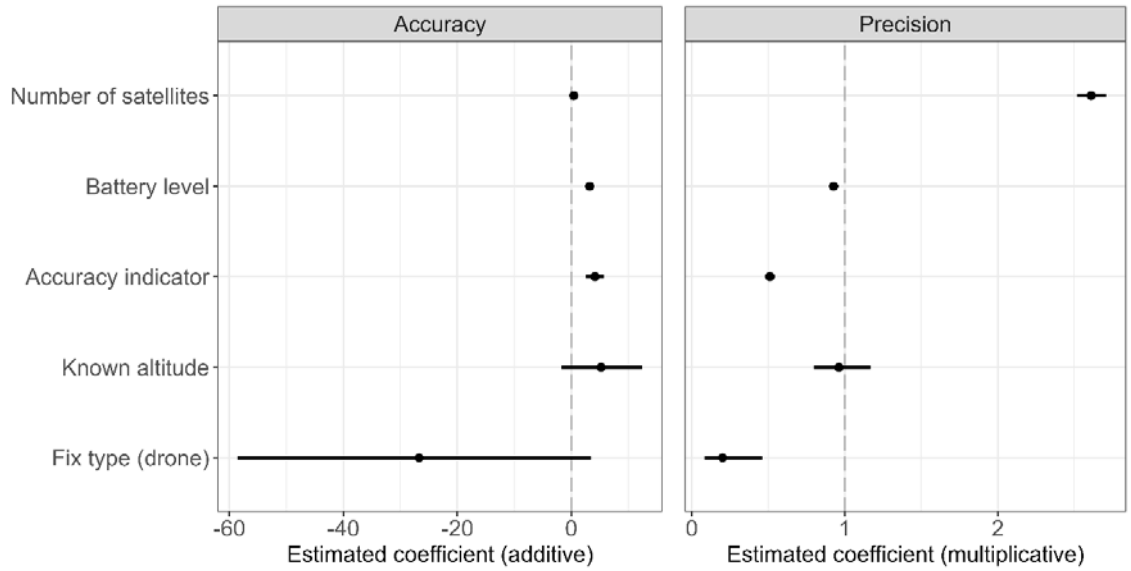
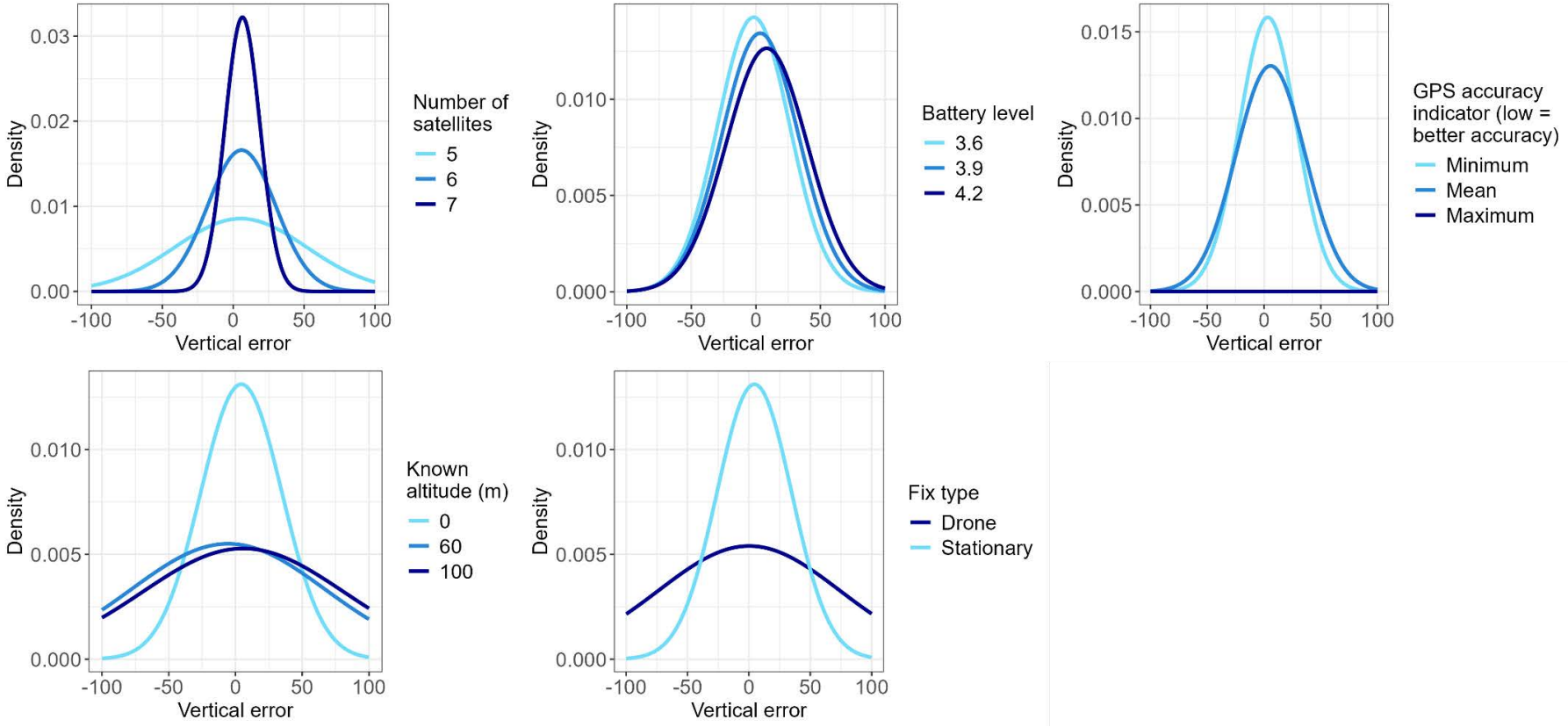


Figure 23: Conditional effects of covariates on predicted vertical error (i.e. discrepancy between altitude estimated by the GPS tags and the known altitude of the drone). Non-focal covariates are fixed to the mean for continuous variables and to the reference level for fix type (stationary), except for known altitude and fix type. In these cases, predictions of 0 m AGL known altitude are conditional on stationary fix type and predictions of 60 m and 100 m AGL are conditional on drone fix type. Conversely, stationary fix types are conditional on altitudes of 0 m AGL, and drone fix types are conditional on 80 m AGL, i.e. an average of 60 m and 100 m.



3.4. LRF trials

3.4.1. Mobile target tracking

Fixes on the fixed wing eBee X and the small rotary Mavic 3E drones proved difficult to obtain in week one, and therefore effort was concentrated on the large rotary drone Matrice M300 in week two. As a result, the total of 670 fixes on mobile targets across both devices mostly contained fixes from the Matrice M300 drone (N=582; Table 6). Drone RTK positions were only recorded while drones were in level flight on pre-determined transects. The majority (80%, N=539) of the obtained LRF fixes fell within these transects, and these were taken into consideration for the analysis (Table 7).

Table 6: Gross laser rangefinder sample sizes obtained during mobile target tracking.

