Received: 28 February 2025 | Accepted: 31 January 2026

DOI: 10.1002/2688-8319.70238

BRITISH ECOLOGICAL SOCIETY Ecological Solutions and Evidence

RESEARCH ARTICLE

Implications of surveyor accuracy in bird flight height estimation for wind farm collision risk assessment

Nicola Largey | Hannah A. Edwards | Darío Fernández-Bellon | Susan Doyle

MKO Research, MKO, Galway, Ireland

Correspondence
Susan Doyle
Email: sdoyle@mkoireland.ie

Funding information
Natural Environment Research Council, Grant/Award Number: 22/RDD/881

Handling Editor: Alejandra Morán-Ordóñez

Abstract

1. The increasing anthropogenic use and modification of the airspace, particularly through the expansion of wind energy, has led to an increase in collision-related mortality of aerial wildlife. Environmental impact assessments rely on accurate bird flight height estimates to predict collision risk, yet these estimates are typically made by surveyors visually estimating flight height, with limited validation of their accuracy.
2. This study evaluates the accuracy of surveyor flight height estimates against measurements from an unmanned aerial vehicle (UAV) and investigates whether accuracy can be improved by the use of annotated maps (spatial cues) and/or prior survey experience.
3. Results indicate that there was generally a discrepancy between surveyor flight height estimates and the UAV measurements, particularly for inexperienced surveyors. Annotated maps reduced variability in estimates and improved accuracy, with the greatest benefit observed for inexperienced surveyors.
4. There was a discrepancy between the number of model predicted collisions based on surveyor estimates and UAV-based predictions, but this was reduced with the use of annotated maps and prior experience. Discrepancies in flight height estimation and subsequent discrepancies in collision predictions resulted in underestimating and overestimating mortality and translated into differences in impact assessment classifications, which could influence wind farm planning and mitigation decisions.
5. *Practical implication*. Overall, our results suggest that while annotated maps are beneficial, particularly for inexperienced surveyors, they are not a substitute for experience. Standardising survey methods through tools like annotated maps can improve data reliability, particularly for less experienced surveyors, ultimately strengthening impact assessments and wind farm management strategies. Beyond the value of spatial cues and surveyor experience, our work underlines the importance of quantifying sources of error in data collection for environmental impact assessments and consequences for subsequent management decisions.

KEYWORDS

annotated maps, collision risk model, ecological surveyor experience, environmental impact assessment, ornithological assessment, vantage point survey, wind energy

1 | INTRODUCTION

The growing anthropogenic use and modification of the airspace has led to increasing conflicts with aerial wildlife (Shamoun-Baranes et al., 2017; Vas et al., 2015). These include disruption of normal wildlife movement and collision-related mortality which can significantly affect species ecology and conservation (Steele & Weston, 2023; Vander-Zanden et al., 2024). In particular, the continued global growth of wind energy facilities to meet renewable energy targets has increased collision risk to bird species (Katzner et al., 2019). Accurate quantification of collision risk is essential to enable adequate assessment of impacts of new wind energy developments and design appropriate mitigation strategies. Bird flight height is a key parameter in the assessment of potential interactions with wind farms and the prediction of collision risk for birds. Flight height data can be generated by a variety of sources (e.g. Largey et al., 2021; Thaxter et al., 2015) but is predominantly collected by surveyors visually estimating bird height from the ground during surveys. An implicit assumption is that the surveyor's estimations of height are precise, thereby facilitating accurate collision risk predictions. However, surveyor height estimates can be subjective, and often, there is no assessment of their precision or accuracy (Thaxter et al., 2015). This uncertainty can have knock-on effects on the accuracy of analyses that use these data, such as collision risk predictions, and on ensuing development and management decisions.

Uncertainty in surveyor estimates can be attributed to a lack of information on the height of fixed objects in the landscape, for example trees and topographical features, as well as variation in bird morphology, and rapid changes in the altitude and speed of the bird (Garvin et al., 2011; Stantial & Cohen, 2015). Although flight height estimate accuracy can be improved with the use of sensor-based methods, such as radar and laser rangefinders (Becker et al., 2020; Harwood et al., 2018), these are generally not required by statutory bodies legislating wind energy developments. This may reflect the high equipment cost and technical expertise requirements, resulting in challenges for incorporating them into guidance and recommendations. There is thus a need for a means of improving surveyor flight height estimates that enables continuity with existing survey protocols, data requirements and guidance, while maintaining accessibility and affordability.

In this study we tested the accuracy of surveyor flight height estimates (hereafter 'flight height estimates') by comparing them to known flight heights of an unmanned aerial vehicle (UAV) and evaluated whether flight height estimate accuracy was improved by the use of spatially annotated maps (maps of a survey area annotated with the height and distance of surrounding landscape features; hereafter 'annotated maps') and/or by prior ornithological vantage point survey experience. We then examined how variation in flight height estimates influenced the predicted collision risk. We tested several hypotheses: (1) we expected that there would be a discrepancy between flight height estimates and the UAV sensor-based measurements; (2) we predicted that surveyors with prior experience would make more accurate flight height estimates than those with no experience; (3) we expected that information on spatial cues provided in annotated maps would improve flight height estimate accuracy; and (4) we expected collision risk model predictions to reflect the differences outlined in the previous three hypotheses.

2 | MATERIALS AND METHODS

2.1 | Data collection

Data were collected through an experimental trial designed to simulate a surveyor's experience during an ornithological vantage point survey. The simulation survey took place at Dangan Sportsground, Galway, Ireland (128480E 227663N), in September 2023. This space was chosen due to its accessibility and range of landscape features similar to real-world bird survey landscapes (e.g. trees, waterbody and topographical variation). The survey was conducted on one occasion over approximately 1.5 h, beginning in full daylight at 11:30 AM.

