METHODOLOGY

How to improve the accuracy of height data from bird tracking devices? An assessment of high-frequency GPS tracking and barometric altimetry in field conditions

Tonio Schaub[1,2,3,4*] , Alexandre Millon[1,5] , Caroline De Zutter[2], Ralph Buij[6], Joël Chadœuf[5], Simon Lee[7,8], Aymeric Mionnet[9] and Raymond Hendrikus Gerardus Klaassen[3,4]

Abstract

Background In the context of rapid development of wind energy infrastructure, information on the flight height of birds is vital to assess their collision risk with wind turbines. GPS tags potentially represent a powerful tool to collect flight height data, yet GPS positions are associated with substantial vertical error. Here, we assessed to what extent high-frequency GPS tracking with fix intervals of 2–3 s (GPS remaining turned on between fixes), or barometric altimetry using air pressure loggers integrated in GPS tags, improved the accuracy of height data compared to standard low-frequency GPS tracking (fix interval ≥ 5 min; GPS turned off between fixes).

Results Using data from 10 GPS tag models from three manufacturers in a field setting (194 tags deployed on free-living raptors), we estimated vertical accuracy based on periods when the birds were stationary on the ground (true height above ground was approximately zero), and the difference between GPS and barometric height in flight. In GPS height data, vertical accuracy was mainly driven by noise (little bias), while in barometric data, it was mostly affected by bias (little noise). In high-frequency GPS data, vertical accuracy was improved compared to low-frequency data in each tag model (mean absolute error (AE) reduced by 72% on average; range of mean AE 2–7 vs. 7–30 m). In barometric data, vertical accuracy did not differ between high- and low-frequency modes, with a bias of − 15 to − 5 m and mean AE of 7–15 m in stationary positions. However, the median difference between GPS and barometric data was smaller in flight positions than in stationary positions, suggesting that the bias in barometric height data was smaller in flight. Finally, simulations showed that the remaining vertical error in barometric and high-frequency GPS data had little effect on flight height distributions and the proportion of positions within the collision risk height range, as opposed to the extensive noise found in low-frequency GPS data in some tag models.

Conclusions Barometric altimetry may provide more accurate height data than standard low-frequency GPS tracking, but it involves the risk of a systematic error. Currently, high-frequency GPS tracking provides highest vertical accuracy and may thus substantially advance the study of wind turbine collision risk in birds.

Keywords Telemetry, Flight height, Vertical accuracy, Vertical error, Wind energy, Wind turbine, Collision risk, Bird of prey, Raptor, Continuous GPS mode

*Correspondence:
Tonio Schaub
tonio.schaub@imbe.fr
Full list of author information is available at the end of the article

Schaub *et al. Animal Biotelemetry*   (2023) 11:31

Page 2 of 16

Background

Flying animals have been shown to suffer mortality from collision with vertical human infrastructures [1, 2]. In particular, collisions of birds with wind turbines can have substantial negative population impacts [3, 4]. This problem is expected to increase in the near future as the number of wind turbines is going to grow worldwide to fulfil the targets for renewable energy production. Therefore, there is an urgent need to quantify collision risk and identify effective mitigation measures reducing the number of casualties. However, this is currently hampered by a lack of accurate data on flight height. These are a prerequisite to reliably quantify the probability of flying within the collision risk height range and the avoidance of wind turbines in birds, two crucial components of collision risk models [5, 6].

Earlier methods to study flight height of birds have been relatively inaccurate (visual observations; [7, 8]) or provided only short sequences of accurate data without bird determination at the species level (radar; [9, 10]). Individual-based tracking by animal-borne GPS tracking devices represents a promising source of flight height data over extended periods [11, 12]. However, GPS positions are associated with inherent horizontal and vertical error. The vertical error can be substantial (mean absolute error up to 30 m [13]) and potentially bias the outcomes of collision risk analyses [14]. Methods have been proposed to account for the error a posteriori within a modelling framework [12, 14]. However, these state-space models require high levels of statistical expertise and computational capacities, and have therefore been little applied until now. Moreover, large errors increase the uncertainty around model outcomes, and particular behavioural aspects like the avoidance of wind turbines by birds require a high level of accuracy for individual data points. For these reasons, it remains critical to increase the vertical accuracy in the raw tracking data.

One possible approach to improve the three-dimensional accuracy of GPS positions is to increase the GPS fix frequency [13, 15]. The highest accuracy is expected for positions obtained when the GPS module does not turn off between successive fixes ("continuous GPS mode"). This occurs when fixes are collected at a high frequency, typically when the time interval between successive GPS fixes is below 5–20 s, depending on the GPS tag model. In this scenario, on average more satellites are used per fix compared to standard low-frequency GPS data collection, where the GPS module is turned off after every fix ("discrete GPS mode"). However, the extent of the accuracy improvement in the high-frequency mode and its consistency across different tag models remains to be demonstrated. Moreover,

a downside of high-frequency GPS tracking is that it is energy demanding, usually depleting the batteries of GPS tags within hours to days (depending on battery size and solar charging conditions).

A second possibility to increase the accuracy of flight height data is the use of barometers (air pressure sensors), which are increasingly integrated into GPS tags. These sensors operate independently from the GPS regarding the height measurement (but still depend on the GPS to determine the horizontal position, necessary to determine the height above ground). The measured air pressure is combined with local weather data in the barometric height formula to determine height [14, 16]. Advantages of barometric altimetry are that it is energy-efficient, barely increasing battery demand compared to GPS fixes without pressure measurement, and that, a priori, accuracy is not related to sampling frequency. However, barometers need to be calibrated and the barometric height calculation requires accurate local weather data. Moreover, the assumptions of the formula regarding the stratification of the atmosphere are not always met in practice [14]. Therefore, it is unclear how barometric altimetry performs under field conditions.

Here, we performed an extensive field test of these two methods to increase the accuracy in flight height data in comparison to standard low-frequency GPS height data. Our study built on a GPS tracking data set of ca. 11 million positions obtained from 194 tags of 10 models from three different manufacturers deployed on four raptor species in France and the Netherlands. Our main approach of quantifying vertical accuracy was based on stationary periods when the birds were positioned on the ground, providing a known true height above ground (i.e. approximately zero). First, we analysed the deviation of GPS and barometric height from true height for these stationary periods and assessed the consistency of the results among GPS tag models. Secondly, to extend the assessment from stationary periods to flight periods, we quantified the deviation between GPS and barometric height for both stationary and flight periods, providing an indirect measure of accuracy. Thirdly, we assessed the credibility of height profiles from high-frequency sampling and identified recurrent error patterns. Fourthly, we quantified the consequences of different levels of error for practical conservation-related questions, using the proportion of positions within the height range of wind turbine rotors as an example. Finally, we provided guidance on how to improve the vertical accuracy of tracking data from GPS tags against the background of the limitations of the different methods.

Schaub *et al. Animal Biotelemetry*   (2023) 11:31

Page 3 of 16

Methods

Data collection

We used data from 194 solar-powered GPS tags which were deployed between 2009 and 2022 on 204 individuals of four raptor species: Montagu's harrier *Circus pygargus*, hen harrier *C. cyaneus*, marsh harrier *C. aeruginosus* and red kite *Milvus milvus* (Additional file 1: Table S1). Birds were captured during the breeding season as adults (*n* = 140) or as nestlings (*n* = 64) close to or on the nest in four study areas in France and the Netherlands (Champagne, Grand Est, Flevoland and Groningen; Additional file 1: Fig. S1).

In the Champagne, Flevoland and Groningen areas, the landscape is open and dominated by intensive arable farming, while in the Grand Est area, it is composed of a mixture of forests, pastures and arable fields. The Flevoland and Groningen areas are flat (standard deviation [SD] of elevation above sea level [a.s.l.]: 1.7 and 1.3 m; mean: −3.1 m and −4.5 m, respectively), whereas the terrain in Champagne and especially in Grand Est is hillier (SD: 33.3 and 63.5 m; mean: 144.0 m and 311.2 m a.s.l.).

Ten different tag models from three manufacturers (*Milsar, Ornitela, UvA-BiTS*) were applied (Additional file 1: Table S1), three of which included a barometric sensor. The *Milsar* and *Ornitela* tags transferred the recorded data remotely via the GSM network, whereas the data from *UvA-BiTS* tags were downloaded using a local antenna system [13]. GPS tags were mounted as backpacks using thoracic x-strap harnesses [17] made from Teflon ribbon. Tags weighed 9.7–24.3 g according to the species, representing on average 3.2% of individual body weight (median 2.9%; SD 1.1%; range 1.7–6.5%; *n* = 207 deployments). There were no indications of adverse tag effects; the tagged birds fulfilled their annual cycle and reproduced as expected.

In spring and summer, 5 min were used as GPS fix intervals as a basic setting during daytime, except for hen harriers in Champagne (15 min). During night, the interval was set to 1–4 h. For autumn and winter, periods of bad weather, and incubation periods in females, the interval was increased to 1–12 h to preserve battery voltage. In addition, high-frequency data were collected using an interval of 3 s in *Ornitela* and *UvA-BiTS* tags and 1 s in *Milsar* tags. With the set interval of 1 s, *Milsar* tags collected GPS fixes at intervals of 2–3 s in practice. These GPS fix intervals were below the manufacturer-specific time thresholds for the continuous GPS mode (< 7 s for *Ornitela*, < 8 s for Milsar, < 16 s for *UvA-BiTS*). High-frequency data were collected mostly during hourly blocks (1–2 h per day), and to a lesser extent using geofences defined around areas of interest (e.g. wind farms, fields with agri-environmental schemes). High- and low-frequency data were similarly distributed across years within tag models (Additional file 1: Fig. S2). In tags with barometric sensor, air pressure measurements were taken alongside every GPS fix. We distinguished four methods of height data collection, i.e. low-frequency GPS (discrete mode), high-frequency GPS (continuous mode), low-frequency barometric and high-frequency barometric.