Target	Device		Total
	Nikon Forestry Pro II	Vectronix Terrapin X	
eBee X	12	29	41
M300	279	303	582
Mavic 3E	27	20	47
Total	**318**	**352**	**670**

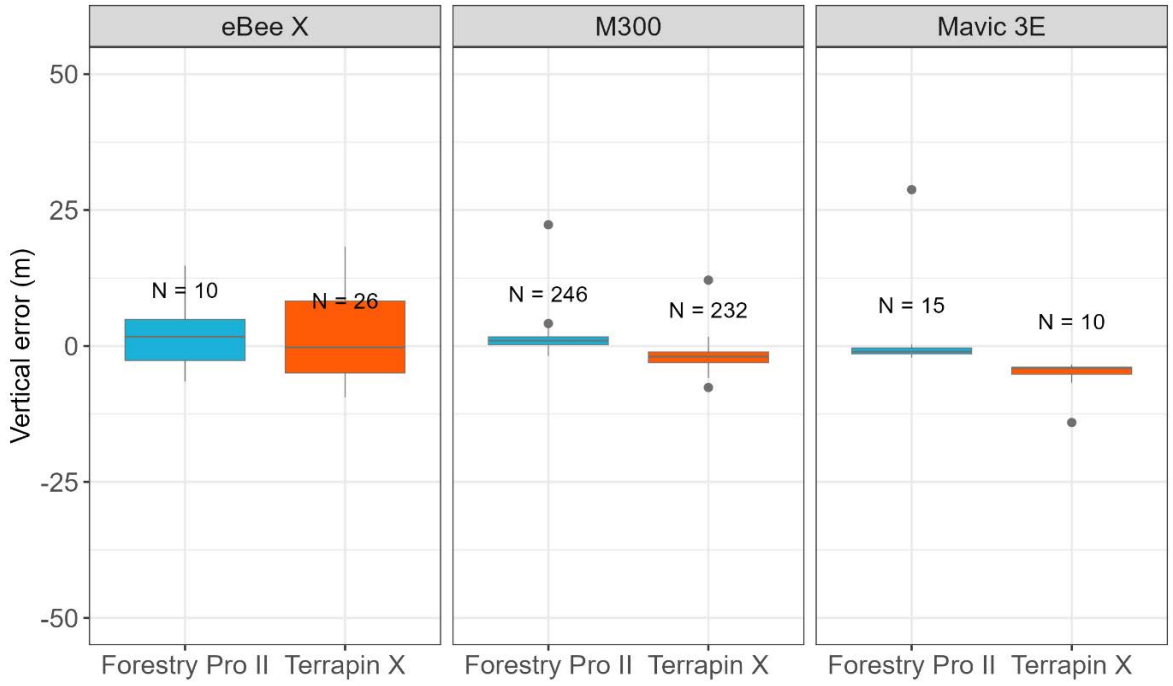
Table 7: Laser rangefinder sample sizes within predefined drone transects.

Target	Device		Total
	Nikon Forestry Pro II	Vectronix Terrapin X	
eBee X	10	26	36
M300	246	232	478
Mavic 3E	15	10	25
Total	**271**	**268**	**539**

Realised transect heights for the rotary drones were within ±10 cm (0.1%) of the nominal heights, however vertical positioning was much less precise for the fixed wing eBee X, which undulated up to ±10 m around the nominal height.

Altitude estimates for moving targets from both LRF models were generally within ±10 m of the reconstructed target heights for the fixed wing eBee X drone, and within ±2 m of the rotary drones, although large outliers occurred. We therefore used a robust regression approach to estimate measurement accuracy (bias) and precision. Robust regression models estimated small, but statistically significant positive bias for the Nikon Forestry Pro II on the M300 (1.00 m for the M300, 95% CI 0.823–1.18) and a slight negative bias on the Mavic 3E (1.01 m, 95%CI -1.63 – -0.38). For the Terrapin X, small but slightly larger negative biases were found for these two targets (-2.06 m, 95% CI -2.35 – -1.76 for the M300; -4.07, 95% CI -4.80 – -3.35 for the Mavic 3E). Measurements of the eBee X flight height for both devices were not significantly different from the drone records, but noting small sample sizes for the eBee and the Mavic 3E. However, precision for both devices was generally on the order of ±10 m (Figure 24), with 90% of measurements of the Nikon Forestry Pro II falling between -4.07 m and 3.39 m of the M300 height, between -7.97 m and 2.05 m for the Mavic 3E, and between -34.4 m and 22.2 m for the eBee X. Precision was comparable for the Terrapin X with 90% of measurements between -9.49 m and 1.29 m for the M300, between -14.1 m and 0.831 m for the Mavic 3E, and between -58.9 m and 28.3 m for the Mavic 3E. Again, noting very small sample sizes for the eBee X and Mavic 3E (Table 7). Effort was not consistent across users and time and only positive hits were recorded – misses were not, so detection rates could not be estimated from these observations.

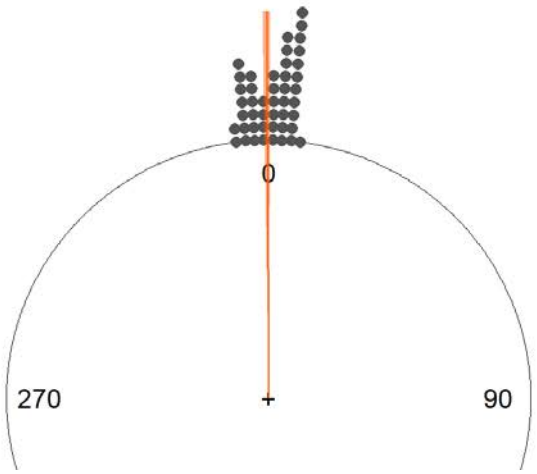
Figure 24: Altitude estimates for moving targets from both LRF models were generally within ±10 m of the reconstructed target heights of the drones. Data shown are limited to LRF hits while the drones were in level flight at predetermined heights.



3.4.2. Magnetic compass accuracy and precision

The azimuth trial on static targets found no evidence for a directional bias of the magnetic compass in the Terrapin X, with a mean angular deviation of 0.48° (95% CI -0.35–1.32), and a circular standard deviation of measurements of ±4.09° (Figure 25). Attempts to fix a target were limited to five and fixes were obtained for all Gannet and Kittiwake targets, but the operators struggled to get hits on Storm Petrel sized targets, obtaining hits for 7/10 light Storm Petrel targets and only 5/12 dark Storm Petrel targets.

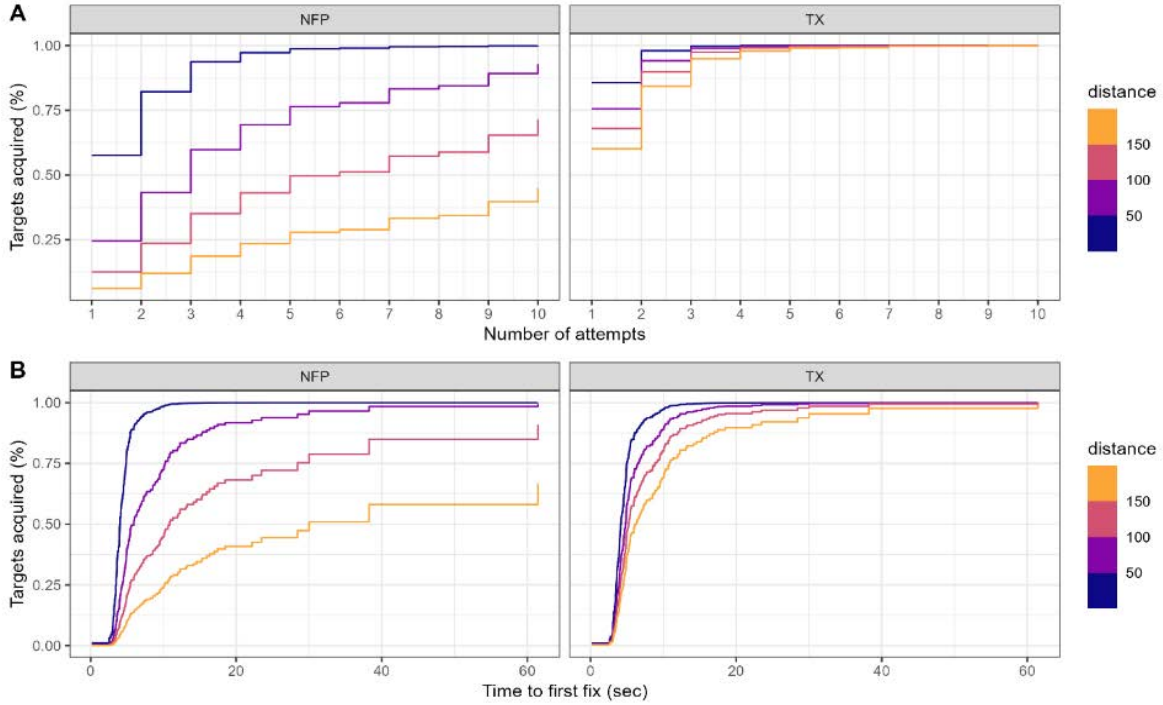
Figure 25: Terrapin X azimuth (compass bearing in degrees) readings for static targets (N=46) on average did not differ significantly from true directions, however, deviations of up to ±5° for individual targets were common. Orange line and shading show mean direction and 95% confidence interval.