Prior to the survey, two annotated maps of the survey area were prepared: one showing a profile photograph of the area to be surveyed with annotated heights, and another showing aerial imagery of the survey area annotated with terrain and landscape height and horizontal distance bands from the observer point (see Supporting Information S1). Maps were developed using a combination of high-resolution landscape photographs and a three-dimensional digital surface model (DSM) of the survey area. The landscape photographs were taken with a camera (Cannon EOS 5D Mark III Full Frame Sensor with a 50 mm Lens; Canon, Tokyo, Japan) following the Landscape Institute Technical Guidance Note 06/19 (Landscape Institute, 2019). These were then annotated with the height (metres) of prominent landscape features (see below for height determination) using Adobe Photoshop (Version 25.0). The DSM was generated during a photogrammetry flight by a camera-mounted UAV (DJI Matrice 30T model with the DJI Pilot 2 application; DJI, Shenzhen, China). The camera captured a sequence of images at 5.5 cm/pixel ground geometrical resolution, with an 80% front overlap and a 70% side overlap. The accuracy of the DSM is reliant on the hovering accuracy of the UAV which can range from 0.1 to 1.5 m in the vertical axis and 0.3–1.5 m in the horizontal axis. The photogrammetric restitution, including the DSM generation, was performed with Agisoft Metashape Professional (Version 2.0.3 [Build 16960]). The elevation of the landscape features was determined using the photographs and DSM in Autodesk 3D Studio Max (Version 26 [Build 26.1.0.2270]). The resulting DSM was overlayed on an aerial map of the survey site. This same map was also annotated with distance bands from the planned participant position in 50-m intervals generated in QGIS (Version 3.32).

Participants from professional, amateur and novice bird survey backgrounds were invited to the survey through word-of-mouth, social media outlets and engagement with relevant organisations (e.g. birdwatching groups). On arrival to the survey site, each participant was assigned a unique identifier number, and their prior ornithological vantage point survey experience was recorded ('prior experience' or 'no experience'). Each participant was provided with a data collection form to enter their flight height and distance estimates and the annotated maps described above. Participants did not have access to optical equipment (e.g. binoculars or telescopes).

During the survey, a UAV with an altimeter (same model described above) was used as a surrogate for a bird with known flight paths and heights. Participants were asked to estimate its distance and height to the nearest metre during two stages. During Stage 1, participants were not provided annotated maps and estimated flight heights based solely on their own judgement. During Stage 2, participants were provided the annotated maps to aid their estimates (both the annotated landscape photograph with feature heights and aerial map with distance bands). Both stages comprised three flight sequences at different distances from the participants to the UAV (50, 150 and 300 m). At the beginning of each sequence, the participants were asked to estimate and record the distance of the UAV. To enable comparison between surveyor estimates and sensor-based measures, surveyors recorded the distance between the UAV and its operator, located among the participants. After the distance estimate was complete, the UAV was flown through 12 randomised heights between 10 and 120 m (remaining at the same distance before moving on to the next sequence). At each height, the UAV hovered for 30 s, allowing the participants to record their flight height estimate in their data collection form. At the end of the sequence, the UAV moved to the next sequence in the next distance band. Separate data collection forms were provided for each stage (without and with annotated maps) to prevent participants referring back to or changing their estimates from the previous stage. Participants had no prior knowledge of the UAV height or distance before either stage, and no knowledge of the annotated maps prior to Stage 2. At the end of the survey, all data forms were collected and digitised for analysis.

2.2 | Analysis

Using data collected from the survey, we defined height estimation error as the participant's estimated height minus the true UAV height. To characterise this error, we calculated four descriptive metrics. These were bias (mean signed difference between heights which indicates whether participants systematically over- or underestimated height), mean absolute error (mean size of the error to provide a measure of typical deviation), root mean square error (mean size of the error but with a greater weight to large deviations to show if extreme values strongly affected accuracy) and median absolute error (to show the central tendency of error size with less sensitivity to outliers). We also expressed bias and mean absolute error as percentages of mean UAV height to contextualise errors relative to the scale of the height. To assess the effects of experience, these were calculated across all participants (mixed experience), for participants with prior ornithological vantage point survey experience (prior experience), and for participants with no ornithological vantage point survey experience (no experience).

We tested the discrepancy between the true height measured by the UAV and flight heights estimated by participants, and the effect of annotated maps and experience on this discrepancy. Vantage point survey methodology recommends that surveyors collect bird flight height data in pre-defined broad height bands rather than to the exact metre to account for observer error (Scottish Natural Heritage, 2017). To quantify this source of observer error, we compared results using the height data to the exact metre to those using the height data categorised into broad height bands. The broad height bands chosen for this study were 0–30 m, 31–90 m and 91 m+, to align with permissible UAV flight heights. This allowed us to assess the accuracy of using exact measurements compared with the height band approach recommended in standard survey practices. All analyses were conducted in the R language and programming environment (Version: 4.3.2; R Core Team, 2024).

The discrepancy between the true UAV height and flight height estimates was tested using two methods depending on data structure. Data to the exact metre were tested using intraclass correlation coefficients (ICC) and data categorised into broad height bands were tested using Cohen's kappa.

Intraclass correlation coefficients (Bartko, 1966) from the package 'irr' (Version 0.84.1; Gamer & Lemon, 2019) represent the agreement between measurements. Koo and Li (2016) provide the following scale for interpretation: $0 < \mathrm{ICC} \leq 0.50$ being 'poor', $0.50 < \mathrm{ICC} \leq 0.75$ being 'moderate', $0.75 < \mathrm{ICC} \leq 0.90$ being 'good' and $0.90 < \mathrm{ICC}$ being 'excellent'. A two-way model measured absolute agreement between the UAV and the average of the participant's estimates, with the significance of the agreement based on an *F*-test.

Weighted Cohen's kappa (κ_w) comparison of agreement (Cohen, 1968; Watson & Petrie, 2010) from the package 'psych' (Version 2.4.2; Revelle, 2024) represents the chance-corrected proportional agreement between measurements based on contingency tables and is thus more powerful than a simple percentage agreement. The interpretation of κ values is not formalised but is generally accepted as follows: $\kappa \leq 0$ being 'poor', $0 < \kappa \leq 0.2$ being 'slight', $0.2 < \kappa \leq 0.4$ being 'fair', $0.4 < \kappa \leq 0.6$ being 'moderate', $0.6 < \kappa \leq 0.8$ being 'substantial' and $0.8 < \kappa \leq 1$ being 'almost perfect' (Landis & Koch, 1977). κ_w was estimated because it considers the magnitude of the difference (e.g. if participant's estimates differed to the UAV by one band versus two bands). A linear weighting structure was used, with all levels of difference weighted equally.

The effect of annotated maps on the discrepancy between the true UAV height and flight height estimates was tested using two methods depending on data structure. Data to the exact metre were tested using a permutation test and data categorised into broad height bands were tested using a Fisher's exact test.

The permutation test assessed whether the variability in the discrepancy differed between Stage 1 (no annotated maps) and Stage 2 (with annotated maps), as a reduction in variability indicates more accurate flight height estimates. Within the data, the categories 'no annotated map' and 'with annotated map' were randomly reassigned to each observation to simulate a scenario in which the observed difference in variability was purely random.

