Ornis Fennica 98: 33–45. 2021

Counting breeding gulls with unmanned aerial vehicles: camera quality and flying height affects precision of a semi-automatic counting method

Alejandro Corregidor-Castro*, Thomas E. Holm & Thomas Bregnballe

*A. Corregidor-Castro, Department of Bioscience, Aarhus University, Grenåvej 14, DK–8410, Rønde, Denmark * Corresponding author's e-mail: alecorregidor@gmail.com*
T.E. Holm, Department of Bioscience, Aarhus University, Grenåvej 14, DK–8410, Rønde, Denmark
T. Bregnballe, Department of Bioscience, Aarhus University, Grenåvej 14, DK–8410, Rønde, Denmark

Received 17 February 2020, accepted 31 January 2021

The use of Unmanned Aerial Vehicles (UAVs) to monitor large colonies of seabirds avoids challenges associated with conventional methods, but manual image processing is expensive. Development of semi-automated analytical methods rely on high image spatial resolution, which requires a trade-off between securing low area coverage and high spatial resolution flying at low altitude vs high area coverage but low spatial resolution flying at higher altitudes. Increasing individual bird detection probabilities requires maximizing contrast between target and background, which can be enhanced using thermal sensors. We applied a semi-automatic analytical method to multispectral UAV derived imagery to count a mixed breeding colony of Herring Gulls (*Larus argentatus*) and Lesser Black-backed Gulls (*L. fuscus*). We trained the computer to detect different image classes by their spectral signature in several orthomosaics obtained from UAV flights at different altitudes using different cameras. Highest agreement with the manual counts was achieved by low flying (20 m) using the highest camera resolution (97.7 ± 1.1% for the Herring Gulls, omission error 2.6%, commission error 0.5%; 94.8 ± 1.8% for Lesser Black-backed Gulls, omission error 6.5%, commission error 1.6%). Method precision varied between trials, confirming the importance of low altitude flying with high quality cameras, and a 40% reduction in detection noise from adding a thermal sensor.

1. Introduction

Monitoring annual abundance of breeding seabirds is fundamental for conservation and long-term studies, as well as tracking ecosystem health by observing population trends (Descamps *et al.* 2011, Chabot & Francis 2016, Lyons *et al*. 2019a). Conventional monitoring of breeding seabirds uses methods such as line transects or point counts. However, the efficacy of such methods relies on observer skills, uncertainties of extrapolations and difficulties regarding the access to view breeding colonies (Brisson-Curadeau *et al.* 2017, Afán *et al.* 2018, Magness *et al.* 2019). High disturbance

levels to breeding birds (Grenzdörffer 2013) may adversely affect individual breeding success or even cause colony failure (Sardà-Palomera *et al.* 2017, Fuller *et al.* 2018). To avoid such bias and impacts, methods to assess abundance and distribution must generate highly accurate data without affecting individual bird behaviour (Sardà-Palomera *et al.* 2017).

Unmanned aerial vehicles (UAVs), popularly known as drones, potentially reduce such bias and impacts. Drone surveys can cover large areas during short duration flights (depending on altitude, e.g., covering a 12.2 ha waterbird colony in 30 minutes; Afán *et al.* 2018, Valle & Scarton 2019) providing high-resolution spatio-temporal data from multispectral sensors (López & Mulero-Pázmány 2019) in inaccessible or dangerous areas (Brisson-Curadeau *et al*. 2017, Inman *et al*. 2019, Valle & Scarton 2019). Drones have now been extensively used to monitor birds (Fretwell *et al.* 2012, Grenzdörffer 2013, Chabot *et al.* 2015, Rush *et al.* 2018, Pfeifer *et al.* 2019). Their use minimizes or eliminates disturbance to birds (Chabot & Francis 2016, Afán *et al.* 2018), producing highly accurate information about numbers of nesting birds and other breeding parameters (Sardà-Palomera *et al.* 2011, Weissensteiner *et al.* 2015, Sardà-Palomera *et al*. 2017).