After removing positions outside of the defined study areas, the dataset comprised 10,777,644 positions with GPS height (2,881,769 from low-frequency and 7,895,875 from high-frequency sampling) and 3,610,374 with barometric height (740,306 from low-frequency and 2,870,068 from high-frequency sampling; Additional file 1: Table S1). The number of height data varied greatly between tags (range for GPS: 111 to 614,099 positions per tag; median: 21,574; mean: 55,555; range for barometric: 3762–388,140; median: 55,086; mean: 97,578), mainly as a consequence of variation in the length of the data collection period (range of the number of days with data per tag: 6–971; median: 125 d; mean: 196.6 d). For Montagu's and marsh harriers (trans-Saharan migrants), all individuals left the study areas in the non-breeding season, thus the dataset included only data from spring and summer. Also for hen harriers and red kites (partial migrants), the majority of individuals left the study areas in winter and for the remaining individuals fewer data could be collected in autumn and winter due to low battery voltage.

Data processing

All data processing and analyses were performed in *R* 4.0.3 [18]. We differentiated between stationary and flight positions based on the instantaneous GPS ground speed which is recorded alongside every GPS position. The distribution of speed values typically shows two modes (one representing stationary and one representing flight positions), and we used the antimode between the two modes as threshold ([19]; Additional file S1). The speed threshold was determined for each combination of species and tag manufacturer, separately for low- and high-frequency data ($1.81–3.83$ m s^{-1} and $0.85–1.86$ m s^{-1} for low- and high-frequency data, respectively).

The GPS height data obtained from the tags were heights above mean sea level (termed height a.s.l. hereafter), i.e. above geoid. However, when comparing height data from different sources, it is important to verify that the same geoid model is used and if not, apply corrections [14]. For *Milsar* and *UvA-BiTS* tags, the manufacturers indicated that EGM96 was used. For the *Ornitela* tags, it was possible to also obtain the height above ellipsoid, i.e. the raw height data above the WGS84 ellipsoid initially determined by the GPS module before application of a geoid model. This led us to notice that the geoid model applied in these tags was biased compared to EGM96 in

Schaub *et al. Animal Biotelemetry*   (2023) 11:31

Page 4 of 16

some study areas. Therefore, to obtain corrected height a.s.l. data, we used the height above ellipsoid data and applied the EGM96 geoid model with resolution of 0.25° [20]. By this correction, the height a.s.l. was offset by a mean of + 3.5 m for Flevoland, + 4.4 m for Groningen, 0.0 m for Champagne and − 1.2 m for Grand Est. For *Milsar* and *UvA-BiTS*, it was not possible to obtain the height above ellipsoid data to apply the same test.

In *Milsar* tags, the GPS height data were internally and irreversibly truncated at sea level. Therefore, the lowest recorded height above ground level (termed height a.g.l. hereafter; see below) in *Milsar* tags was − 197.5 m, whereas much lower values were obtained from the other tag models (Additional file 2: Table S3), likely leading to an underestimation of the vertical error in *Milsar* data.

The calculation of barometric height based on the pressure measurements of the tags was performed using the barometric formula describing the relationship of air pressure with height above a reference level under different meteorological conditions [21]: $z = -\frac{T_0}{L} * (1 - \left(\frac{P}{P_0}\right)^{-\frac{LR_0}{g}})$, where z is the height above the reference level, T_0 is the temperature at reference height, L is the temperature lapse rate, P is the pressure at height z (measured by the tag), P_0 is the pressure at reference height, R_0 is the specific gas constant (287.05 J K^{-1} kg^{-1}) and g is the standard acceleration of free fall (9.81 m s^{-1}). We obtained data on T_0, L and P_0 from the global weather model ECMWF ERA5 with a temporal resolution of 1 h and a spatial resolution of 0.25° [22, 23]. The tracking data were annotated with ERA5 data using the Environmental-Data Automated Track Annotation System (Env-Data) provided by Movebank [24], which included an interpolation of the ERA5 data to the timestamp and horizontal position of each GPS fix. The resulting height above the ERA5 model surface was transformed into height a.s.l. (see Additional file 1 for details).

Both for GPS and barometric height, we transformed height a.s.l. into height a.g.l. by applying the European Digital Elevation Model (EU-DEM, v1.1) with a resolution of 25 m [25]. EU-DEM is based on the EEG2008 geoid, but the difference between EEG2008 and EGM96 (used in the GPS data and the weather model) was negligible for our study areas (mean absolute difference 0.12 m, maximum difference 0.65 m).

Identification of stationary positions on the ground

Our assessment of vertical accuracy was based on positions when the birds were stationary on the ground, as for these positions the true height a.g.l. was known. Note that in fact, the true height a.g.l. was not zero, but the height of the back of the bird where the

tag was attached. However, as in our study species this difference was small (15–30 cm), we applied zero as true height. To identify these "ground positions", two different approaches were adopted. For the three species of harriers which are known to sit on the ground most of the time when being stationary, and for which the landscape in the study areas was relatively homogeneous with low occurrence of vertical structures (large-scale open agricultural areas), we used the digital national topographic maps *BD TOPO* for France [26] and *TOP10NL* for the Netherlands [27]. Positions at > 50 m from vertical structures (trees, hedgerows, buildings, electric pylons) were classified as ground positions (see Additional file 1 for details). The proportion of ground positions amongst stationary positions varied between 82.9 and 99.8% for the combinations of species and study area.

Contrary to the harriers, red kites are known to perch on trees or other vertical structures most of the time when stationary. Moreover, the landscape in the red kite study area was more heterogeneous with more vertical structures (more interspersed trees, hedgerows and forests; more field margins with fence poles), which were only partially included in the digital national geographic maps. Therefore, we applied a more restrictive approach by classifying the perching habitat manually by visual inspection of satellite images. We identified continuous stationary periods during daytime consisting of ≥ 2 subsequent positions in low-frequency and 20 positions in high-frequency data, with < 50 m between subsequent positions. Periods were defined as ground periods if all positions were on agricultural fields and if the mean coordinates were > 20 m away from any vertical structures or field margins visible in the satellite image. (Note that the more restrictive classification approach allowed to reduce the threshold distance compared to the harrier case.) Out of the 2400 inspected periods (random sample), 15.7% were classified as ground periods, comprising 31,948 individual positions (8.8% of the classified positions).

Estimation of vertical accuracy and comparison between methods and tag models

Conceptually, we considered the error in the height data from stationary position on the ground on three levels, i.e. trueness, precision and accuracy [28]. Trueness refers to the deviation of the average of the measured values from a reference value (bias or systematic error), which we described using the deviation of the mean, and the median, from the true height, i.e. zero. Precision refers to the deviation of individual measurements from the average (noise or random error),

Schaub *et al. Animal Biotelemetry* (2023) 11:31

Page 5 of 16

which we described using the mean, median and 95% quantile of absolute error (AE) and the root mean square error (RMSE), all with the median as reference. Accuracy refers to the combination of precision and trueness, i.e. the deviation of individual measurements from the true value, which we described using the same parameters as for precision, but with the true height, i.e. zero, as reference.

To reduce temporal auto-correlation in the high-frequency data for the statistical analyses, we subsampled the tracking data to a minimum interval of 5 min. As positions at the beginning of high-frequency blocks and short stationary periods were overrepresented in the subsampled data and these had higher than average vertical error, we removed the first minute of every high-frequency block and stationary periods consisting of < 5 subsequent positions before subsampling to prevent bias.

To statistically compare vertical accuracy on the three levels across methods and tag models, we applied hierarchical bootstrapping [29]. We chose this non-parametric method to estimate confidence intervals because the distributions of the height data had very long tails (see Results), which prevented the use of parametric methods like linear models (residual distributions remained unsatisfactory after log or Box–Cox transformation of the response variable). For each of the 26 combinations of method and tag model, we resampled at the first hierarchical level (individual tags) with replacement, and then without replacement at the second hierarchical level (individual height data within each resampled tag) following Ren et al. [29]. In this way, 1,000 bootstrap replicates were constructed for each combination for six parameters of interest, i.e. mean and median error with true height as reference (trueness), mean and median absolute error with median height as reference (precision), and mean and median absolute error with true height as reference (accuracy). We used the mean and the range between the 2.5% and 97.5% quantiles across the replicates as estimate and confidence interval. We considered differences between groups to be significant when the confidence intervals did not overlap.

Visual inspections of high-frequency tracking data

To assess the credibility of height profiles in high-frequency tracks across stationary and flight positions, and to identify potential error patterns, we carried out visual inspections of individual high-frequency tracks. A graph of height a.g.l. over time was produced for every track of at least 100 consecutive high-frequency positions ($n = 9993$ high-frequency tracks).