Variances were then recalculated for the two groups based on the permuted data, and the mean difference in variances was calculated. A total of 1000 iterations were run to create a null distribution of differences in variance representing no effect of map use. The observed mean difference was then compared against this null distribution to calculate a two-tailed *p*-value representing how extreme the observed difference was compared with the null distribution. A *p*-value ≤0.05 indicates that the observed difference is unlikely to occur by chance.

The Fisher's exact test contingency analysis assessed whether the difference in the discrepancy was associated with the annotated maps for data categorised into broad flight height bands. A *p*-value ≤0.05 was considered statistically significant. Cramer's V was used to assess the strength of the association using the package 'vcd' (Version 1.4–12; Meyer et al., 2023). The interpretation of V is subjective (Akoglu, 2018; Lee, 2016) but given the low degrees of freedom in this analysis is conservatively: $0 < V \le 0.10$ being 'negligible', $0.10 < V \le 0.30$ being 'weak', $0.30 < V \le 0.50$ being 'moderate' and $V > 0.50$ being 'strong'.

2.3 | Collision risk model

A mock Collision Risk Model following Band et al. (2007) methodology was conducted using the data collected during the survey to understand how flight height estimate errors affect collision risk predictions. For each participant, two collision risk models were conducted using their estimates from Stage 1 (no annotated maps) and their estimates from Stage 2 (with annotated maps) and then compared with a collision risk model based on the corresponding UAV measurements. Collision risk models first calculate the predicted number of transits of a bird across a risk zone using the total duration the bird spent at 'risk' height bands, which was 31–90 m in our scenario. The model then calculates a collision rate based on the number of transits, the ecology of the species in question and the specifications of the facility (i.e. turbine dimensions, surveyor viewshed area, survey effort, species flight speed and style, latitudinal photoperiod, species biometrics and species collision avoidance rate). These remaining model parameters were set to a default value according to a mock facility and bird species, designated as common buzzard *Buteo buteo* (see Supporting Information S2 for all parameter values). To provide a more intuitive result, the collision rate was multiplied by the number of years a facility would normally operate for (25 years) to provide the predicted number of bird collisions.

The discrepancy between the predicted number of bird collisions from participant and corresponding UAV data were then calculated. The effect of annotated maps on this discrepancy was tested using a Wilcoxon signed-rank test for paired analysis (the discrepancy in Stage 1 [no annotated maps] compared with the discrepancy in Stage 2 [with annotated maps]). This test was conducted for each participant experience level group to compare the effect for different experience levels. As above, *p*-values ≤0.05 were considered statistically significant and the *r*-value was used to assess the effect size. There is no formal interpretation of r, but it is recommended (following Tomczak & Tomczak, 2014) as: $0.10 < 0.3$ being 'small', $0.30 < 0.5$ being 'moderate' and ≥0.5 being 'large'.

In environmental impact assessments, the impact of a development on bird populations is typically measured by the percentage change in mortality from baseline conditions, such as the percentage increase in background mortality that might result from collisions with wind turbines. Guidance from Percival (2003) categorises the magnitude of impacts as follows: a 'low' magnitude impact corresponds to a 1%–5% change in mortality from baseline, 'medium' as 5%–20%, 'high' as 20%–80% and 'very high' as >80%. To demonstrate the practical implications of errors in flight height estimates, participant experience and the use of annotated maps, we used the results of the mock collision risk model to determine the magnitude of the impact of our mock wind energy facility on mortality levels in the local buzzard population. First, we calculated the number of local buzzards that would die from natural causes using a background mortality rate of 0.1 (Kenward et al., 2000). This calculation was done for a range of baseline population sizes (10–1000 buzzards) to examine how variations in collision predictions could influence the impact magnitude and, ultimately, development decisions. From this range, we identified three specific population sizes (550, 140 and 35 buzzards) where differences in impact magnitudes, and thus development decisions, may become apparent. Second, we calculated the percentage increase in mortality by dividing the mean buzzard collision predictions for each experience group by the buzzard background mortality and expressing the result as a percentage. The corresponding impact magnitude categories were then assigned based on the thresholds defined by Percival (2003).

3 | RESULTS

A total of 35 participants took part in the survey, 11 of whom had prior ornithological vantage point survey experience. A total of 2386 flight height estimates were collected for analysis (1231 during Stage 1 and 1155 during Stage 2). Descriptive metrics of height estimation error are presented in Table 1. Inexperienced participants without annotated maps generally overestimated heights, producing large mean absolute errors (i.e. greater deviation from the true UAV height). In contrast, experienced participants without maps showed little systematic bias and smaller mean absolute errors. Providing annotated maps reduced errors in both groups, although inexperienced participants still slightly overestimated heights, but with markedly lower mean absolute error.

3.1 | Accuracy of flight heights and ornithological experience

When data were to the metre, flight height estimates had a low accuracy (ICC[A, 2] = 0.321, 95% confidence intervals [CI] 0.254–0.381 or 'poor agreement'). Flight height estimation accuracy was higher for participants with prior experience (ICC[A, 2] = 0.772, CI

TABLE 1 Summary of height estimation error (in metres) according to use of annotated maps and experience.

Map	Participant group	*N*	*B*	Mean AE	RMSE	Median AE	%*B*	%MAE
No map	Mixed experience	1231	37.86	63.67	146.52	20	59.48	100.02
	Prior experienced	386	−1.91	27.56	37.86	20	−3.01	43.38
	No experience	845	56.03	80.16	174.99	25	87.95	125.83
With map	Mixed experience	1155	4.40	26.20	46.07	15	7.22	43.02
	Prior experienced	363	−11.48	19.40	26.66	13	−18.84	31.85
	No experience	792	11.67	29.32	52.62	15	19.16	48.14

Note: *N* indicates the number of estimates in each condition. Bias (*B*), mean absolute error (mean AE), root mean square error (RMSE) and median absolute error (median AE) are presented, along with bias and mean absolute error expressed as a percentage of mean unmanned aerial vehicle (UAV) height (%*B* and %MAE, respectively).