3.4.3. Time to first fix experiment

Effort required to get an LRF fix on a static target was found to vary significantly between devices and target distances. All operators managed to fix targets at any range with the Terrapin X device, generally requiring fewer than three attempts (Figure 26). Time to first fix was generally less than 10 seconds at 25 m and increased to less than 40 seconds at 200 m (Figure 26). Targets were more difficult to acquire with the Nikon Forestry Pro II. While all operators acquired all targets at 25 m within seven attempts, only 45% of targets were acquired at 200 m within 10 attempts (Figure 26). Time to first fix was not significantly different to the Terrapin X at 25 m (< 10 seconds), but increased more rapidly with distance to < 40 seconds at 75 m and beyond 40 seconds larger distances (Figure 26).

Figure 26: Acquisition of static targets was faster overall and required fewer attempts with the Terrapin X rangefinder. Plots show predictions from a Cox proportional hazard model. Nikon Forestry Pro II (NFP), Terrapin X (TX)



4. Discussion

This study comprised four field trials (onshore and offshore LiDAR-coupled DAS trials, GPS telemetry trials and LRF trials) with the aim of comparing and investigating sources of uncertainty associated with the use of different methods for estimating seabird flight heights.

For the onshore LiDAR surveys, smaller and dark static targets were detected less frequently than larger and light targets. These findings were also reflected for the moving targets (in which the smaller drone was detected less frequently than the larger drones) and potentially in the offshore trials (in which Gannet and Kittiwake comprised the majority of flying bird detections, but noting that baseline frequencies of different species were unknown). Consequently, density estimates for small and dark-plumaged species that are uncorrected are likely to be underestimated in real-world surveys due to lower detection rates. However, accuracy and precision of LiDAR-estimated heights were very similar across static target and drone types in relation to the scale of realised flight heights. These results show that although LiDAR sampling efficiency is species dependent, LiDAR measured heights for detected birds have similar accuracy and precision across different species. The LiDAR instrumentation used by Supplier 2 also facilitated the estimation of flight speed (which is also required for collision risk modelling), and estimated flight speeds for birds detected during the offshore survey were consistent with those recently estimated using animal-borne GPS technologies. However, further validation work is needed to provide confidence in the estimated flight speeds derived using this method.

For Supplier 1, detection rates from onshore DAS surveys were overall greater than those from the LiDAR data. For Supplier 2, detection rates were generally similar between the LiDAR and DAS data, except for dark Storm Petrel targets for which detection was better in the LiDAR data. These results are most plausibly explained by variations in each supplier's data processing and analysis procedures, e.g. Supplier 1 utilised higher image resolution but lower LiDAR point density compared to Supplier 2. Accuracy and precision of flight height estimations were substantially reduced in the size-based DAS method used by Supplier 1 compared to the LiDAR measurements for both static and moving targets, with size-based DAS estimates generally having measurement uncertainties two to three orders of magnitude larger than LiDAR measurements, i.e. 10 seconds or metres rather than centimetres. As expected from theoretical considerations (Boersch-Supan *et al*. 2024, Brighton *et al*. 2025), this finding was generally more exaggerated for small and dark targets and higher survey altitudes, which result in photogrammetric measurement errors that are proportionally larger relative to the true target size. Photogrammetric errors were identified as the primary driver of DAS height errors in the onshore study, where they had a consistently larger influence then target size variation. Overall, these findings suggest that LiDAR surveys may result in improved estimations of flight height distributions, whereas use of DAS may achieve a greater sampled volume. However, since the Supplier 1 onshore survey was conducted in suboptimal weather conditions, DAS height estimates may potentially be worse than would be expected in improved weather conditions (e.g. potentially due to degraded image quality). In addition, comparison of results between LiDAR and size-based DAS flight height estimates is likely to be influenced by the specific survey equipment, configuration, and image processing methods used. The size-based DAS results presented here are based on the methodological approach followed by a single supplier. Other suppliers utilise slightly different methods to estimate flight height from imagery (e.g. Humphries *et al*. 2023), and although the present study confirms theoretical predictions about the scale of uncertainty in photogrammetric flight height estimates, developments aiming to improve population-level inferences from imprecise individual flight height measurements produced by size-based methods are ongoing (e.g. Forster *et al*. 2024). Size-based DAS height estimates were not provided for Supplier 2, precluding comparison of results between the two commissioned suppliers.

Detection rate and accuracy/precision of height estimation varied between the two aircraft suppliers. This highlights that differing weather conditions and equipment types had a strong influence on survey results, introducing unknown levels of uncertainty. This uncertainty could potentially be quantified by performing onshore calibration using targets of known heights before conducting offshore surveys. Such calibration surveys could also be used to investigate other sources of uncertainty such as reasons for failed matches between LiDAR and imagery detections, causes of multipath interference and noise LiDAR hits, and validation of detection of birds sitting on the surface of the water. However, such onshore trials are likely to represent 'best-case' results since target locations are known and influences of noise inherent to at-sea surveys would not be present. This also highlights the need to carry out actual surveys in the best environmental conditions possible to limit the impact of noise, although doing so is also likely to introduce sampling bias by limiting observations to flight height behaviours that are realised in fair weather (Feather *et al*. 2024), given that weather conditions influence flight behaviour (Davies *et al*. 2024).

The results of the GPS telemetry trial generally showed that precision of height estimates was poorer and more variable than measurement accuracy. As discussed for size-based estimates above, this has important implications for the estimation of flight height distributions, which are of particular interest for collision risk estimation, as distributional estimates are much more sensitive to measurement uncertainty than estimates of mean flight heights. Our findings show that covariates potentially impacting precision of flight height estimation, such as the number of visible satellites for the fix, need to be accounted for when estimating flight height distributions from telemetry data, as demonstrated by previous studies (Ross-Smith *et al*. 2016, Péron *et al*. 2020).

The LRF trials demonstrated that although the LRF devices generally performed to specification for determining ranges to, and heights of, static and moving targets, obtaining fixes was an inconsistent process and sampling efficiency was often low, in part because the device interfaces were not designed to facilitate the rapid and accurate logging of moving targets. Lack of exact time recording was a particular hurdle for both evaluated devices in the validation study. This, and the quantification of detection rates and effective sampling volumes, are elements that require further work to significantly improve the utility of this method.