Effect of error on flight height distributions and proportion of positions at collision risk height range based on simulations

To assess the effect of error on flight height distributions and derived flight parameters relevant for conservation, we performed simulations by adding different levels of bias or noise to two example flight height distributions from high-frequency GPS data, from red kites in Grand Est (tag model *OT-25*) and from marsh harriers in Groningen (tag models *4C.L* and *6C.L*; Additional file 1: Table S1). As an example of a derived parameter, we used the proportion of positions at the height range of wind turbine rotors, which is a commonly used input parameter in collision risk assessments [30, 31]. We applied 50–200 m a.g.l. as collision risk height range (CRHR), representing the height range of the rotors of most modern wind turbines. Concerning the proportion of positions within the CRHR, the two example datasets represented the extremes among our study species, with 37.4% of positions within the CRHR in the red kites data, compared to 4.2% in the marsh harrier data (distribution modes around 22 and 1 m a.g.l., respectively; Additional file 3: Fig. S8). Note that the flight height data used here are not free of error, but it is sufficiently small (see Results) not to be problematic for this illustrative purpose.

To clearly separate the effect of precision (noise) and trueness (bias), we applied both types of error separately. For bias, we applied both the mean error found for each combination of tag model and method in this study based on stationary positions on the ground (26 values; Additional file 2: Table S3), and a theoretical range of bias between −20 and 20 m with increments of 1 m. These levels of bias were added to the flight positions as a constant.

Regarding noise, first, we applied the empirical error distributions found in the 10 tag models, with the median per combination of tag model and method as reference ("precision"), on the two flight height distributions. We added an error randomly drawn from the error distributions to each flight position. Secondly, we applied theoretical error distributions to illustrate the effect of gradually increasing error. We applied exponential distributions for the AE $F(x) = \lambda e^{-\lambda x}$ with rate parameter $\lambda = \frac{1}{\bar{x}}$ where $\bar{x}$ (i.e. mean AE) was varied between 1 and 40 m (increments of 1 m), and normal distributions with standard deviation varying between 1 and 50 m (increments of 1 m), corresponding to a mean AE of 0.8–39.9 m. The range of mean AE for the theoretical distributions was chosen so that it covered the range of mean AE in relation to the median in the empirical distributions (1.3–29.5 m; see Results), with some extension

Schaub *et al. Animal Biotelemetry*   (2023) 11:31

Page 6 of 16

towards higher values which could be present in GPS tag model not studied here. Note that the exponential error distributions generally matched the empirical error distributions better than the normal distributions. For each flight position, we added or subtracted a randomly drawn value from the exponential or normal error distributions (random choice of algebraic sign in the exponential distributions).

For both empirical and theoretical error distributions, we plotted the relative increase of the proportion of positions within the CRHR, compared to the baseline where no additional error was applied, against the mean absolute error of the error distributions.

Results

Estimation of vertical accuracy based on stationary periods

Overall, the distributions of error around true height from stationary positions on the ground showed a clear mode (Fig. 1). The medians of the recorded height a.g.l. were close to zero in the GPS data (−3.8 to 4.3 m), while barometric height data had a reduced trueness with median height a.g.l. between −15.0 and −4.9 m (Fig. 2; Additional file 2: Table S3). Trueness did not differ significantly between low- and high-frequency sampling within GPS or barometric data in most tag models (Fig. 2).

By contrast, there was a much higher variation in precision (error around median height) in low-frequency GPS data between tag models, with median AE ranging from 2.6 to 17.4 m (mean across tag models ± SD 6.3 ± 4.6), compared to high-frequency GPS data (range of median AE 1.0–4.0 m; mean 2.4 ± 1.0) and to both low- and high-frequency barometric data (range of median AE 2.8–4.2

[mean 3.5 ± 0.7] and 2.3–3.5 m [mean 2.9 ± 0.6], respectively). Most importantly, in low-frequency GPS data, the median AE around the median was on average 2.6 times larger than in high-frequency GPS data (median 2.3; range 1.5–6.2). In barometric data, regardless the sampling frequency, precision was similar to high-frequency GPS data or slightly higher (Fig. 2; Additional file 2: Table S3). Large outliers with absolute height above median > 50 m occurred regularly in low-frequency GPS data (on average 6.6% of positions; range 0.3–17.4%), whereas these were much scarcer in high-frequency GPS data (mean 0.4%; range 0.0–1.5%), and nearly absent in barometric data (mean 0.1%; range 0.0–0.2%; Additional file 2: Table S5). In every tag model, the mean AE was higher than the median AE, especially in low-frequency GPS data, reflecting the long tails of the AE distributions. Therefore, differences between low- and high-frequency GPS data increased when considering mean instead of median AE (mean AE in low-frequency data on average 8.1 times larger than in high-frequency data; median 3.8; range 2.0–19.9; Additional file 3: Fig. S3).

Also regarding overall accuracy, low-frequency GPS data had larger errors (with true height as reference) than high-frequency GPS data in all tag models. Median AE ranged from 3.3 to 18.9 m in low-frequency GPS data (mean 6.8 ± 4.8 m), and from 1.2 to 4.0 in high-frequency data (mean 2.9 ± 0.9 m; median AE on average 2.4 times larger in low-frequency data; median 1.9; range 1.4–6.5). Mean AE ranged from 7.4 to 29.9 m in low-frequency GPS data (mean 18.9 ± 18.9 m), and from 1.5 to 7.0 in high-frequency data (mean 3.9 ± 1.7 m; mean AE on average 6.6 times larger in low-frequency data; median 3.4;
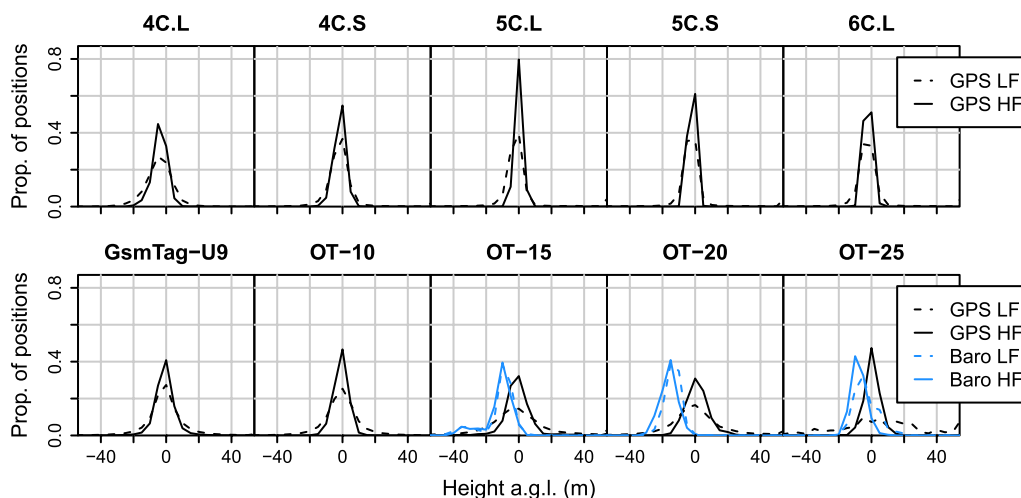

Fig. 1 Distributions of the recorded heights above ground level from stationary positions on the ground for each tag model and method. The lines connect the proportions of positions per height class of 5 m (central class centred around zero). Height data < − 50 and > 50 m a.g.l. not shown. *Prop.* proportion, *LF* low frequency, *HF* high frequency
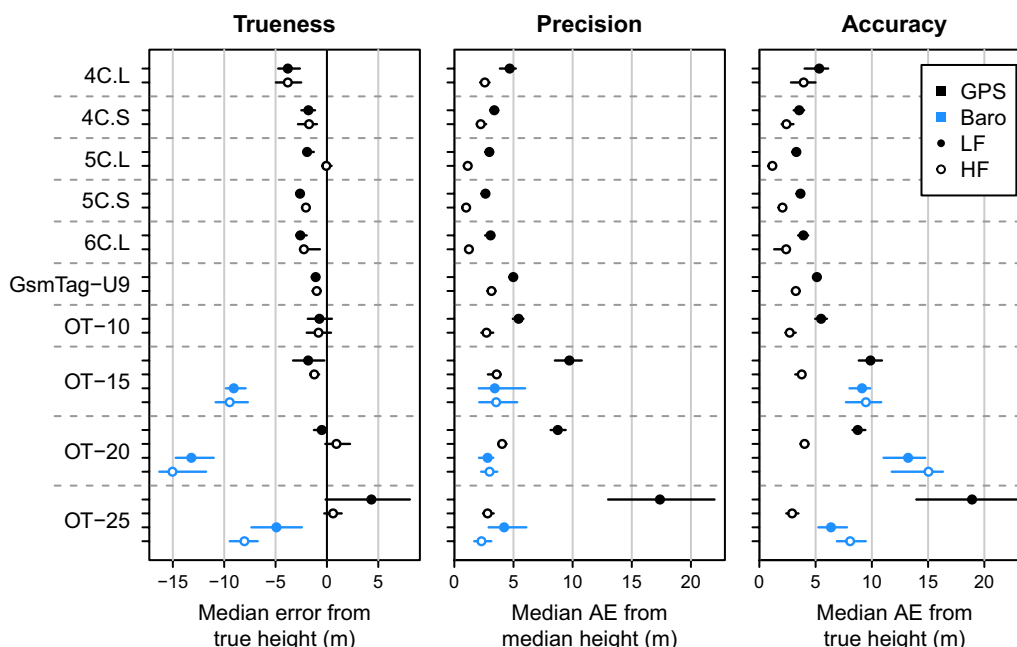
Fig. 2 Estimates and 95% confidence intervals for trueness, precision and accuracy for each combination of GPS tag model and method based on hierarchical bootstrapping. Parameters used: median error with true height as reference (equivalent to median height a.g.l.; trueness), median absolute error with median height as reference (precision), median absolute error with true height as reference (accuracy). *a.g.l.* above ground level, *AE* absolute error

range 1.6–19.0; Additional file 2: Table S3). The difference between high- and low-frequency GPS data was significant in all cases (except for one tag model for median AE; Fig. 2, Additional file 3: Fig. S3). In barometric data, accuracy did not differ between high- and low-frequency data in any tag model (Fig. 2, Additional file 3: Fig. S3). Median AE varied between 6.4 and 15.0 m (mean 10.2 ± 3.3 m) and mean AE between 6.8 and 15.2 m (mean 11.3 ± 3.2 m). The results for barometric height data in comparison to high- and low-frequency GPS data were mixed among tag models, with barometric data being less accurate (based on median AE) than low-frequency GPS in *OT-20*, similarly accurate than low-frequency GPS in *OT-15* and intermediate between low- and high-frequency GPS in *OT-25* (Fig. 2).