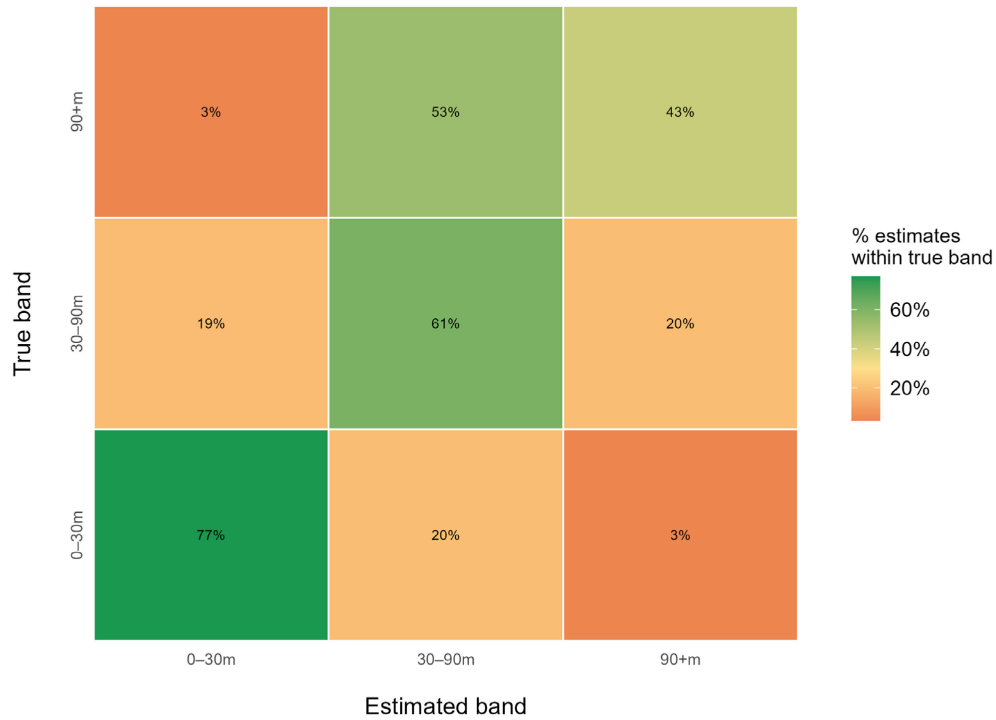

FIGURE 1 Agreement between participant height estimates and true height when data are categorised into broad bands. Shading indicates the percentage of correct estimates, ranging from lightest (100% of estimates in correct band) to darkest (0% of estimates in correct band).

0.730–0.807 or 'good agreement') than for participants with no experience (ICC[A, 2] = 0.279, CI 0.184–0.362 or 'poor agreement').

When data were categorised into broad height bands, agreement increased, indicating higher accuracy (κ_w = 0.58, CI 0.55–0.61 or 'moderate agreement'). As with data to the metre, accuracy was higher for participants with prior experience (κ_w = 0.62, CI 0.58–0.67 or 'substantial agreement') than participants with no experience (κ_w = 0.56, CI 0.53–0.60 or 'moderate agreement').

3.2 | Effect of annotated maps on discrepancy

The significance of the variability in the discrepancy between the true UAV height and the flight height estimates differed depending on the use of annotated maps. The permutation test revealed that the use of annotated maps significantly reduced variability (Figure 1; Table 2). The observed difference in variability was not significant for participants with prior experience. However, for participants with no prior experience, the use of annotated maps significantly reduced variability.

When data were considered in broad height bands, the overall discrepancy between the true UAV height band and participant estimated height bands was lower when using annotated maps (number of pairs = 2, V = 0.091, p < 0.001), and the magnitude of this effect was 'negligible' (Figure 2). Again, this varied according to experience group, with participants with no previous experience (n = 24) showing a significant increase in accuracy with the use of maps (number of pairs = 2, V = 0.095, p = 0.001), while for participants with prior experience (n = 11), this change was non-significant (number of pairs = 2, V = 0.080, p = 0.12).

Participant group	Observed difference	Permuted differences	CI	*p*-value
Mixed experience	6494.97	−109.88	−6224.885925.84	**0.03**
Prior experience	402.32	−19.46	−539.77528.81	0.20
No experience	9287.44	−93.60	−9084.798817.31	**0.04**

TABLE 2 Difference in variability (as measured by the variance in the discrepancy between the true unmanned aerial vehicle (UAV) height and estimated participant height) between Stage 1 (no annotated map) and Stage 2 (with annotated map).

Note: The mean observed difference in variability, mean permuted differences and 95% confidence interval (CI) based on the permutation distribution are shown, along with the *p*-values. A significant *p*-value indicates that the observed difference is extreme relative to the distribution of permuted differences. Significant *p*-values are in bold.

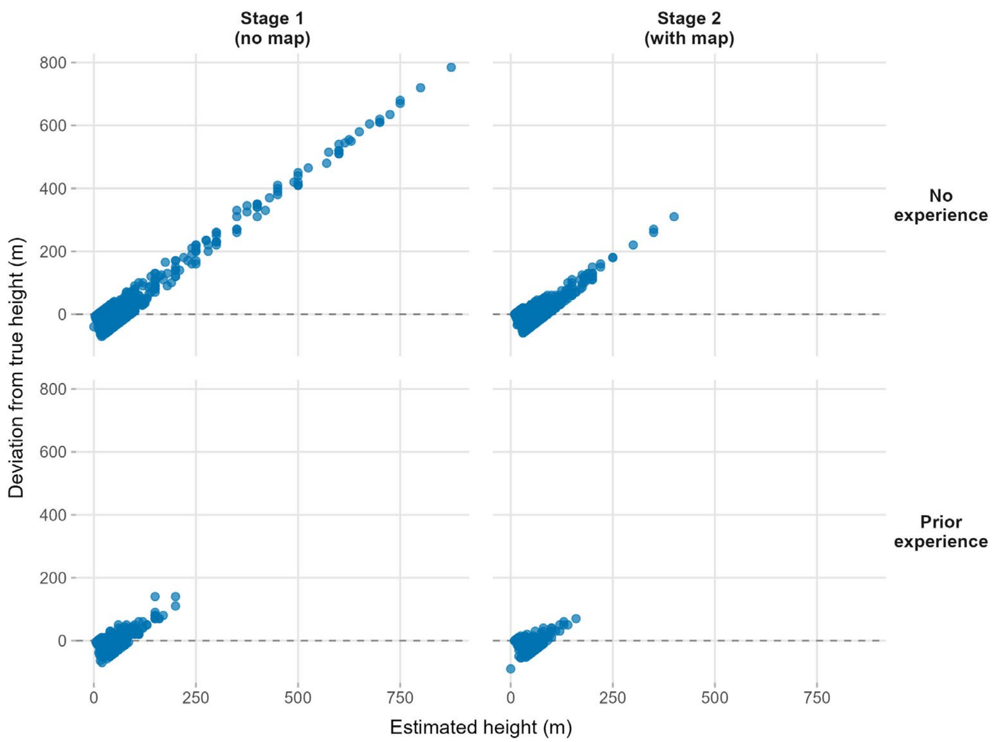

FIGURE 2 Deviation from true height (estimated height − true height) plotted against estimated height. Each point represents an individual estimate. The dashed horizontal line at zero represents perfect agreement (positive values indicate overestimation; negative values indicate underestimation.