There were a number of limitations to the trials that potentially influenced the results of this study. Performance of tags in the telemetry trial was poor overall in comparison to ongoing field deployments of similar tags (BTO, unpublished data); in particular, the average number of satellites tracked was low compared to previous tag deployments on birds. For stationary tags, this may be a result of tags being positioned close to a group of hangars and metal buildings at the airfield perimeter, potentially resulting in multipath interference and hence unreliable location fixes which are likely to be of greater uncertainty than typically expected for these tags. For the tags on drones, it is possible that interference from the drones themselves may have limited the number of visible satellites. Although the existence and potential extent of this interference is currently unknown, additional experiments aiming to quantify potential drone interference are planned. In addition, the drone positional tracking by onboard RTK devices was not achieved at the planned temporal resolution and extent. As a result, drone tracks required interpolation to achieve one estimated fix position per second. Uncertainty regarding the actual position of the drone may therefore have influenced the results of the GPS telemetry and LRF trials. For the eBee drone in particular, timestamp-offsets between the GPS tags and the interpolated drone fixes may have impacted the reliability of results, especially since the drone was undulating, so vertical error was within the range of 1 m rather than 1–10 cm like the rotary drones. Future studies should ensure recording of RTK GPS data is appropriately set up for drone trials. In addition, height of the sea surface can vary by several metres between the crest and trough of the wave and across tidal cycles, leading to significant differences between flight altitudes AMSL and AISL, which is a further cause of uncertainty over derived flight height of seabirds (Feather *et al.* 2025).

Finally, for the DAS-coupled LiDAR surveys, successful detection of sitting and free flying birds offshore could not be verified for either supplier since no ground truthing of targets was possible during the study and supplier image analysis and LiDAR processing methods were not subject to external quality assurance. The stark difference between the proportion of sitting birds recorded in the offshore surveys (51.7% for Supplier 1% vs. 0.2% for Supplier 2) is difficult to reconcile with ecological expectations about behavioural variability in the survey area at this time of the year and highlights a possible requirement for a validation procedure to establish how reliably a survey system and processing approach is able to detect birds sitting on the water.

Overall, this study has highlighted potential sources of uncertainty relating to different methods for estimating flight height distributions. Based on the results of the field trials, accuracy and precision of LiDAR height estimates outweighed size-based DAS height estimates, although the DAS trials resulted in better detection probabilities (depending on equipment and data processing methods used). Comparison of airborne survey data was influenced by a combination of environmental conditions, sensor configuration differences between suppliers, and supplier-specific processing workflows. Although comparison of LiDAR and drone heights generally remained within 1 m, size-based DAS height and detection rate uncertainties varied more noticeably. These findings support the value of survey-day calibration checks to verify the accuracy and precision of the LiDAR data for individual surveys.

Our results highlight the need for calibration and validation of monitoring approaches. Future surveys should aim to quantify potential sources of uncertainty so they may be accounted for when estimating the shape of flight height distributions.

5. References

Agostinelli, C. & Lund, U. 2024. 'R package "circular": Circular Statistics (version 0.5-1)'. Available at: https://CRAN.R-project.org/package=circular.

Ainley, D., Porzig, E., Zajanc, D. & Spear, L.B.. 2015. Seabird flight behavior and height in response to altered wind strength and direction. *Marine Ornithology* **43**: 25–36.

Boersch-Supan, P.H., Brighton, C.H., Thaxter, C.B. & Cook, A.S. 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.* (2023). *Marine Biology* **171**: 122.

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.

Bürkner, P.-C. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* **80**: 1–28.

Cook, A.S.C.P., Thaxter, C.B., Davies, J., Green, R.M.W., Wischnewski, S. & Boersch-Supan, P. 2023. Understanding seabird behaviour at sea part 2: improved estimates of collision risk model parameters. Scottish Government. Available at: https://www.gov.scot/publications/understanding-seabird-behaviour-sea-part-2-improved-estimates-collision-risk-model-parameters/pages/1/.

Davies, J.G., Boersch-Supan, P.H., Clewley, G.D., Humphreys, E.M., O'Hanlon, N.J., Shamoun-Baranes, J., Thaxter, C.B., Weston, E. and Cook, A.S. 2024. Influence of wind on Kittiwake *Rissa tridactyla* flight and offshore wind turbine collision risk. *Marine Biology* **171**: 191.

van Erp, J.A., van Loon, E.E., De Groeve, J., Bradarić, M. and Shamoun-Baranes, J. 2023. A framework for post-processing bird tracks from automated tracking radar systems. *Methods in Ecology and Evolution* **15**: 130–143.

Feather, A., Rhoades, J. & Boersch-Supan, P.H. (in prep.) ReSCUE Project Gap and Power Analysis – Preliminary report with a focus on LiDAR coverage. British Trust for Ornithology, Thetford.

Feather, A.P., Burton, N.H.K., Johnston, D.T. & Boersch-Supan, P.H. *et al*. 2025. 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* **780**. British Trust for Ornithology, Thetford.

Forster, J., Macleod, K. & Scott, M. 2024. Response to: "Natural body size variation in seabirds provides a fundamental challenge for flight height determination by single-camera photogrammetry". *Marine Biology* **171**: 120.

Fox, A.D., Desholm, M., Kahlert, J., Christensen, T.K. & Krag Petersen, I.B. 2006. Information needs to support environmental impact assessment of the effects of European marine offshore wind farms on birds. *Ibis* **148**: 129–144.

Grant, M.L., Bond, A.L. & Lavers, J.L. 2022. The influence of seabirds on their breeding, roosting and nesting grounds: a systematic review and meta-analysis. *Journal of Animal Ecology* **91**: 1,266–1,289.

Humphries, G.R.W., 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. *Marine Biology* **170**: 18.

Kaldellis, J.K. & Zafirakis, D. 2011. The wind energy (r)evolution: a short review of a long history. *Renewable Energy* **36**: 1,887–1,901.

Lane, J.V., Jeavons, R., Deakin, Z., Sherley, R.B., Pollock, C.J., Wanless, R.J. & Hamer, K.C.. 2020. Vulnerability of Northern Gannets to offshore wind farms; seasonal and sex-specific collision risk and demographic consequences. *Marine Environmental Research* **162**: 105196.

Largey, N., Cook, A.S., Thaxter, C.B., McCluskie, A., Stokke, B.G., Wilson, B. & Masden, E.A. 2021. Methods to quantify avian airspace use in relation to wind energy development. *Ibis* **163**: 747–764.

Lato, K.A., Stepanuk, J.E., Heywood, E.I., Conners, M.G. & Thorne, L.H. 2022. Assessing the accuracy of altitude estimates in avian biologging devices. *PLOS ONE* **17**: e0276098.

Masden, E.A., Cook, A.S., McCluskie, A., Bouten, W., Burton, N.H. & Thaxter, C.B. 2021. When speed matters: the importance of flight speed in an avian collision risk model. *Environmental Impact Assessment Review* **90**: 106622.

Masden, E.A. & Cook, A.S.C.P. 2016. Avian collision risk models for wind energy impact assessments. *Environmental Impact Assessment Review* **56**: 43–49.

Mosbech, A., Johansen, K.L., Davidson, T.A., Appelt, M., Grønnow, B., Cuyler, C., Lyngs, P. & Flora, J. 2018. On the crucial importance of a small bird: the ecosystem services of the Little Auk (*Alle alle*) population in Northwest Greenland in a long-term perspective. *Ambio* **47**: 226–243.

Paleczny, M., Hammill, E., Karpouzi, V. & Pauly, D. 2015. Population trend of the world's monitored seabirds, 1950–2010. *PLOS ONE* **10**: e0129342.