Difference between GPS and barometric height

For stationary positions, regardless of sampling frequency, the difference between GPS and barometric height was on average larger than zero (range of median difference: 4.9–16.4 m; Fig. 3, Additional file 2: Table S6), i.e. barometric height was on average lower than GPS height. However, the median difference was smaller or even slightly negative for flight positions (range of median difference: −1.1 to 7.6 m; Fig. 3). The
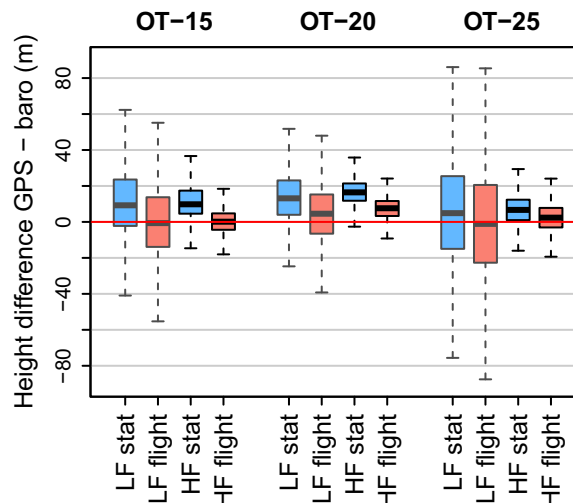

Fig. 3 Difference between GPS and barometric height for three tag models (*OT-15*, *OT-20* and *OT-25*) for high- and low-frequency sampling (LF/HF) and stationary ("stat") and flight positions ("flight"). Thick horizontal lines indicate medians and boxes the ranges between the 1st and the 3rd quartiles; whiskers extend to the most extreme data point at a distance of no more than 1.5 times the interquartile range from the box; data points outside of whiskers were omitted. See Additional file 2: Table S6 for sample sizes

Schaub *et al. Animal Biotelemetry* (2023) 11:31

Page 8 of 16

median difference also changed with (barometric) height a.g.l. in low-frequency data, becoming negative in positions > 0–40 m a.g.l. in two of the three tag models, indicating that barometric height was on average higher than GPS height (Fig. 4; see Additional file 3: Fig. S4–S6 for graphs showing the entire height range and data from the other two tag models). In absolute terms, the difference between GPS and barometric height data was smallest in high-frequency flight positions (range of median absolute difference: 4.6–8.1 m; compared to 8.0–16.5 m in stationary high-frequency positions and 11.6–21.6 m in low-frequency positions; Additional file 2: Table S6).

In line with the differences found between GPS and barometric height data in flight, the flight height distributions based on barometric data appeared to be shifted by a few metres compared to those from high-frequency GPS data in two of the tag models, whereas the shapes of the distributions were similar (Fig. 5, Additional file 3: Fig. S9). By contrast, in low-frequency GPS data, the flight height distributions differed remarkably from those of the three other methods by being flattened out, showing a less pronounced peak.

Description of high-frequency tracking data including recurrent error patterns

Overall, high-frequency tracks from all tag models showed realistic flight movements, in line with flight patterns expected for our study species. Thermal ascent flights were easily discernible by zig-zag patterns in the horizontal plane, and commonly alternated with descending gliding flights (Fig. 6b). The height sequences of barometric height and GPS height were generally very close to each other (Fig. 6c, Additional file 3: Fig. S7).

Nevertheless, we identified three recurrent error patterns in GPS height data from high-frequency tracks, with variable frequency across the three tag manufacturers. First, GPS height often showed a quick increase or decrease at the beginning of a high-frequency bout. When barometric data were available, this frequently coincided with a conspicuous offset of the GPS height compared to the barometric height which disappeared usually within 30–60 s (Fig. 6c, Additional file 3: Fig. S7c-d). Secondly, gradual drifting of GPS height during stationary periods was observed in *Ornitela* tags, mainly at a scale < 30 m (Fig. 6c, Additional file 3: Fig. S7b). Thirdly, height data from *Milsar* tags included "spikes", i.e. individual and easily discernible outliers, normally at a scale < 50 m (typically 20–50 spikes per hour; Fig. 6d).

The pattern of changes in the difference between GPS and barometric height in relation to movement (stationary vs. flight) was also observed during the visual inspection of high-frequency tracks, with abrupt changes coinciding with the moments of take-off and landing (Additional file 3: Fig. S7d).
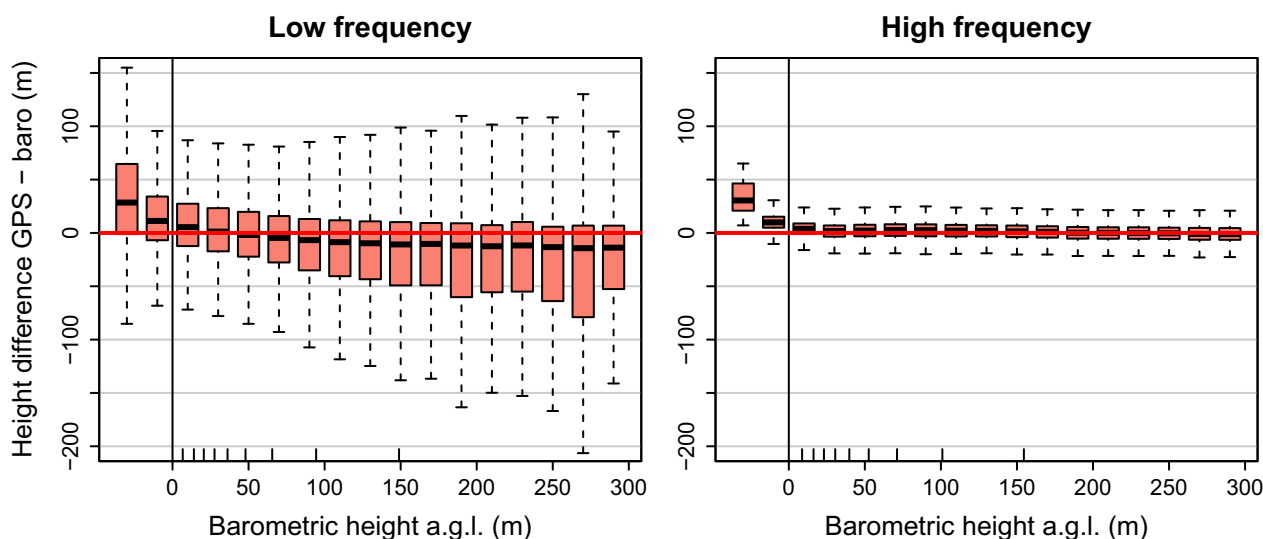

Fig. 4 Difference between GPS and barometric height in relation to height above ground level (a.g.l.; range −40 to 300 m) for *OT-25* GPS tags deployed on red kites in low- and high-frequency sampling (bins of 20 m). Only flight positions were considered. Thick horizontal lines indicate medians and boxes the ranges between the 1st and the 3rd quartiles; whiskers extend to the most extreme data point at a distance of no more than 1.5 times the interquartile range from the box; data points outside of whiskers were omitted. Sample size: 126,278 positions for low-frequency; 1,032,956 for high-frequency. Tick marks above the x-axis indicate deciles
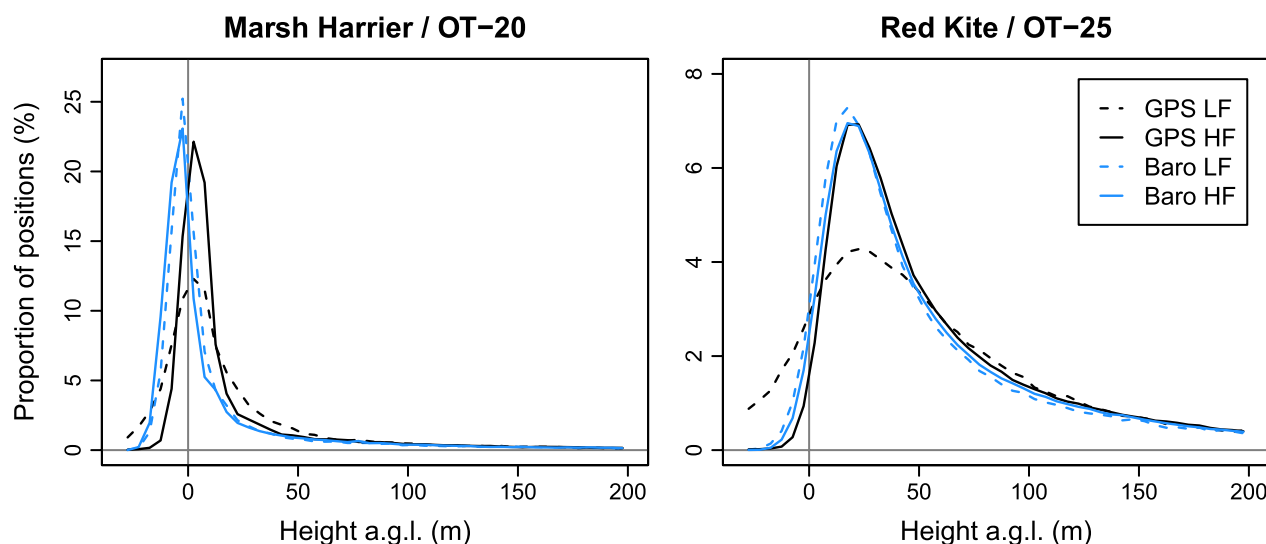
Fig. 5 Flight height distributions of marsh harriers and red kites based on either GPS or barometric height data from *OT-20* and *OT-25* GPS tags collected using either low- (LF) or high-frequency (HF) sampling (height classes of 5 m). Note different scales of the y-axis between panels. Height data < − 30 m and > 200 m above ground level (a.g.l.) not shown