Application of drone technology, however, requires substantial post-survey data processing time (Descamps *et al.* 2011, Inman *et al.* 2019). Traditionally, this has been conducted manually; counting birds and/or nests on images, which is highly time-consuming, and in some cases may negate efficiency gains from data collection (Chabot & Francis 2016, Lyons *et al.* 2019b). Widespread adoption of such new technologies is dependent on their exceeding both the efficiency and accuracy of traditional methods (Hodgson *et al.* 2018). Towards this aim, different automated and semi-automated approaches to image analysis have been developed (Hollings *et al.* 2018, Rush *et al.* 2018), which have reduced processing times up to eight-fold (Lyons *et al.* 2019b).

Typically, these (semi-)automated methods use point process algorithms, spectral thresholding, and combinations of spectral properties and predictive modelling (see Lyons *et al.* 2019b), to differentiate the targets (i.e., breeding birds) from the remaining elements of the image. One of the main drivers for automatic detection in aerial imagery is the spatial resolution of the image. To achieve the most efficient species identification in object-oriented approaches, it is recommended to work at an image resolution of 0.5–1 cm/px (Grenzdörfer 2013, Dulava *et al.* 2015, Afán *et al.* 2018, Barr *et al.* 2018, Rush *et al*. 2018). However, this level of resolution requires the surveyor to trade-off flying at lower altitude to increase image spatial resolution against flying higher and covering larger areas (Inman *et al.* 2019). The option of flying at low altitude and thereby getting images of high spatial resolution may become constrained by battery life (especially if there is limited access to spare batteries or to charge batteries in the field) and the lower threshold of flying height below which the target species (or other species) change behaviour or even leave the nest due to the possible disturbance produced by the UAV.

To use the spectral properties, the imagery needs to generate consistently high contrast between target objects and their background, i.e., the spectral signature of the target needs to be constant (i.e., in shape and colour, Hollings *et al.* 2018, Lyons *et al.* 2019b) and stand out from the rest of the image (Chabot & Francis 2016). A common challenge is to differentiate elements with similar spectral signatures in the image, for instance mistaking pale rocks for bright birds (Chabot & Francis 2016; Lee *et al*. 2019). In such situations, thermal sensors offer the potential to increase individual detection probability (since endotherms tend to be warmer objects than their surroundings, i.e., polar bears and sea ice, Burke *et al.* 2019, Chabot *et al.* 2019). Despite falling costs, this technology remains expensive (Kays *et al.* 2019), and there are still hardware and software challenges due to the novelty of the technology (Scholten *et al.* 2019). Low resolution of thermal sensors compared to RGB cameras (Kays *et al.* 2019) and difficulties recognizing species using thermal images alone also restricts their application (Goodenough *et al.* 2018). Use of thermal sensors thus needs a complementary RGB sensor (Kays *et al.* 2019), especially when applied where several species breed in sympatry (Chabot & Francis 2016, Afán *et al.* 2018).

In this study, we conducted a drone-based

survey and applied semi-automated methods to imagery to count a mixed breeding colony of Herring Gulls (*Larus argentatus*) and Lesser Black-backed Gulls (*L. fuscus*) located on an uninhabited island in the Danish part of the Wadden Sea. Using ArcMap 10.5, we trained the computer to detect different elements based on their spectral signature in several orthomosaics obtained from a series of UAV flights at different altitudes. We tested the effect of camera quality on the efficacy of applying the semi-automatic counting of gulls and whether adding a thermal infrared (TIR) band to the method improved the differentiation of the target elements from similar objects in the surroundings.

2. Material and methods

2.1 Study site

We collected the UAV imagery in May 2018 and 2019 on Langli (55° 30' N, 8° 18' E) located in Ho Bugt close to the city of Esbjerg, Denmark (Fig. 1), a 80 ha island, *c.* 2 km long and 0.5 km wide at its widest. Langli is a part of an internationally important area for waterbirds protected under the Ramsar Convention and the EU Birds Directive. The island is uninhabited, open to public access only between 16 July and 15 September. Langli is flat with marshy areas at the northern and southern ends, but mostly covered in dunes rising to 14–16 m high in the center of the island. Dry marsh areas have a high coverage of Common Velvet Grass (*Holcus lanatus*) and Red Fescue (*Festuca rubra*) with Common Saltmarsh Grass (*Puccinellia maritima*) and Sea Lavender (*Limonium vulgare*) dominating wetter areas. Herring Gull (HG) has bred on the island since before 1900, whereas the Lesser Black-backed Gull (LBBG) population has only established since the late 1970's (Møller 1978). Both HG and LBBG are common in Denmark and most of Europe. The Danish populations of HG and LBBG are estimated to number c. 90,000 and c. 5,000 pairs, respectively (Bregnballe & Lyngs 2014, T. Bregnballe, unpubl. data), often breeding in mixed colonies (Calladine 1997) as is the case on Langli.