Péron, G., Calabrese, J.M., Duriez, O., Fleming, C.H., García-Jiménez, R., Johnston, A., Lambertucci, S.A., Safi, K. & Shepard, E.L.. 2020. The challenges of estimating the distribution of flight heights from telemetry or altimetry data. *Animal Biotelemetry* **8**: 5.

R Core Team. 2024. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at: https://www.R-project.org.

Rahman, A., Farrok, O. & Haque, M.M. 2022. Environmental impact of renewable energy source based electrical power plants: solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic. *Renewable and Sustainable Energy Reviews* **161**: 112279.

Ross-Smith, V.H., Thaxter, C.B., Masden, E.A., Shamoun- Baranes, J., Burton, N.H., Wright, L.J., Rehfisch, M.M. & Johnston, A. 2016. Modelling flight heights of lesser black-backed gulls and great skuas from GPS: a Bayesian approach. *Journal of Applied Ecology* **53**: 1,676–1,685.

Shields, M.A., Dillon, L.J., Woolf, D.K. & Ford, A.T. 2009. Strategic priorities for assessing ecological impacts of marine renewable energy devices in the Pentland Firth (Scotland, UK). *Marine Policy* **33**: 635–642.

Signa, G., Mazzola, A. & Vizzini, S. 2021. Seabird influence on ecological processes in coastal marine ecosystems: an overlooked role? A critical review. *Estuarine, Coastal and Shelf Science*. **250**: 107164.

Stephan, T., Enkelmann, E. & Kroner, U. 2023. Analyzing the horizontal orientation of the crustal stress adjacent to plate boundaries. *Scientific Reports* **13**: 15590.

Thaxter, C.B., Ross-Smith, V.H. & Cook, A.S.C.P. 2015. How high do birds fly? A review of current datasets and an appraisal of current methodologies for collecting flight height data: Literature review. *BTO Research Report* **666**. British Trust for Ornithology, Thetford.

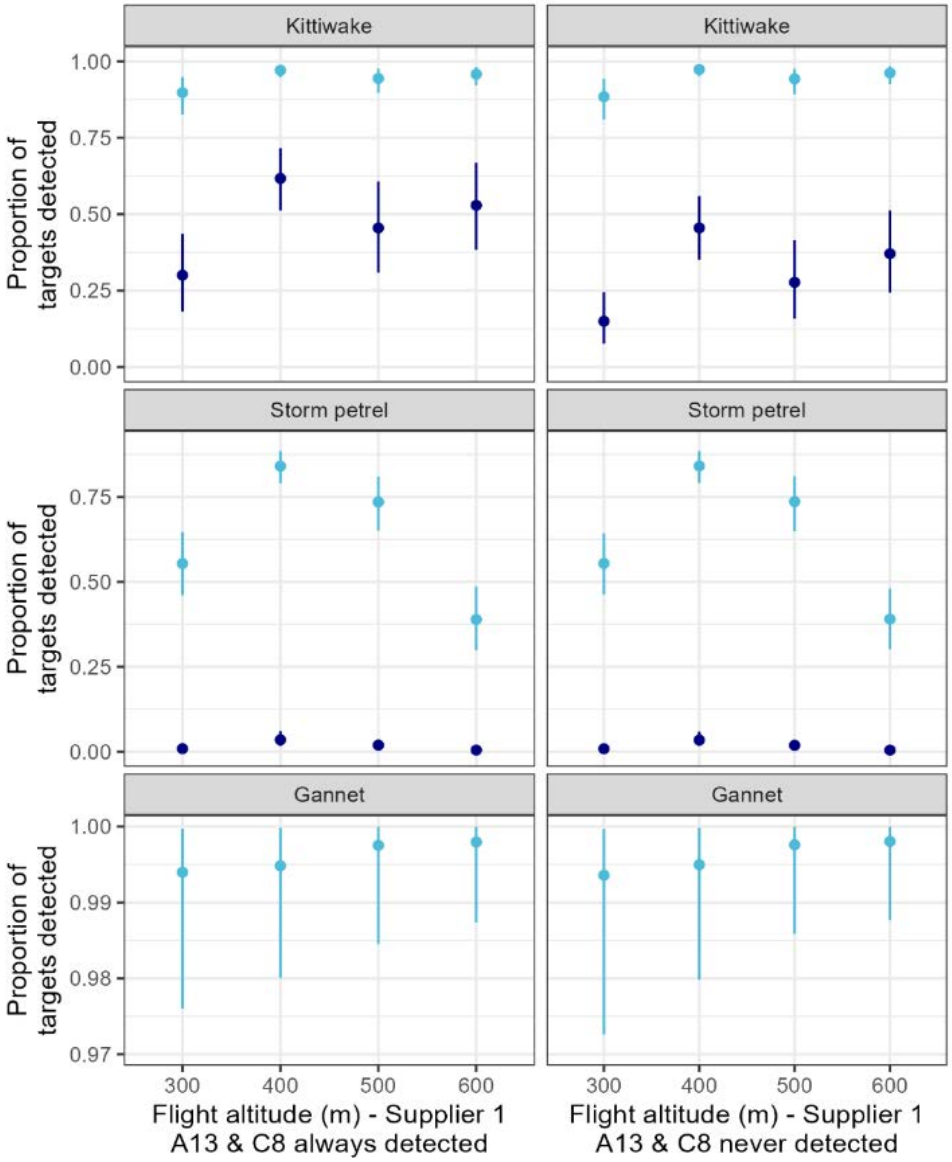
Therneau, T. 2024. A Package for Survival Analysis in R. R package version 3.8-3'. Available at: https://CRAN.R-project.org/package=survival.

Therneau, T.M. & Grambsch, P.M. 2000. *Modeling Survival Data: Extending the Cox Model*. Springer (Statistics for Biology and Health), New York.

Appendix 1: Influence of detection of targets A13 and C8

For two dark Kittiwake targets positioned on tripods (A13 and C8), some LiDAR hits detected the tripods rather than the targets themselves during the Supplier 1 survey. Genuine hits on these targets could not be distinguished from the tripod hits, so these targets were removed from the analysis as their true detection rates were unknown. To test the potential impact of this uncertainty on the model predictions, we reran the Supplier 1 detection models with these targets included. Two scenarios were run to test the difference between these targets being classified as 'detected' or 'not detected' in all flight passes. Classifying the targets as 'not detected' resulted in overall lower detection probability for dark Kittiwake compared to classifying the targets as 'detected' (Figure A1). However, trends of the influence of target species, colour and flight altitude on target detection rate were unaffected.

Figure A1: Conditional estimates (including 95% credible intervals) of proportion of targets detected for the Supplier 1 LiDAR survey, depending on whether targets A13 and C8 (both dark Kittiwake) were classified as 'always detected' or 'not detected', according to target species, target colour and flight altitude.



Appendix 2: Supplementary results

Onshore targets

The results in Section 3.1.1 for accuracy and precision of height estimation are shown for static targets (i.e. targets that were attached to fixtures rather than the washing line) or, for the Supplier 1 LiDAR data, had multiple LiDAR hits (Figures 8 to 10). For completeness, here we include results for suspended targets and those that only received one LiDAR hit.

For the Supplier 1 LiDAR data (Figure A2-1), credible intervals of predicted height were wider and precision of height measurement was lower (i.e. higher standard deviation) for the suspended targets compared to the static targets. For the Supplier 1 size-based DAS method (Figure A2-2), target type had little influence over accuracy and precision of target height, although credible intervals for precision were slightly wider for suspended targets. For the Supplier 2 LiDAR data (Figure A2-3), height estimates were less accurate and precise for suspended targets compared to static targets.