3.3 | Collision risk model

The mock collision risk model showed that participant-based data generally predicted fewer collisions than UAV-based data for the 25-year period (Figure 3). The model based on UAV data, which predicted 3.32 buzzard collisions, is compared below to the different participant-based models (Table 3). Data from participants without annotated maps predicted a mean of 2.60 ± 0.20 collisions (0.72 fewer), while data from participants using annotated maps resulted in a significantly more accurate predicted mean of 3.34 ± 0.17 collisions (0.02 more). However, the difference in the discrepancy varied according to experience level. In the group with no experience, data from flight height estimates without maps resulted in a predicted mean of 2.46 ± 0.27 collisions (0.86 fewer) which significantly improved with the use of annotated maps to a predicted mean of 3.17 ± 0.21 collisions (0.16 fewer). In contrast, for participants with prior survey experience, the absolute discrepancy relative to the UAV prediction was similar with and without maps (3.72 ± 0.25 collisions [0.40 more] versus 2.90 ± 0.26 collisions [0.42 fewer]), although the within-participant map treatment effect (a directional shift from over- to underestimation) was statistically significant.

Our results indicated that the impact of the mock wind farm on a baseline population of 550 buzzards was categorised as 'medium' when using UAV derived heights (6.04% increase in mortality; Supporting Information S2). Data collected by participants with prior experience and by inexperienced participants using annotated maps slightly under- or overestimated the increase in mortality (5.27%–6.76%) but was still categorised as a 'medium' impact on the population. However, data from inexperienced participants without annotated maps further underestimated the change in mortality and resulted in the impact being categorised as 'low' (4.47% increase in mortality). This pattern of participant-based data underestimating the increase in mortality, and data from inexperienced participants

BRITISH ECOLOGICAL SOCIETY Ecological Solutions and Evidence

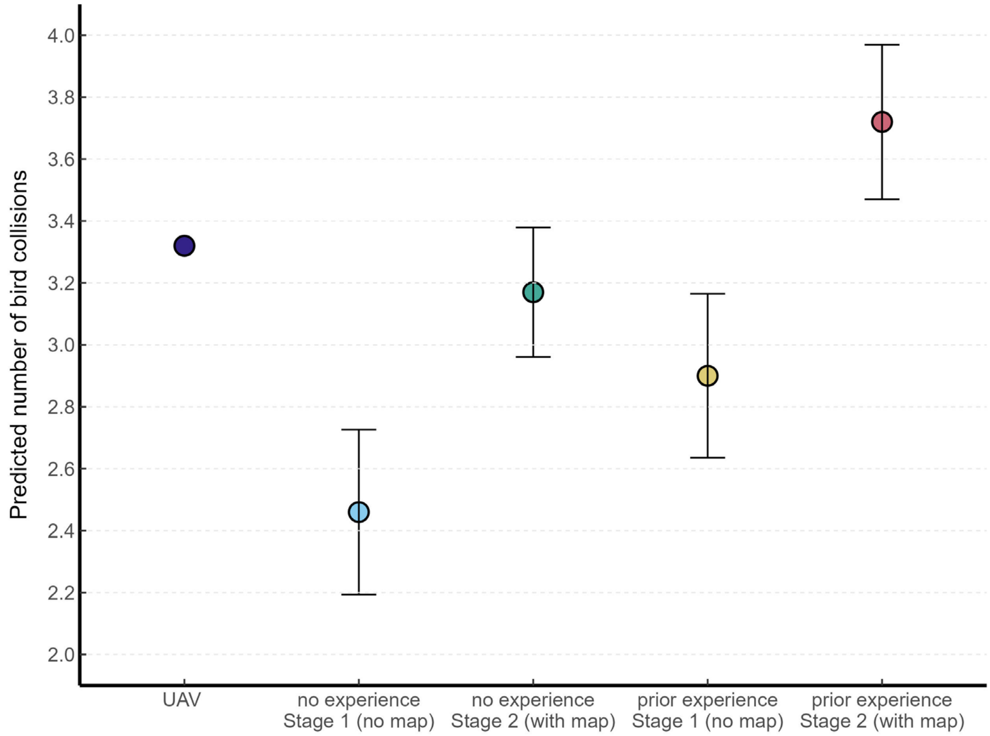

FIGURE 3 Mock collision risk model results. The number of bird collisions over the lifetime of the facility predicted from data collected by the sensor-based unmanned aerial vehicle (UAV) compared with the predictions based on participant flight height estimates. The mean prediction and standard error are presented for participant data collected with and without annotated maps and for the different experience levels (no experience and prior experience).

TABLE 3 Difference in the discrepancy between the unmanned aerial vehicle (UAV) and participant derived predicted number of bird collisions between Stage 1 (no annotated maps) and Stage 2 (with annotated maps) of the survey.

Participant group	*n*	Discrepancy no map	Discrepancy with map	Magnitude	Rating	*p*-value
Mixed experience	35	−0.72	0.02	0.44	Moderate	**<0.001**
Prior experience	11	−0.42	0.40	0.52	Large	**0.01**
No experience	24	−0.86	−0.16	0.39	Moderate	**0.01**

Note: The number of pairs (*n*), the average discrepancy in the number of bird collisions between the UAV and participants in Stage 1 and Stage 2, the magnitude of the effect (*r*), magnitude rating interpretation and significance (*p*-value) are presented for the three experience level groups.

without annotated maps resulting in a lower impact category were consistent for baseline population sizes of 140 and 35 buzzards.