2.2 Drone survey and protocol

We conducted several flights with three different drones to test the semi-automatic counting method. On 30 May 2018, we used a fixed wing UAV (Event 38 model E384) with a Sony ILCE-QX1 camera flying for 2.5 hours to cover the entire island at 50 m altitude (1 battery change). In 2019, we flew two types of quadcopters, a DJI Matrice 210 and a DJI Phantom 4 Pro. The DJI Matrice 210 was flown with two different cameras, a DJI Zenmuse X5S with a DJI MFT 15 mm F/1.7 ASPH lens, and a DJI Zenmuse XT2 TIR camera

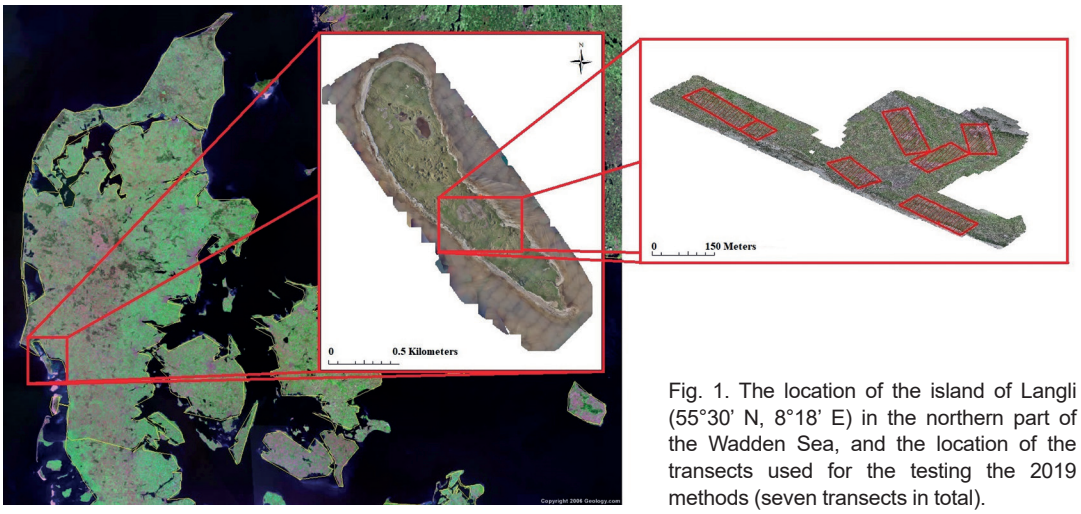

Fig. 1. The location of the island of Langli (55°30' N, 8°18' E) in the northern part of the Wadden Sea, and the location of the transects used for the testing the 2019 methods (seven transects in total).

19 mm lens including a 4k visual camera (dual sensor). The DJI Phantom 4 Pro had a camera with a standard 8.8 mm / 24 mm (35 mm format equivalent) F / 2.8–F / 11 lens, thus achieving a lower image quality than the one equipped in the DJI Matrice 210.

We first conducted a survey with the DJI Matrice 210 Zenmuse X5S at 20 m altitude on 14 May 2019 from 17:33 to 18:47 (74 min, 2935 images, 2 battery changes). On 15 May 2019 from 07:14 to 09:52, we flew the DJI Phantom 4 Pro at 15 m (158 min, 3967 images, 5 battery changes). We flew the DJI Matrice 210 with the XT2 camera, i.e., with both RGB and TIR cameras, at 50 m altitude on 15 May 2019 from 04:56 to 06:42 (106 min, 1901 dual images, 4 battery changes), to test how the addition of the TIR sensor affects the detection of gulls. All the information regarding the different flights has been summarized in Appendix I.