For the Supplier 1 LiDAR survey (Figure A2-4), number of LiDAR hits had no observable effect on the mean or precision of target height estimation.

Figure A2-1: Conditional estimates (including 95% credible intervals) of the predicted mean error (A) and precision (B) of estimated target height for the Supplier 1 LiDAR survey, according to target species, colour, flight altitude (m AGL) and target type (suspended=on washing line; static=attached to fixture). Results are shown for targets with multiple LiDAR hits.

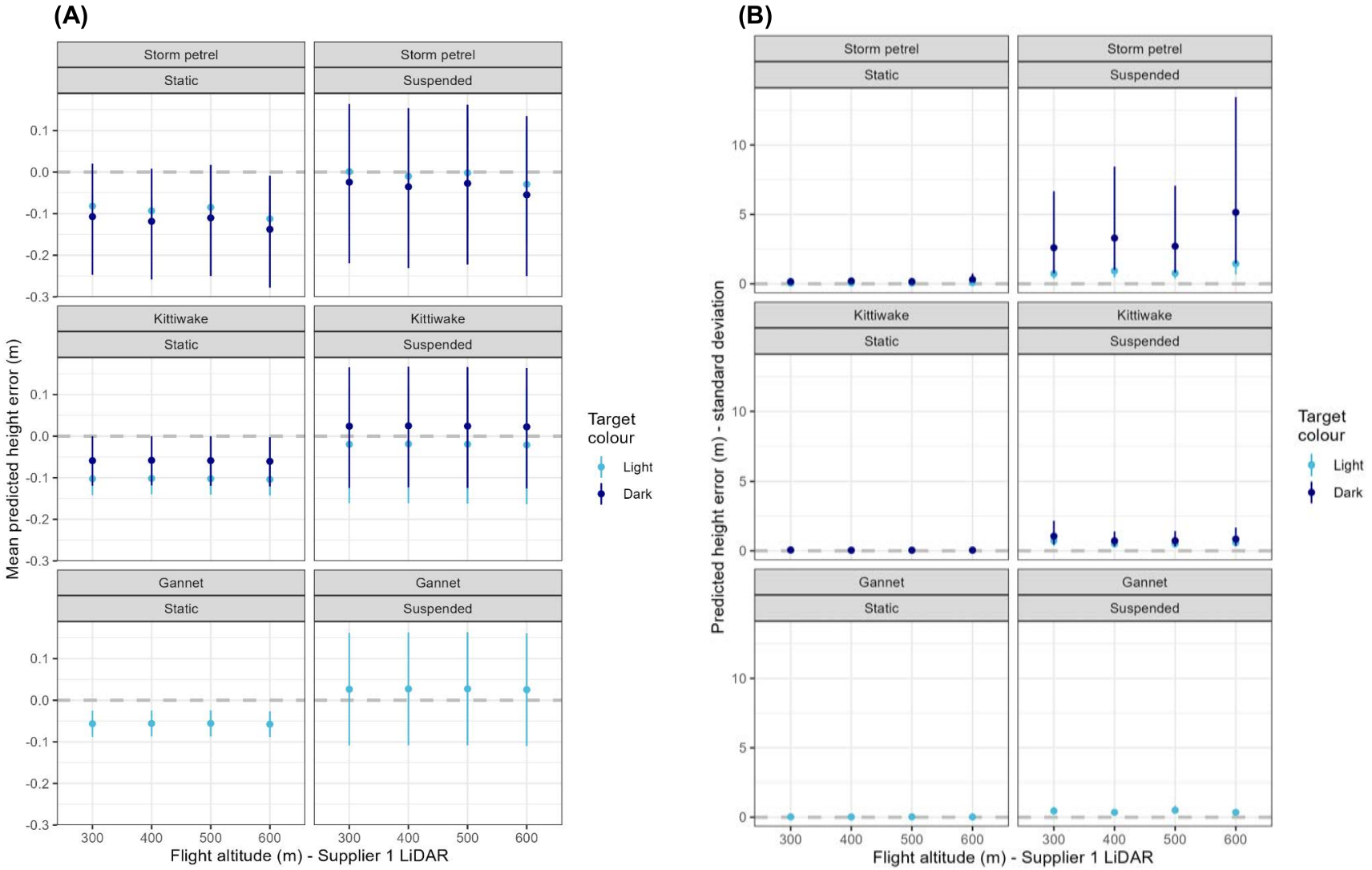


Figure A2-2: Conditional estimates (including 95% credible intervals) of the predicted mean error (A) and precision (B) of estimated target height for the Supplier 1 DAS survey, according to target species, colour, flight altitude (m AGL) and target type (suspended=on washing line; static=attached to fixture).