4 | DISCUSSION

We used an experimental approach to test the effect of spatial cues (annotated maps) and experience on flight height estimation accuracy and variability. We also quantified the effect of using broad height bands instead of exact flight heights on accuracy. Our results indicate that the accuracy of flight height estimates was influenced by both surveyor experience and the use of spatially annotated maps, although overall there was a discrepancy between surveyor flight height estimates and the UAV sensor-based measurements (Hypothesis 1). In line with our predictions, participants with prior ornithological vantage point experience were more accurate than those with no experience (Hypothesis 2). This reflects previous work which has shown that surveyor experience improves the accuracy of distance measurements (Baird & Burkhart, 2000). Flight height estimation was generally less accurate and more variable when surveyors did not have access to annotated maps (Hypothesis 3), with surveyors often estimating above the true UAV height. The absence of familiar spatial features has been similarly linked to inaccurate height estimation in simulated environments (Oravetz, 2009; Park et al., 2021), while a tendency for people to over-estimate the height of objects was described by Yang et al. (1999). While annotated maps improved the accuracy and reduced the variability of flight height estimates for all participants, this effect was most apparent for surveyors with no prior experience and was limited for participants with prior experience. Prior survey experience may provide participants with a frame of reference or ability to account for changes in the perceived size of the UAV, leading to improved flight height estimates. These differences in accuracy and variability stemming from the use of annotated maps, particularly for inexperienced surveyors, had a significant effect on final bird collision predictions. Collision prediction accuracy reflected the improvements in flight height estimate accuracy due to surveyor experience and the use of annotated maps (Hypothesis 4).

Experience plays a crucial role in survey accuracy. This has been shown for estimation of quantities, such as abundance (Johnston et al., 2018; Williams et al., 2006), distance (Baird & Burkhart, 2000; Sunde & Jessen, 2013) and height (Butt et al., 2013). In some cases, additional training of surveyors may help reduce the discrepancy in estimated quantities (Garel et al., 2005; Gibson et al., 1955; Gibson & Bergman, 1954). Our findings suggest that in the case of height estimation for ornithological surveys, experienced surveyors may intuitively use landscape features for reference, resulting in improved flight height estimates and in turn collision risk predictions. This likely acts as an equivalent of the annotated maps that we provided. However, as these references themselves rely on estimations of landscape object heights, annotated maps may also be beneficial for experienced surveyors.

Annotated maps significantly reduced the variability of exact flight height estimate discrepancy for surveyors with no experience. Similarly, when flight heights were categorised into height bands, the use of annotated maps significantly improved their accuracy. While these same patterns held for surveyors with prior experience, the relationships were weaker and/or non-significant. This suggests that the use of the annotated maps is more impactful for inexperienced surveyors. The frame of reference provided by the annotated maps helps the surveyor place flight height estimates within the broad predefined bands used in field data collection. Training has been found to improve the accuracy of distance estimates (Gibson et al., 1955; Gibson & Bergman, 1954), and our findings indicate that providing annotated maps may be a useful tool to accelerate the learning process and enable surveyors to error check their flight height estimates compared with known object heights. Nonetheless, while annotated maps can enhance the accuracy of inexperienced surveyors' flight height estimates and help bridge the experience gap, they should not be seen as a substitute for experience.

Collision rate predictions in this study indicate that differences in flight height estimates (no prior experience versus prior experience, without annotated maps versus with annotated maps) can also impact collision risk model predictions. As expected, the use of annotated maps (particularly for inexperienced surveyors) resulted in more accurate bird collision predictions that were closer to those estimated from the UAV data. Participants with no experience and no maps were the least accurate, underestimating collisions by 0.86 compared with the UAV-based predictions. However, using maps significantly improved their accuracy, reducing the discrepancy to 0.16 fewer collisions than the UAV data. While modest in absolute terms, these differences can have important practical implications for environmental impact assessments, where accurate collision risk predictions are essential for estimating the potential impact of new developments (Cramp et al., 1988). Following the standard impact assessment processes, we used these predictions to determine the impact categories (i.e. 'low', 'medium' and 'high' magnitude) of our mock wind energy facility on the mortality levels of a hypothetical buzzard population. Regardless of the size that the buzzard population was set at, data from participants generally underestimated the increased mortality to buzzards caused by the hypothetical wind turbines. In particular, data from participants with no experience and no annotated maps consistently resulted in lower impact categories than those based on UAV data. However, when these participants had access to annotated maps, their derived impact categories aligned with those of experienced participants and UAV predictions. These differences have tangible real-world implications, as the impact category calculated in the environmental impact assessment process determines whether mitigation measures will be proposed for a development. Although our study focused on observer-derived error, we note that the use of fixed thresholds within collision risk models can amplify small differences in input values, occasionally leading to disproportionate changes in impact categorisation. To address this, recent studies have recommended uncertainty-aware approaches, such as Monte Carlo CRM runs (Masden, 2015), Bayesian analyses (New et al., 2015) and population models incorporating stochasticity and demographic trends (Horswill et al., 2022; Schippers et al., 2020).

Our findings can be easily integrated into current standard survey protocols recommended by statutory bodies (e.g. Scottish Natural Heritage, 2017). While sensor-based tools may offer quantifiable accuracy and precision for measurements over visual estimates (e.g. Becker et al., 2020; Harwood et al., 2018), logistical considerations, such as accessibility and expense often prohibit their use (e.g. Largey et al., 2021). Furthermore, in the terrestrial/onshore context, the data collected from such tools are still commonly analysed in a banded flight height model structure similar to data collected from visual estimates (Band et al., 2007; and more recently, Band, 2024). Therefore, increasing the accuracy and precision of the data with expensive sensor-based methods without adjusting the model to incorporate such data may prove to be a sub-optimal trade-off between accuracy and cost. While advancements in collision risk modelling are being developed and applied offshore (Johnston et al., 2014; McGregor et al., 2018), such models are not yet typically applied to onshore developments. Annotated maps may therefore represent a viable means of improving surveyor flight height estimates while minimising cost.

While our findings underline the value of experimental approaches to test the accuracy of surveyor flight height estimates, the results should be considered within the appropriate context. Firstly, the UAV hovered at pre-defined flight heights, which does not reflect the behaviour of most bird species. Secondly, the level of experience from participants was a binary value (prior experience versus no experience) rather than quantified (e.g. number of years of experience). Further research may help reveal the experience threshold beyond which the benefit of annotated maps becomes negligible. Thirdly, using annotated maps requires sufficient landscape features for height references, therefore the usefulness of annotated maps may vary across landscapes. Finally, it has been shown that height estimation is more difficult with smaller targets (e.g. Harwood

BRITISH ECOLOGICAL SOCIETY Ecological Solutions and Evidence

et al., 2018). The dimensions of the UAV (47 cm × 58.5 cm) in this study represented medium-to-large bird species surveyors would encounter. Future consideration could be given to experimental designs with a range of UAV dimensions.