Effect of error on flight height distributions and proportion of positions at collision risk height range based on simulations

When applying additional bias to flight height data of red kites and marsh harriers, the effect on the proportion of positions within the CRHR was similar in both species (Fig. 7). The levels of bias found in the different GPS tag models in this study in stationary positions lead to a relative change of the proportion at risk height of − 22.1% to + 8.3% in red kites, and − 24.2% to + 9.0% in marsh harriers.

When applying additional noise, the flight height distributions were flattened out with less pronounced peaks (Additional file 3: Fig. S8), similarly to the empirical flight height distributions based on low-frequency GPS data (Fig. 5). The proportion of positions within the CRHR generally increased with increasing additional noise (Fig. 7). The effect of noise depended on the flight height distribution of the considered species. In marsh harriers (steep flight height distribution with low mode; Additional file 3: Fig. S8), the proportion of positions within the CRHR was overestimated by > 50% in six out of ten applied empirical error distributions from low-frequency GPS data, with a maximum of 209.5% (Fig. 7). By contrast, in red kites showing a flatter flight height distribution with mode closer to the CRHR compared to marsh harriers, the proportion within the CRHR was only overestimated by up to 12.0%.

Discussion

Based on a data set consisting of ca. 11 million GPS positions collected using 194 tags of 10 GPS models from three manufacturers, we found substantial differences in accuracy between different methods of collecting height data (low-frequency GPS, high-frequency GPS, low-frequency barometric, high-frequency barometric). In GPS data, the vertical error consisted mainly of noise rather than bias, whereas the barometric data mainly suffered from bias, with relatively little noise. Notably, overall accuracy was improved in high-frequency (continuous-mode) compared to low-frequency (discrete-mode) GPS height data. In barometric data, vertical accuracy was intermediate in stationary positions, but likely the bias was smaller in flight.

Importantly, using simulations based on our empirical data, we showed that the degree of error found in low-frequency GPS data can significantly bias the outcomes of practical applications of the data in some conditions. More specifically, noise in the height data can lead to a significant increase of the proportion of positions within the collision risk height range (CRHR). This would in turn lead to an important overestimation of wind turbine collision mortality when implemented in collision risk models [6]. In other words, this confirmed that the low accuracy in low-frequency GPS data can be a genuine problem in the study of collision risk of birds with
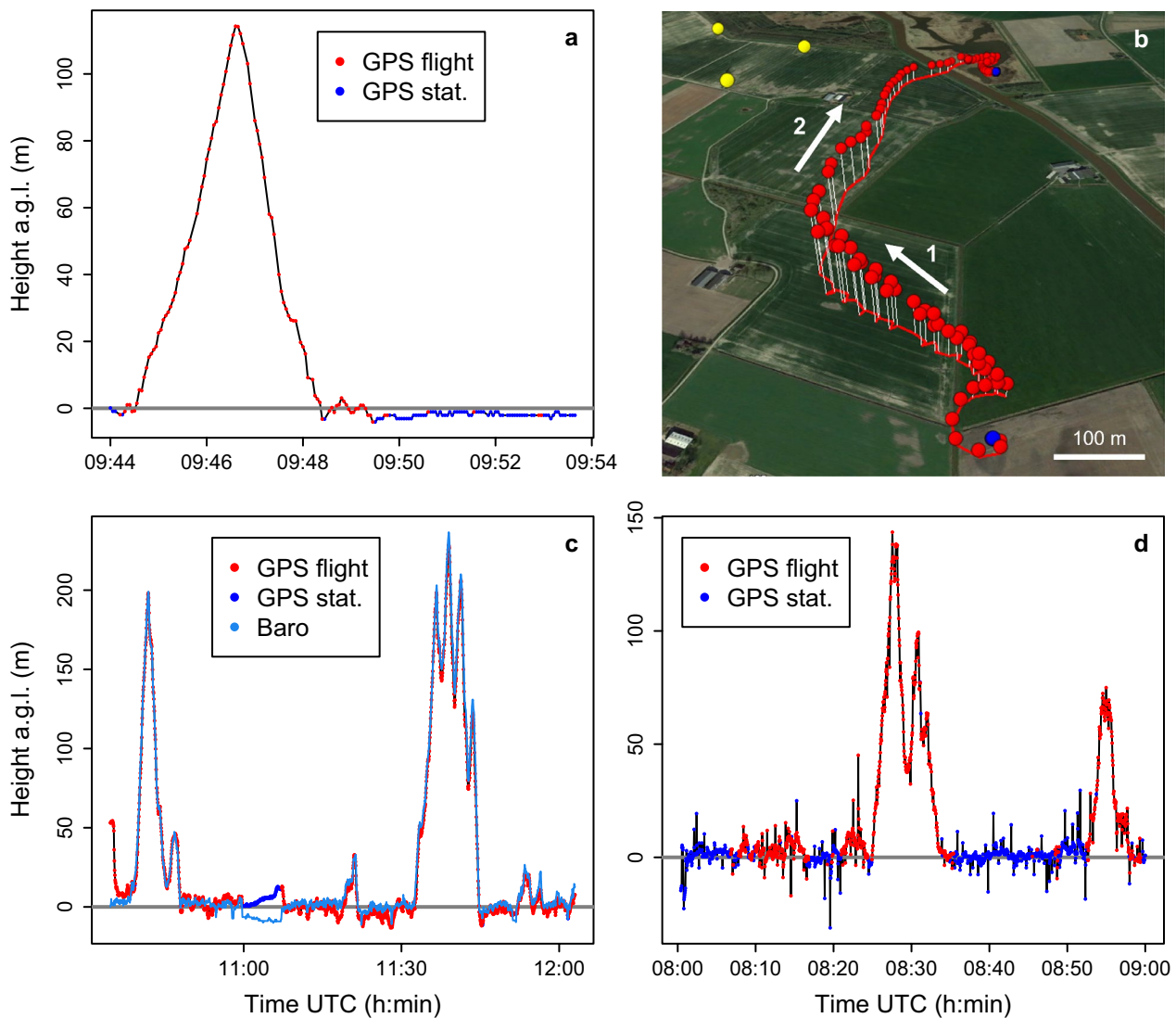
Fig. 6 Typical examples of height profiles of high-frequency tracks: **a** example showing realistic height profile in high-frequency sampling (thermal ascent flight from ground level to c. 110 m a.g.l., descending gliding flight back to approximate ground level; marsh harrier, tag model *UvA-BiTS 6C.L*); **b** 3D representation of the same track showing zig-zag pattern during thermal ascent (1) and straight descending gliding flight (2); yellow points: wind turbines; satellite image: *Google Earth*; **c** example showing closeness between GPS and barometric height, accuracy time lag in GPS height data in the first minute of the sequence (quickly decreasing GPS height) and drift of GPS height in stationary periods (marsh harrier, *Ornitela OT-15*); **d** example showing "spikes" in GPS height data from *Milsar* tags (hen harrier, *Milsar GsmTag-U9*). Note different time scales on the *x*-axis between panels (**a**, **b** 10 min, **c** 90 min, **d** 60 min). *a.g.l.* above ground level, *stat.* stationary

wind turbines and other vertical human infrastructures. By contrast, the effect of the remaining error in high-frequency GPS data and barometric data on the proportion of positions within the CRHR was small.