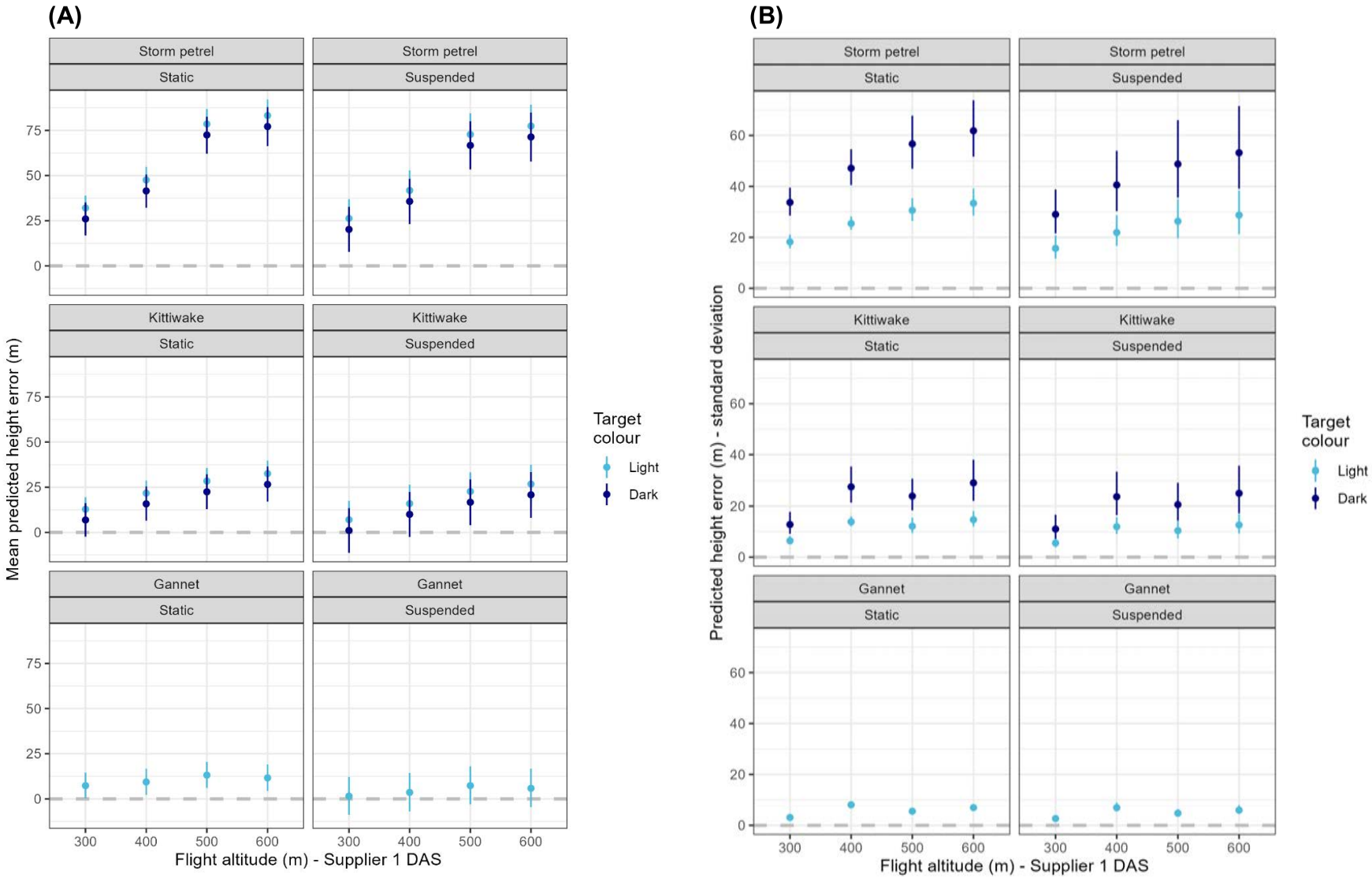


Figure A2-3: Conditional estimates (including 95% credible intervals) of the predicted mean error (A) and precision (B) of estimated target height for the Supplier 2 LiDAR survey, according to target species, colour, flight altitude (m AGL) and target type (suspended=on washing line; static=attached to fixture).

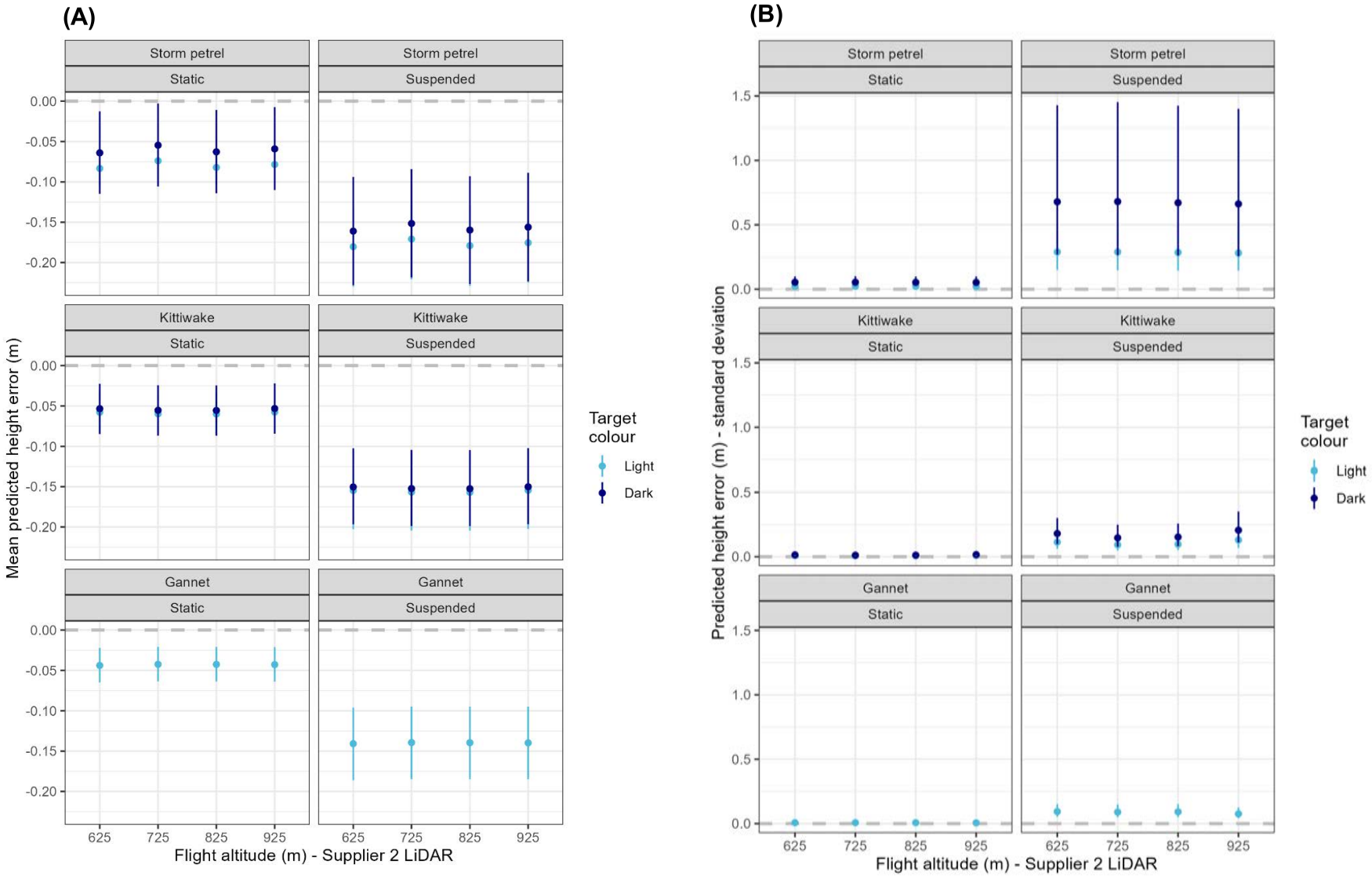
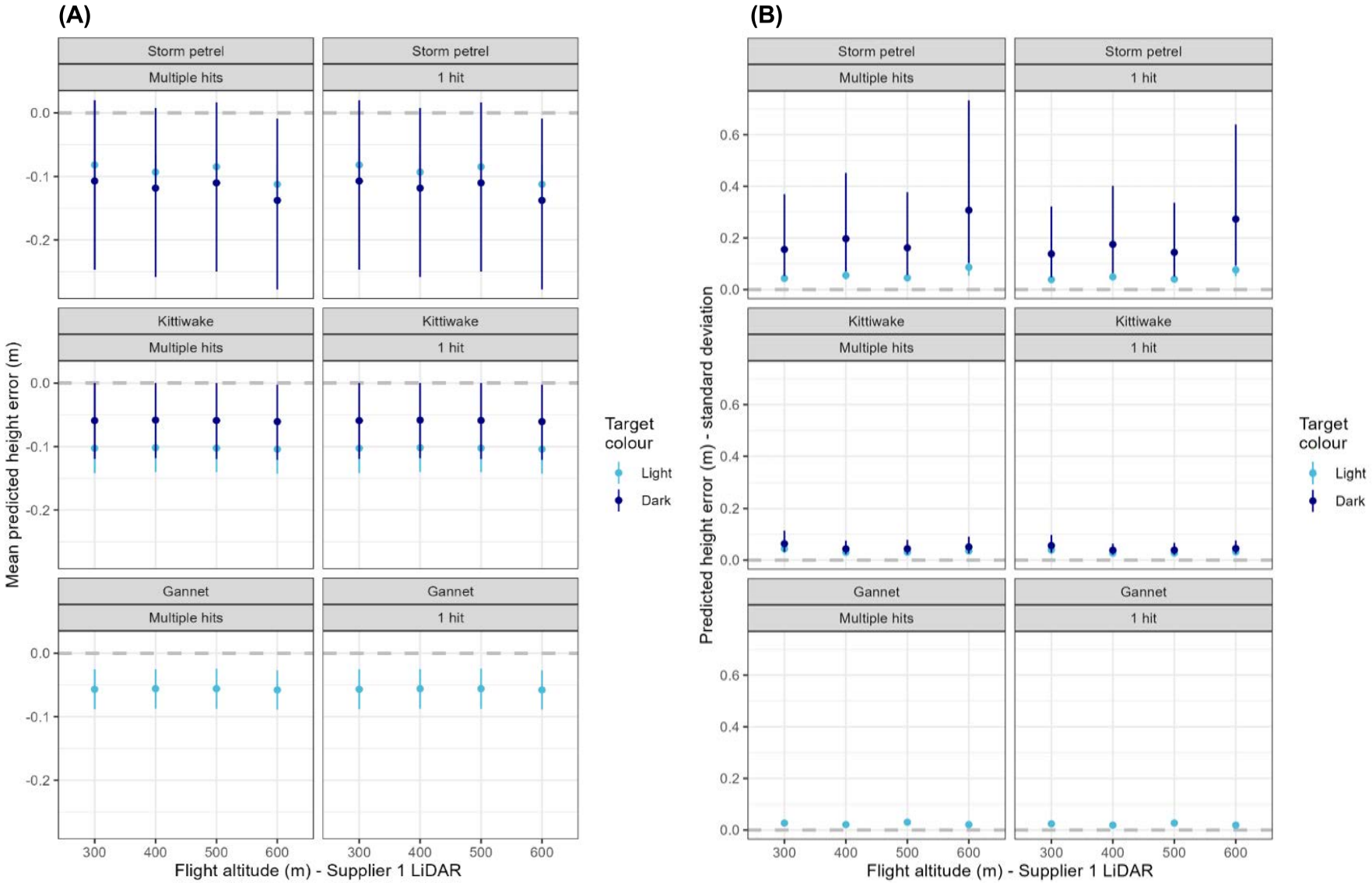