4.1 | Recommendations

The variability in the accuracy of flight height estimates highlights the need for reliable and standardised methods to determine bird flight heights. Our results suggest that annotated maps can significantly reduce variability, improving accuracy and subsequently bird collision predictions, particularly for inexperienced surveyors, but they should not be considered as a substitute for experience. Annotated maps could be used during training to enhance surveyors' skills and in the field to reduce variation in flight height estimates between surveyors of different experience levels. However, prioritising experienced surveyors is crucial, particularly in areas lacking landscape features, to help minimise errors. Ultimately, the routine use of annotated maps as a standardisation tool can enhance and complement, but not replace, the expertise of experienced surveyors. Given the variability in collision predictions described here, appropriate reporting of collision risk models can provide insight into associated uncertainties. To ensure transparency and reliability of impact assessments, collision risk models should clearly state how data were obtained (including surveyor experience and use of tools like annotated maps) and where possible include measures of uncertainty by quantifying surveyor accuracy. Our work underscores not only the value of experience and the use of spatial cues to improve the accuracy and confidence of collision predictions, but also the importance of testing and evaluating the accuracy of measurements which are subject to observer error and are ultimately used for management and planning decision-making.

AUTHOR CONTRIBUTIONS

Nicola Largey and Susan Doyle conceived the study, designed the methodology, collected the data and conducted the statistical analysis. Nicola Largey led the writing of the manuscript and finalised it together with Susan Doyle and Hannah A. Edwards. Susan Doyle, Hannah A. Edwards and Darío Fernández-Bellon reviewed the manuscript. All authors critically contributed to the drafts and gave their final approval for the publication.

ACKNOWLEDGEMENTS

The authors would like to thank Pietro Salvagno, Jack Workman and James Newell for their valuable contribution to the data collection process for this study. We are grateful to all the participants who volunteered their time to the fieldwork, as well as staff at the University of Galway, especially Brian Saunders, for hosting the fieldwork. The authors would also like to thank their colleagues at MKO, especially John Hynes and Pat Roberts, for critical discussion and comments to improve the manuscript. We thank the Sustainable Energy Authority of Ireland for their financial support of this project.

CONFLICT OF INTEREST STATEMENT

The authors confirm that they have no conflict of interest to disclose.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/2688-8319.70238.

DATA AVAILABILITY STATEMENT

Data are available from Dryad Digital Repository: http://doi.org/10.5061/dryad.t76hdr8fn (Largey et al., 2026).

ORCID

Darío Fernández-Bellon https://orcid.org/0000-0002-8016-6515
Susan Doyle https://orcid.org/0000-0002-3465-5123

REFERENCES

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Baird, R. W., & Burkhart, S. M. (2000). Bias and variability in distance estimation on the water: implications for the management of whale watching. IWC Meeting Document SC/52/WW1 (Vol. 10), Scientific Committee of the International Whaling Commission. https://wwhandbook.iwc.int/en/downloadable-resources/searchable-database-of-scientific-literature/bias-and-variability-in-distance-estimation-on-the-water-implications-for-the-management-of-whale-watching

Band, W. (2024). Using a collision risk model to assess bird collision risks for onshore wind farms. NatureScot Research Report 909. https://www.nature.scot/doc/naturescot-research-report-909-using-collision-risk-model-assess-bird-collision-risks-onshore-wind

Band, W., Madders, M., & Whitfield, D. P. (2007). Developing field and analytical methods to assess avian collision risk at wind farms. In *Birds and wind farms: risk assessment and mitigation* (pp. 259–275). Quercus. https://www.natural-research.org/application/files/4114/9182/2839/Band_et_al_2007.pdf

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, *19*, 3–11.

Becker, F. K., Millikin, R. L., & Leslie, A. J. (2020). Using radar technology to assess visual monitoring accuracy of Cape Vulture movements. *Ostrich*, *91*(1), 83–90. https://doi.org/10.2989/00306525.2019.1664673

Butt, N., Slade, E., Thompson, J., Malhi, Y., & Riutta, T. (2013). Quantifying the sampling error in tree census measurements by volunteers and its effect on carbon stock estimates. *Ecological Applications*, *23*(4), 936–943. https://doi.org/10.1890/11-2059.1

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Cramp, S., Simmons, K., Ferguson-Lees, I., Gillmor, R., Hollom, P., Hudson, R., Nicholson, E., Ogilvie, M., Olney, P., Voous, K., & Wattel, J. (1988). *Handbook of the birds of Europe, the Middle East and North Africa*. Oxford University Press.

Gamer, M., & Lemon, J. (2019). *irr*: *Various coefficients of interrater reliability and agreement*. R Package. https://CRAN.R-project.org/package=irr

Garel, M., Cugnasse, J. M., Gaillard, J. M., Loison, A., Santosa, Y., & Maublanc, M. L. (2005). Effect of observer experience on the monitoring of a mouflon population. *Acta Theriologica*, *50*, 109–114. https://doi.org/10.1007/BF03192623

Garvin, J. C., Jennelle, C. S., Drake, D., & Grodsky, S. M. (2011). Response of raptors to a windfarm. *Journal of Applied Ecology*, *48*(1), 199–209. https://doi.org/10.1111/j.1365-2664.2010.01912.x

Gibson, E. J., & Bergman, R. (1954). The effect of training on absolute estimation of distance over the ground. *Journal of Experimental Psychology*, *48*(6), 473–482. https://doi.org/10.1037/h0055007

Gibson, E. J., Bergman, R., & Purdy, J. (1955). The effect of prior training with a scale of distance on absolute and relative judgments of distance over ground. *Journal of Experimental Psychology*, *50*(2), 97–105. https://doi.org/10.1037/h0048518

Harwood, A. J., Perrow, M. R., & Berridge, R. J. (2018). Use of an optical rangefinder to assess the reliability of seabird flight heights from boat-based surveyors: Implications for collision risk at offshore wind farms. *Journal of Field Ornithology*, *89*(4), 372–383. https://www.jstor.org/stable/44994135

Horswill, C., Miller, J. A. O., & Wood, M. J. (2022). Impact assessments of wind farms on seabird populations that overlook existing drivers of demographic change should be treated with caution. *Conservation Science and Practice*, *4*, e12644.