Accuracy in GPS height

We found that GPS height data were more accurate in the high-frequency (continuous) mode than in the standard low-frequency (discrete) mode in all the considered tag models. This can be explained by an increased number of satellites used for the GPS fixes in the high-frequency mode (about twice as many satellites used per fix compared to the low-frequency mode; mean ± SD 12.4 ± 3.3 vs. 6.5 ± 2.0; Additional file 3: Fig. S10). The notable differences which we found in the accuracy of GPS height data between tag models, especially in low-frequency GPS data (range of mean AE 7.4–29.9 m), might partly be due to technical differences in the GPS modules used in the tags, like the application of additional global navigation satellite systems (GNSS) in addition to GPS (e.g.
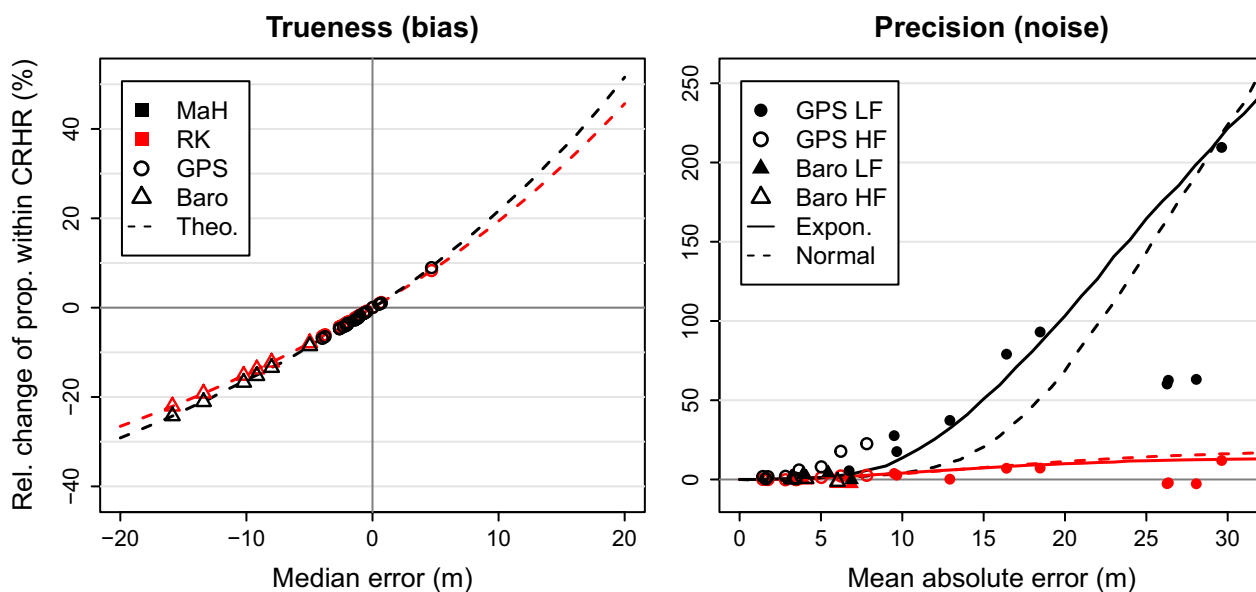
Schaub *et al. Animal Biotelemetry* (2023) 11:31

Page 11 of 16

Trueness (bias)

Precision (noise)

Fig. 7 Relative change of the proportion of positions within the collision risk height range (50–200 m) in marsh harriers (black) and red kites (red) when applying different degrees of bias (trueness) or noise (precision) to the height data. Points represent empirical error distributions found in different GPS tag models in either GPS or barometric height data from either low-frequency or high-frequency sampling. Lines represent gradually increasing bias for trueness and increasing noise based on theoretical error distributions (exponential or normal) for precision. *Expon.* exponential, *MaH* marsh harrier, *RK* red kite, *Baro* barometric, *LF* low-frequency, *HF* high-frequency

GPS+GLONASS in *Ornitela* tags as opposed to GPS only in *Milsar* tags) or different internal settings (for example regarding time-to-fix). The year of data collection could also affect the positional accuracy of GPS data, as over the years, more satellites have been added to the orbit. However, in our case, there has only been a slight increase of the number of satellites over the years (Additional file 3: Fig. S10).

There is also a large variation among results on vertical accuracy of GPS tags reported in earlier studies, and our results generally fell within these ranges. In low-frequency sampling, Bouten et al. [13] reported a mean AE in relation to true height of 20.8–26.3 m with 10-min intervals and 4.0 m with 1-min intervals, while Péron et al. [14] indicated a mean AE of 27 m with 1-min intervals, Acácio et al. [15] of 9.7 m with 60 min and 5.0 m with 1-min intervals and Heuck et al. [16] a 95% quantile of AE of 33 m (compared to 20–161 m in our data; Additional file 2: Table S4). Note that we did not consider differences in accuracy between different intervals in low-frequency GPS tracking here, as opposed to some of the studies cited. Regarding high-frequency GPS data, reference data are scarce, but Bouten et al. [13] reported mean AE of 1.4–2.8 m with 6-s intervals (compared to 1.5–7.0 m in our data) and Thaxter et al. [32] whisker ranges of 11–14 m with 10 s intervals (compared to 8–33 m in our data; Additional file 2: Table S4).

Visual inspection of height profiles of high-frequency tracks indicated some recurrent error patterns in the high-frequency GPS height data (accuracy time lag at the beginning of high-frequency sequences, spikes, drift in stationary periods). However, these concerned only a relatively small proportion of positions, or stationary positions only. The error arising from the accuracy time lag and spikes could be reduced with relatively simple methods. For example, applying a moving average with a window of nine data points to the high-frequency GPS data of *Milsar GsmTag-U9* tags reduced the 95% quantile of AE from 13.1 to 10.7 m (Additional file 4). The accuracy time lag has been also reported in earlier studies, where it was found to last 10–35 s [33, 34]. This problem can be solved by removing the first part of high-frequency sequences (Additional file 4). The finding of increasing accuracy within the first portion of high-frequency GPS sequences is in opposition to the suggestion that there could be a constant initial error that is maintained during the entire high-frequency sequence [14].

Accuracy in barometric height and pathways for improvement

Our results on vertical accuracy in barometric height data were mixed. On the one hand, in stationary positions, barometric height had a substantial bias compared to true height. On the other hand, the closeness of GPS

and barometric height in high-frequency data (Figs. 4, 6) and the similarity between flight height distributions obtained from high-frequency GPS and barometric data (Fig. 5) suggest that the accuracy of barometric height data for flight positions is relatively high, and that it indeed represents an improvement compared to low-frequency GPS data.

It has been described earlier that the vertical error in barometric height data consists to a large extent of a bias related to weather conditions and calibration, as opposed to the error dominated by random noise in the GPS height data [14, 16]. This implies also the absence of extreme outliers in the barometric height data (this study, [16]). As in our data, the bias in barometric height data from stationary positions reported by Heuck et al. [16] was negative (median of -22.6 m). Regarding the precision in barometric height data, Heuck et al. [16] reported a 95% quantile of AE in relation to median height of only 1.3 m in barometric data in a stationary experiment (tags not deployed on birds), compared to 9.5–26.5 m in our data. With drone experiments, Lato et al. [35] found a mean vertical error of only 1.6 m in barometric data and Péron et al. [14] reported a RMSE of 22 m between barometric and GPS height for low-frequency data. The latter is also considerably lower than the RMSE between the two types of data in our study (43–92 m). These differences could be explained by the longer time span during which our data were collected, implying a wider range of weather conditions in which air pressure was measured. Moreover, we cannot exclude that the increased error in barometric height data in our study resulted from the fact that we evaluated accuracy in a field setting, with tags deployed on free-living birds. For example, there might be effects of heat radiation of the birds, moisture or dirt on the pressure readings.

An important aspect of our results which has, to our knowledge, not been described earlier is that the difference between GPS height and barometric height differed systematically between stationary and flight positions. Possibly, the difference could be due to an effect of movement on the air pressure measurement, which could be related to differences in wind speed and temperature between moving and stationary states, or the fact that often the tag is partially covered by feathers when the birds are stationary, possibly impairing the measurements. The difference between stationary and flight positions implies that a correction of the barometric height data based on stationary tests [16] might not be optimal for flight positions. Moreover, the difference between GPS height and barometric height changed along the height gradient in low-frequency data, with barometric height on average exceeding GPS height for recorded barometric heights > 0–40 m a.g.l. This could be caused

by an altitudinal bias in the barometric height data. However, if this was the case, we would expect this pattern to be also present in high-frequency data. We could not exactly retrace how the altitudinal pattern arose, but the comparison of flight height distributions obtained from the different methods (Fig. 7) suggested that the distributions based on barometric data both from low- and high-frequency sampling were shifted by an approximately constant offset compared to high-frequency GPS data, at least within the height range relevant for wind turbine collision risk (below 300 m). Therefore, a correction offset based on the mean difference between GPS height and barometric height in flight positions could be a way of aligning barometric and GPS height data. However, mean GPS height during flight might not be free of bias either [35]. Experiments with drones might help to verify if this correction approach is indeed effective, but note that also in such experiments, obtaining reference data for true height in flight is not trivial (but see [35]). The correction should optimally be conducted for each tag separately, as we obtained indications that the bias in barometric data differs between individual tags (unpublished data), similarly to Heuck et al. [16]. It should be noted that even though the accuracy in barometric height data might be improved with further corrections, our results also suggest that a bias of a few metres, which probably remained in the barometric height data without corrections, might not have major implications for the proportion of positions within the collision risk height range, as opposed to extensive noise as in low-frequency GPS height data (Fig. 7).

Table 1 Overview of the sources of error in GPS and barometric height data, either regarding the determination of height above ground itself or regarding the identification of stationary periods on the ground on which we based our assessment of vertical accuracy

Source of error	GPS height	Barometric height
Determination of height above ground		
GPS height above geoid/ellipsoid	X	
Air pressure measurement		X
Weather data		X
Simplifications in barometric height formula		X
Horizontal GPS position	X	X
Digital elevation model	X	X
Identification of stationary periods on the ground		
GPS speed	X	X
Classification of stationary/flight positions	X	X
Habitat classification	X	X

Sources of error

The aim of this study was to assess the overall vertical accuracy occurring in a practical field setting. However, we do want to stress that the accuracy we described here in fact represented a combination of different sources of error, not only related to the height measurement itself, but also to the digital elevation model (DEM) and to the identification of stationary periods on the ground (occasional erroneous classification of flight positions as stationary; Table 1). Therefore, we expect that the true vertical error itself is somewhat smaller than indicated here. In addition, it has been reported that horizontal and vertical accuracy in GPS data is higher when tags are moving [34, 36]. This suggests that our estimation of accuracy based on stationary positions is conservative when transferred to flight positions.