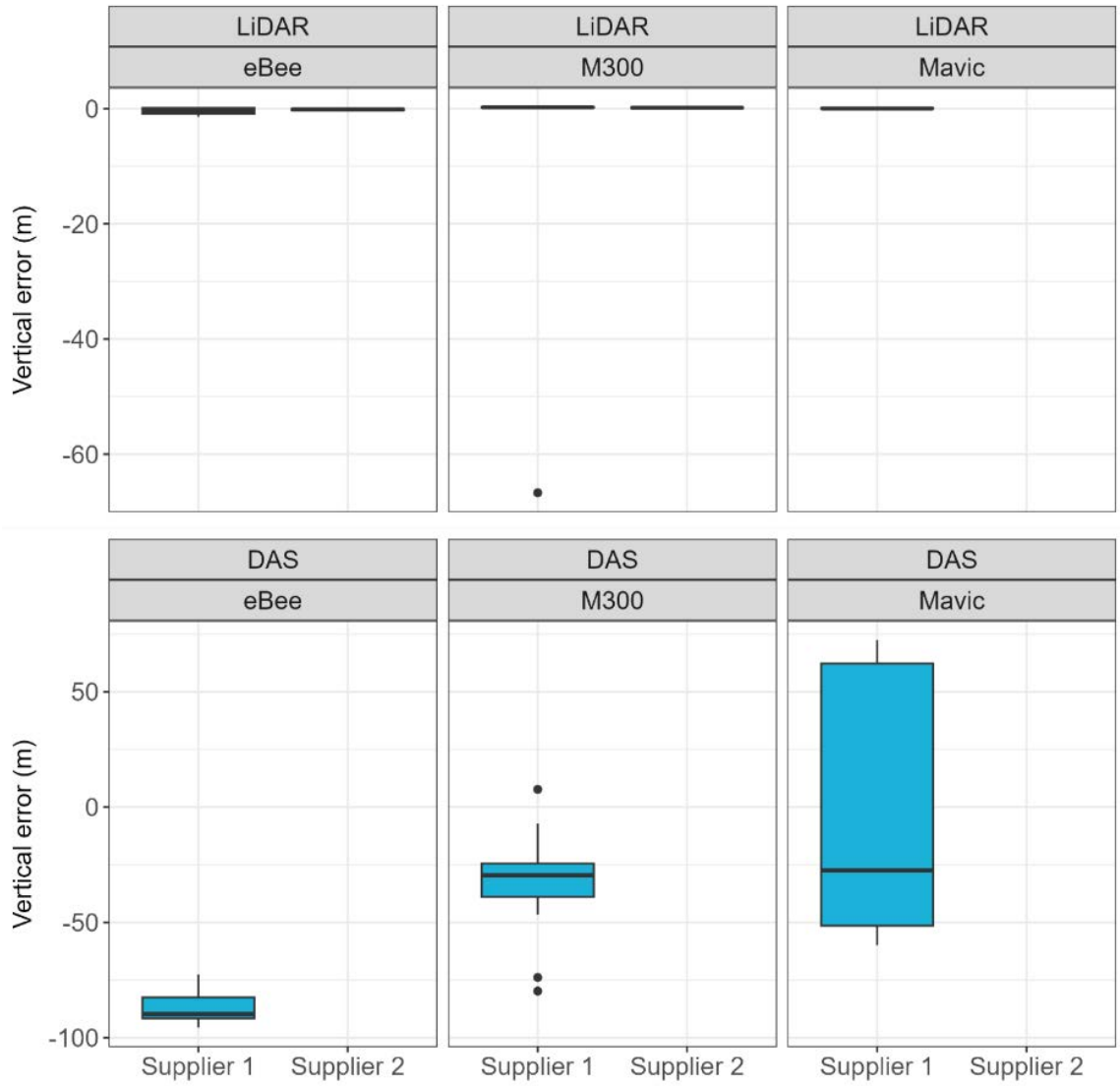
Figure A2-4: Conditional estimates (including 95% credible intervals) of the predicted mean error (A) and precision (B) of estimated target height for the Supplier 1 LiDAR survey, according to target species, colour, flight altitude (m AGL) and number of hits.



Onshore moving targets

For ease of visualisation, one observation with an extreme negative vertical error was removed from Figure 17. This observation was made at an aircraft flight height of 300 m which was erroneously configured. Figure A2-5 shows vertical error between LiDAR and DAS-estimated heights and known altitude of the drone with this outlier included.

Figure A2-5: Deviation in flight heights (m) between heights estimated by LiDAR or DAS and the known height of the drone, with extreme observation included. Note the y-axes differ between the rows.



Cover images, Philipp Boersch-Supan/BTO

Evaluating aerial LiDAR and other approaches to avian flight height measurement – ReSCUE Project Validation Study Report

The ReSCUE project investigates the accuracy and reliability of seabird flight height data to mitigate impacts of offshore wind farms on seabird populations. Through a set of trials, the projects set out to validate Light Detection and Ranging (LiDAR)-coupled Digital Aerial Surveys, and to a lesser extent size-based methods. Additional technologies evaluated in the trials included bird-borne Global Positioning System (GPS) telemetry tags and human-operated laser rangefinders.

We found that the comparative performance of the two LiDAR-coupled Digital Aerial Survey suppliers varied based on system differences and weather conditions during surveys, particularly in terms of detection rates from different sensors. LiDAR flight height measurements, however, proved to be accurate and precise for both suppliers, with measurement uncertainties on the scale of centimetres. In contrast, size-based DAS estimates had uncertainties on the order of tens of metres.

Findings from this study will directly inform the development of best practice guidance for seabird flight height data collection and analysis, supporting impact assessments for offshore wind farms while minimising ecological risks to seabird populations.

Suggested citation: Rhoades, J., Feather, A., Harwood, A., Banks, A. & Boersch-Supan, P.H. 2025. Evaluating aerial LiDAR and other approaches to avian flight height measurement – ReSCUE Project Validation Study Report. *BTO Research Report* **796**, BTO, Thetford, UK.

Offshore Wind Evidence + Change Programme

THE CROWN ESTATE

Reducing Seabird Collisions Using Evidence

NATURAL ENGLAND

ISBN 978-1-912642-95-3

9 781912 642953 >

BTO Birds Science People