Johnston, A., Cook, A. S. C. P., Wright, L. J., Humphreys, E. M., & Burton, N. H. (2014). Modelling flight heights of marine birds to more accurately assess collision risk with offshore wind turbines. *Journal of Applied Ecology*, *51*(1), 31–41. https://doi.org/10.1111/1365-2664.12191

Johnston, A., Fink, D., Hochachka, W. M., & Kelling, S. (2018). Estimates of observer expertise improve species distributions from citizen science data. *Methods in Ecology and Evolution*, *9*(1), 88–97. https://doi.org/10.1111/2041-210X.12838

Katzner, T. E., Nelson, D. M., Diffendorfer, J. E., Duerr, A. E., Campbell, C. J., Leslie, D., Vander-Zanden, H. B., Yee, J. L., Sur, M., Huso, M. M., & Braham, M. A. (2019). Wind energy: An ecological challenge. *Science*, *366*(6470), 1206–1207. https://www.science.org/doi/10.1126/science.aaz9989

Kenward, R. E., Walls, S. S., Hodder, K. H., Pahkala, M., Freeman, S. N., & Simpson, V. R. (2000). The prevalence of non-breeders in raptor populations: Evidence from rings, radio-tags and transect surveys. *Oikos*, *91*(2), 271–279. https://doi.org/10.1034/j.1600-0706.2000.910207.x

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Landscape Institute. (2019). *Technical guidance note 06–19 visual representation of development proposals*. Landscape Institute. https://www.landscapeinstitute.org/wp-content/uploads/2019/09/LI_TGN-06-19_Visual_Representation-1.pdf

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*(3), 747–764. https://doi.org/10.1111/ibi.12913

Largey, N., Edwards, H. A., Fernández-Bellon, D., & Doyle, S. (2026). Data from: Implications of surveyor accuracy in bird flight height estimation for wind farm collision risk assessment. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.t76hdr8fn

Lee, D. K. (2016). Alternatives to P value: Confidence interval and effect size. *Korean Journal of Anesthesiology*, *69*(6), 555–562. https://doi.org/10.4097/kjae.2016.69.6.555

Masden, E. A. (2015). Developing an avian collision risk model to incorporate variability and uncertainty. *Scottish Marine and Freshwater Science*, *6*, 14. https://doi.org/10.7489/1659-1; http://www.scotland.gov.uk/Resource/0048/00486433.pdf

McGregor, R., King, S., Donovan, C., Caneco, B., & Webb, A. (2018). *A stochastic collision risk model for seabirds in flight*. Report No. HC0010-400-001. Marine Scotland Science, Marine Scotland, Scottish Government. https://www.gov.scot/publications/stochastic-collision-risk-model-for-seabirds-in-flight/

Meyer, D., Zeileis, A., Hornik, K., & Friendly, M. (2023). *Vcd: Visualizing categorical data*. R package version 1.4-12.

New, L., Bjerre, E., Millsap, B., Otto, M. C., & Runge, M. C. (2015). A collision risk model to predict avian fatalities at wind facilities: An example using golden eagles, *Aquila chrysaetos*. *PLoS One*, *10*, e0130978.

Oravetz, C. (2009). *Human estimation of slope, distance, and height of terrain in simulated lunar conditions* (Ph.D. Thesis). Massachusetts Institute of Technology.

Park, H., Faghihi, N., Dixit, M., Vaid, J., & McNamara, A. (2021). Judgments of object size and distance across different virtual reality environments: A preliminary study. *Applied Sciences*, *11*(23), 11510. https://doi.org/10.3390/app112311510

Percival, S. M. (2003). *Birds and wind farms in Ireland: A review of potential issues and impact assessment*. Ecology Consulting.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality*. Northwestern University. R package.

Schippers, P., Buij, R., Schotman, A., Verboom, J., van der Jeugd, H., & Jongejans, E. (2020). Mortality limits used in wind energy impact assessment underestimate impacts of wind farms on bird populations. *Ecology and Evolution*, *10*, 6274–6287.

Scottish Natural Heritage. (2017). *Recommended bird survey methods to inform impact assessment of onshore wind farms*. Scottish Natural Heritage. https://www.nature.scot/doc/recommended-bird-survey-methods-inform-impact-assessment-onshore-windfarms

Shamoun-Baranes, J., van Gasteren, H., & Ross-Smith, V. (2017). Sharing the aerosphere: Conflicts and potential solutions. In P. Chilson, W. Frick, J. Kelly, & F. Liechti (Eds.), *Aeroecology*. Springer. https://doi.org/10.1007/978-3-319-68576-2_18

Stantial, M. L., & Cohen, J. B. (2015). Estimating flight height and flight speed of breeding Piping Plovers. *Journal of Field Ornithology*, *86*(4), 369–377. https://www.jstor.org/stable/24617942

Steele, W. K., & Weston, M. A. (2023). Sex and age bias in Australian magpies struck by aircraft. *Birds*, *4*(4), 295–302. https://doi.org/10.3390/birds4040025

Sunde, P., & Jessen, L. (2013). It counts who counts: An experimental evaluation of the importance of observer effects on spotlight count estimates. *European Journal of Wildlife Research*, *59*, 645–653.

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 No. 666. British Trust for Ornithology.

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, *1*(21), 19–25.

Vander-Zanden, H. B., Nelson, D. M., Conkling, T. J., Allison, T. D., Diffendorfer, J. E., Dietsch, T. V., Fesnock, A. L., Loss, S. R., Ortiz, P. A., Paulman, R., & Rogers, K. H. (2024). The geographic extent of bird populations affected by renewable-energy development. *Conservation Biology*, *38*, e14191.

Vas, E., Lescroël, A., Duriez, O., Boguszewski, G., & Grémillet, D. (2015). Approaching birds with drones: First experiments and ethical guidelines. *Biology Letters*, *11*(2), 20140754. https://doi.org/10.1098/rsbl.2014.0754

Watson, P. F., & Petrie, A. (2010). Method agreement analysis: A review of correct methodology. *Theriogenology*, *73*(9), 1167–1179. https://doi.org/10.1016/j.theriogenology.2010.01.003

Williams, I. D., Walsh, W. J., Tissot, B. N., & Hallacher, L. E. (2006). Impact of observers' experience level on counts of fishes in underwater visual surveys. *Marine Ecology Progress Series*, *310*, 185–191.

Yang, T. L., Dixon, M. W., & Proffitt, D. R. (1999). Seeing big things: Overestimation of heights is greater for real objects than for

BRITISH ECOLOGICAL SOCIETY Ecological Solutions and Evidence

objects in pictures. *Perception*, *28*(4), 445–467. https://doi.org/10.1068/p2854

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Supporting Information S1. Surveyor accuracy.

Supporting Information S2. Collision risk model.

How to cite this article: Largey, N., Edwards, H. A., Fernández-Bellon, D., & Doyle, S. (2026). Implications of surveyor accuracy in bird flight height estimation for wind farm collision risk assessment. *Ecological Solutions and Evidence*, *7*, e70238. https://doi.org/10.1002/2688-8319.70238