At several stages of our analysis, we came across problems of obtaining raw data from the GPS tags. For example, the raw height above ellipsoid data were only available for one out of three manufacturers, and in the case of another manufacturer, height above geoid was truncated at zero, precluding negative values. These limitations potentially bias error assessments as conducted here, but can also have implications for analyses of flight height. Therefore, we want to call on manufacturers to make raw data (unprocessed height above ellipsoid for GPS and raw pressure measurements for barometers) available throughout, in line with Péron et al. [14].

The large differences in vertical accuracy across tag models, especially in low-frequency GPS data, and the need for correction of the barometric height data as found here, underline the importance of testing the accuracy of GPS tags. Assessing accuracy using data from tags already deployed on birds, as done here, has the advantages that it can be applied a posteriori and that it integrates the in situ conditions of data collection. However, it requires the possibility to identify periods during which the true height of the birds is approximately known, like stationary periods on the ground (this study) or on the sea surface [32], which is not possible for every species. This approach is also restricted to non-flight positions. Un-deployed tags can be tested with stationary experiments [13, 14, 16], experiments where tags are moved horizontally [34], or using drones [35]. Approaches based on stationary data have the disadvantage that results may not be fully applicable to flight data (see above). Approaches with moving tags have the disadvantage that the true height is difficult to determine, but drones with laser altimeter represent a promising new method to solve this issue [35].

Effect of error on proportion of positions within the collision risk height range

Using simulations, we showed that both bias and noise in the height data can lead to a bias in the proportion of positions within the collision risk height range (CRHR). However, the potential effect of noise was much larger than the effect of bias (up to $+210\%$ with noise compared to up to -24% with bias). Moreover, the effect of noise differed strongly between the two considered species, with a strong overestimation of the proportion within the CRHR in marsh harriers at the highest levels of noise, but only a small overestimation in red kites. This can be explained by differences both in the shape of the flight height distributions and in the location of the mode in relation to the CRHR in the two species, with a very steep distribution with a mode located relatively far from the CRHR in marsh harriers and a broader distribution with the mode being closer to the CRHR in red kites. It is important to note that extensive noise could not only lead to an overestimation of the proportion of positions within the CRHR, but also to an underestimation, most probably in cases where the mode of the flight height distribution falls within the CRHR (e.g. larger soaring birds like short-toed eagle *Circaetus gallicus*, unpublished data). This would in turn lead to an underestimation of wind turbine collision risk.

We would like to stress that low-frequency GPS data does not necessarily produce erroneous outcomes. In our empirical flight height data, the difference in the proportion of positions within the CRHR between high-frequency and low-frequency GPS data was surprisingly small in some tag model-species combinations (Additional file 2: Table S7). The effect of noise on the results will depend on (1) the level of noise in the data (which we showed to vary between tag models); (2) the true flight height distribution (e.g. marsh harrier vs. red kite) and (3) the question for which the data are applied (e.g. definition of the CRHR). However, in practice, neither the exact level of noise nor the true flight height distribution are normally known, making it difficult to predict the effect of noise on the outcomes.

Pros and cons of high-frequency GPS tracking and barometric altimetry

Our study showed that the use of high-frequency GPS tracking resulted in the highest vertical accuracy amongst the considered methods. Additionally, this method provides the advantage of an increased horizontal accuracy ([13]; Table 2). Moreover, the high temporal resolution enables the use of high-frequency GPS tracking data for detailed analyses of 3D flight trajectories with many potential applications, e.g. regarding habitat use and foraging behaviour [37] or the use of thermal uplifts [38,

Schaub *et al. Animal Biotelemetry*   (2023) 11:31

Page 14 of 16

Table 2 Advantages and disadvantages of high-frequency GPS tracking and barometric altimetry, compared to standard low-frequency GPS tracking

	High-frequency GPS tracking	Barometric altimetry
Advantages		
Increased vertical accuracy	X	X
Increased horizontal accuracy	X	
Potential for detailed trajectory analyses	X	
Disadvantages		
High battery demand	X	
Intensive management of tag settings required	X	
Sampling only during restricted time periods	X	
Sampling bias towards periods of good solar charging conditions	X	
Weather data required		X
Reduced accuracy in unstable weather conditions		X
Increase of tag weight		X

39]. In particular, the study of wind turbine avoidance by birds requires high positional accuracy both in the horizontal and vertical dimension, and reliable information on this aspect is urgently needed to improve the predictions of mortality from wind turbine collisions [5]. High-frequency GPS tracking could play an important role to fill this knowledge gap [30, 32].

The main disadvantage of high-frequency GPS tracking is the high battery demand, which implies that this type of data can only be collected during restricted time periods. The collection of high-frequency GPS data depends on solar charging conditions, which poses the risk of a sampling bias by an underrepresentation of circumstances with poor solar charging, for example in relation to time of day, weather, season or sex (underrepresentation of females due to reduced movement during the breeding season). However, whenever representative results on the vertical niche of a bird species are required, it is important to sample across the aforementioned variables in an unbiased way. Note that the extent of the problem of battery demand and sampling bias might depend on the behaviour of the study species (e.g. depending on time spent flying and habitat) and the climatic conditions in the study area (e.g. less problematic in tropical areas).

The application of high-frequency GPS tracking has been facilitated by the possibility of remotely modifying tag settings, mainly through the GSM network in recent tag models. However, to date, the monitoring of battery voltage levels and the activation of the high-frequency mode often have to be performed manually, which requires a considerable time investment on a daily basis and might discourage researchers from applying high-frequency settings. Note that the automatic initiation of the high-frequency mode when battery voltage reaches a

defined threshold is already an available option in some manufacturers at present, but this potentially leads to a strong bias towards good solar charging conditions. In this respect, it would be a considerable step forward if tag manufacturers could provide more complex programming options for tag settings (for example, when a defined voltage threshold is reached, scheduling a one-hour sequence of high-frequency sampling for a random time on the next day). Another example of a promising avenue in this context is automatic flight detection, i.e. automatic application of high-frequency tracking when the bird is in flight, and low-frequency tracking when the bird is stationary, which is already available in some tag models [34, 39], but not yet fully efficient for all bird species (unpublished data).

Barometric data have the advantage of reduced battery demand compared to high-frequency GPS data (Table 2). In fact, barometric measurements are recorded alongside every GPS fix with a negligible increase in battery consumption. This makes it much easier to obtain flight height data without the aforementioned sampling biases. A disadvantage of barometric altimetry is the additional weight of the pressure sensor. For example, the lightest GPS tag with pressure sensor of the manufacturer *Ornitela* currently weighs 20 g, preventing its use on smaller species such as Montagu's and hen harrier.

Conclusions

The recent advancements of the GPS tracking technology have opened many opportunities for the study of animal movements. However, it has remained challenging to obtain accurate flight height data from GPS tags. At the same time, this data is urgently needed to accurately predict the collision risk of birds with wind

Schaub *et al. Animal Biotelemetry*   (2023) 11:31

Page 15 of 16

turbines and identify effective mitigation measures. Based on a field assessment using data from GPS tags deployed on free-living birds, we confirmed that GPS height data from standard low-frequency GPS tracking is associated with substantial error, blurring flight height distributions and potentially leading to an important bias in parameters relevant for bird conservation. Barometric altimetry may provide more accurate height data, but there is the risk of a systematic error which is difficult to resolve fully. Dedicated experiments are needed, especially elucidating the behaviour of barometric height in relation to movement (stationary vs. flying), to derive an effective correction method for barometric height data. Most importantly, we showed that high-frequency (continuous-mode) GPS tracking substantially improves vertical accuracy compared to low-frequency (discrete-mode) GPS tracking. It can be seen as a complementary approach to statistical modelling techniques accounting for the vertical error [12, 14]. Moreover, it has the additional advantage that it enables detailed 3D trajectory analyses, notably with respect to wind turbine avoidance. However, care should be taken to collect the high-frequency data in an unbiased, representative way.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40317-023-00342-1.

> **Additional file 1**. Supplementary details on methods.
>
> **Additional file 2**. Supplementary results (tables).
>
> **Additional file 3**. Supplementary results (figures).
>
> **Additional file 4**. Additional treatment of high-frequency GPS data.

Author contributions

TS, AMil, RB, SL, AMio and RK did fieldwork. TS, AMil, CZ and RK conceptualised the study. TS performed data processing and data analysis, with support from AMil and JC. TS and RK wrote the first draft. All authors contributed to subsequent drafts and approved the final manuscript.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

GPS tracking was approved by the responsible authorities CRBPO (personal projects 633 and 987), Wageningen University & Research (project 2017.D-0045.004), University of Groningen (permits 5869B and 6429B) and LAVES Niedersachsen (permit 33.12-42502-04-14/1550), and was performed following all relevant guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Aix Marseille Univ, CNRS, IRD, Avignon Univ, Institut Méditerranéen de Biodiversité et d'Écologie, BP 80, 13545 Aix-en-Provence, France. [2] ENGIE Lab CRIGEN, 4 rue Joséphine Baker, 93240 Stains, France. [3] Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, PO Box 11103, 9700 CC Groningen, The Netherlands. [4] Dutch Montagu's Harrier Foundation, Berkenweg 1, 9471 VA Zuidlaren, The Netherlands. [5] Groupe d'Études et Protection des Busards, Maison Forestière de Blinfey, 52110 Beurville, France. [6] Wageningen Environmental Research, Droevendaalsesteeg 3A, 6708 PB Wageningen, The Netherlands. [7] Natural England, Sterling House, Dix's Field, Exeter, UK. [8] Centre for Ecology and Conservation, University of Exeter, Penryn, UK. [9] LPO Champagne-Ardenne, Der Nature, Ferme des Grands Parts, 51290 Outines, France.

References

1. De Lucas M, Perrow MR. Birds: Collision. In: Perrow MR, editor. Wildlife and wind farms, conflicts and solutions. Volume 1: Onshore: Potential effects. Exeter: Pelagic Publishing; 2017. p. 155–90.
2. Thaxter CB, Buchanan GM, Carr J, Butchart SHM, Newbold T, Green RE, et al. Bird and bat species' global vulnerability to collision mortality at wind farms revealed through a trait-based assessment. Proc R Soc B Biol Sci. 1862;2017(284):20170829.
3. Bellebaum J, Korner-Nievergelt F, Dürr T, Mammen U. Wind turbine fatalities approach a level of concern in a raptor population. J Nat Conserv. 2013;21(6):394–400.
4. Carrete M, Sánchez-Zapata JA, Benítez JR, Lobón M, Donázar JA. Large scale risk-assessment of wind-farms on population viability of a globally endangered long-lived raptor. Biol Conserv. 2009;142(12):2954–61.
5. Chamberlain DE, Rehfisch MR, Fox AD, Desholm M, Anthony SJ. The effect of avoidance rates on bird mortality predictions made by wind turbine collision risk models. Ibis. 2006;148:198–202.

Schaub *et al. Animal Biotelemetry*   (2023) 11:31

Page 16 of 16

6. Masden EA, Cook ASCP, McCluskie A, Bouten W, Burton NHK, Thaxter CB. When speed matters: the importance of flight speed in an avian collision risk model. Environ Impact Assess Rev. 2021;90:106622.

7. Hull CL, Muir SC. Behavior and turbine avoidance rates of eagles at two wind farms in Tasmania, Australia. Wildl Soc Bull. 2013;37(1):49–58.

8. Johnston A, Cook ASCP, Wright LJ, Humphreys EM, Burton NHK. Modelling flight heights of marine birds to more accurately assess collision risk with offshore wind turbines. J Appl Ecol. 2014;51(1):31–41.

9. Fijn RC, Krijgsveld KL, Poot MJM, Dirksen S. Bird movements at rotor heights measured continuously with vertical radar at a Dutch offshore wind farm. Ibis. 2015;157(3):558–66.

10. Aschwanden J, Stark H, Peter D, Steuri T, Schmid B, Liechti F. Bird collisions at wind turbines in a mountainous area related to bird movement intensities measured by radar. Biol Conserv. 2018;220:228–36.

11. Pfeiffer T, Meyburg BU. Flight altitudes and flight activities of adult Red Kites (*Milvus milvus*) in the breeding area as determined by GPS telemetry. J Ornithol. 2022;163(4):867–79.

12. Ross-Smith VH, Thaxter CB, Masden EA, Shamoun-Baranes J, Burton NHK, Wright LJ, et al. Modelling flight heights of lesser black-backed gulls and great skuas from GPS: a Bayesian approach. J Appl Ecol. 2016;53(6):1676–85.

13. Bouten W, Baaij EW, Shamoun-Baranes J, Camphuysen CJ. A flexible GPS tracking system for studying bird behaviour at multiple scales. J Ornithol. 2013;154(2):571–80.

14. Péron G, Calabrese JM, Duriez O, Fleming CH, García-Jiménez R, Johnston A, et al. The challenges of estimating the distribution of flight heights from telemetry or altimetry data. Anim Biotelem. 2020;8(5):1–13.

15. Acácio M, Atkinson PW, Silva JP, Franco AMA. Performance of GPS/GPRS tracking devices improves with increased fix interval and is not affected by animal deployment. PLoS ONE. 2022;17(3):e0265541.

16. Heuck C, Sommerhage M, Stelbrink P, Höfs C, Geisler K, Gelpke C, et al. Untersuchung des Flugverhaltens von Rotmilanen in Abhängigkeit von Witterung und Landnutzung unter besonderer Berücksichtigung vorhandener Windenergieanlagen im Vogelschutzgebiet Vogelsberg - Abschlussbericht. Im Auftrag des Hessischen Ministerium für Wirtschaft, Energie, Verkehr und Wohnen; 2019. https://landesplanung.hessen.de/sites/landesplanung.hessen.de/files/2022-11/flugverhaltenrotmilan_abschlussbericht_200206.pdf. Accessed 29 Mar 2023.

17. Anderson D, Arkumarev V, Bildstein K, Botha A, Bowden CGR, Davies M, et al. Practical guide to methods for attaching research devices to vultures and condors. Vulture News. 2020;78a:1–72.

18. R Core Team. R: a language and environment for statistical computing. R version 4.0.3 (2020-10-10). Vienna, Austria: R Foundation for Statistical Computing; 2020.

19. Kölzsch A. Extract Two Movement Speeds v1.1 (6). MoveApps; 2022. https://github.com/movestore/Flight-Speed-Estimation.git. Accessed 10 Jan 2023.

20. Agisoft. Global Geoid Models. https://www.agisoft.com/downloads/geoids. Accessed 2 Mar 2023.

21. ISO. ISO 2533-1975. Standard atmosphere. International Organization for Standardization; 1975.

22. Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, et al. ERA5 hourly data on pressure levels from 1959 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS); 2018. https://doi.org/10.24381/cds.bd0915c6. Accessed 1 Nov 2022.

23. Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, et al. ERA5 hourly data on single levels from 1959 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS); 2018. https://doi.org/10.24381/cds.adbb2d47. Accessed 1 Nov 2022.

24. Dodge S, Bohrer G, Weinzierl R, Davidson S, Kays R, Douglas D, et al. The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data to facilitate research of external factors effects on movement. Mov Ecol. 2013;1:1–14.

25. EEA. European Digital Elevation Model (EU-DEM), version 1.1. European Environment Agency; 2017. https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1. Accessed 6 Feb 2022.

26. IGN. BD TOPO®. https://geoservices.ign.fr/bdtopo. Accessed 15 Sept 2022.

27. PDOK. Dataset: Basisregistratie Topografie (BRT) TOPNL en TOPraster Historie. https://www.pdok.nl/downloads/-/article/basisregistratie-topografie-brt-historie. Accessed 14 Dec 2022.

28. OIML. International vocabulary of metrology - Basic and general concepts and associated terms (VIM), 3rd ed. International Organization of Legal Metrology; 2012.

29. Ren S, Lai H, Tong W, Aminzadeh M, Hou X, Lai S. Nonparametric bootstrapping for hierarchical data. J Appl Stat. 2010;37(9):1487–98.

30. Schaub T, Klaassen RHG, Bouten W, Schlaich AE, Koks BJ. Collision risk of Montagu's Harriers *Circus pygargus* with wind turbines derived from high-resolution GPS tracking. Ibis. 2020;162(2):520–34.

31. Band W, Madders M, Whitfield DP. Developing field and analytical methods to assess avian collision risk at wind farms. In: de Lucas M, Janns GFE, Ferrer M, editors. Birds and wind farms risk assessment and mitigation. Madrid: Quercus; 2007. p. 259–75.

32. Thaxter CB, Ross-Smith VH, Bouten W, Masden EA, Clark NA, Conway GJ, et al. Dodging the blades: new insights into three-dimensional area use of offshore wind farms by lesser black-backed Gulls *Larus fuscus*. Inter-Res Mar Ecol Prog Ser. 2018;587:247–53.

33. Corman A-M, Garthe S. What flight heights tell us about foraging and potential conflicts with wind farms: a case study in Lesser Black-backed Gulls (*Larus fuscus*). J Ornithol. 2014;155:1037–43.

34. Grünkorn T, Welcker J. Erhebung von Grundlagendaten zur Abschätzung des Kollisionsrisikos von Uhus an Windenergieanlagen im im nördlichen Schleswig-Holstein. BioConsult SH; 2019. https://bioconsult-sh.de/site/assets/files/1803/endbericht_uhutelemetrie_20191128.pdf. Accessed 29 March 2023.

35. Lato KA, Stepanuk JEF, Heywood EI, Conners MG, Thorne LH. Assessing the accuracy of altitude estimates in avian biologging devices. PLoS ONE. 2022;17(10):1–20.

36. Zheng Z, Rasouli S, Timmermans H. Evaluating the accuracy of GPS-based taxi trajectory records. Procedia Environ Sci. 2014;22:186–98.

37. Schlaich AE, Klaassen RHG, Bouten W, Both C, Koks BJ. Testing a novel agri-environment scheme based on the ecology of the target species, Montagu's Harrier *Circus pygargus*. Ibis. 2015;157:713–21.

38. Duriez O, Peron G, Gremillet D, Sforzi A, Monti F. Migrating ospreys use thermal uplift over the open sea. Biol Lett. 2018;14(12):20180687.

39. Harel R, Horvitz N, Nathan R. Adult vultures outperform juveniles in challenging thermal soaring conditions. Sci Rep. 2016;6:27865.

Publisher's Note