The take-off and landing site for each of the UAV surveys was located close to the old field station on Langli (55° 30' 43.7" N, 8° 18' 58.6" E). This location allowed us to control the UAVs from a position (> 100 m) where we avoided any unnecessary disturbance to the colonies (Vas *et al.* 2015), while keeping drones within the visual line of sight (VLOS) of the pilot. Flights were conducted during the breeding season just before hatching, to increase the probability that one of the parents would be lying down on a nest with eggs. We decided to avoid flying below 15 m because earlier studies experienced that the behaviour of HG and LBBG appeared unaffected by UAV survey flights as long as the UAV was flying at 15 m height or more (Holm *et al.* 2018, Rush *et al*. 2018, Holm *et al*. 2019).

2.3 Flight control and image processing

In 2018, we used the Mission Planner open source autopilot (Ardupilot 2019) as ground station software to control the flight. Photos were taken automatically with an overlap of 65–75% and stitched using Drone2Map from ESRI (ESRI 2019). In 2019, we used DroneDeploy Pro to both control the flight and process the imagery (dronedeploy.com). With this software, we could upload the transect images and stitch them together in groups of up to 1,000 pictures. Once images were uploaded and processed, the stitched images were exported as an orthomosaic TIF-file. As expected, the final resolution of the images varied according to survey altitude. For the RGB pictures at 50 m, the final resolution was 1.20 cm/px, for the TIR imagery it was 7.10 cm/px, and for the 20 and 15 m flights, the final resolution was 0.53 cm/px.

2.4 Supervised classification

In order to test how drone imagery could be processed effectively, we attempted the use of a semi-automatic method for counting individuals and differentiating species, as suggested by Rush *et al.* (2018). To develop this method, we conducted a supervised classification. This kind of classification is based on the creation of a training set, in which a sample of pixels of the image are assigned to different classes. These classes are then differentiated from each other by the unique spectral signature of the pixels that are contained in them, thus creating well-defined classes of fundamental importance for classification accuracy (Foody & Mathur 2006, Richards & Jia 2006).

For the first round of classifications, we selected six random areas (100 × 60 m) from the large orthomosaic obtained from the Langli flight of the fixed-wing drone in 2018. In 2019, we added an extra spectral band to the analysis by flying the DJI Matrice 210 equipped with both RGB and TIR cameras over previously ground counted transects at 50 m (N = 7; Fig. 1) to compare with imagery from RGB cameras at 15 m (DJI Phantom 4 Pro) and at 20 m (DJI Matrice 210). These transects were individually delimited with four cones (one on each corner of the seven transects) that were used as Ground Control Points (GCPs) to georeference the imagery. We analyzed all of these images following the same semi-automatic procedure.

2.4.1 Supervised classification (RGB)

For each of the six selected areas from 2018, we established a training set using the *training sample manager tool* in ArcGIS v10.5 (ESRI 2019). We

obtained the specific pixels of each gull species using polygons and we assigned them to type classes each as follows: i) heads, ii) HG shaded back, iii) HG bright back, iv) LBBG shaded back, v) LBBG bright back. We analyzed the histograms and scatterplots per band and class, and retested and retrained each class when overlapping of these parameters was present, in order to increase the accuracy of the method. Following this process, we used the *maximum likelihood tool* to conduct the supervised classification of the different classes (Fretwell *et al.* 2012, Grenzdorffer 2013, Rush *et al.* 2018) with a reject fraction of 0.01, which was enough to exclude the majority of unwanted pixels. The signature set was specific to each area as this specificity increased the ability to recognize individuals (Rush *et al*. 2018). We converted the classification output into polygons in order to operate each of the classes individually, choosing the "gull candidates" by selecting the polygons classified as "backs" and their proximity to the polygons classified as "heads", as well as to each other ("shaded backs" together with "bright backs"). By gathering these polygons together, we could create "gull-like" polygons of known area and perimeter, where the category of the predominant pixels determined the gull species. Polygons with unrealistic values were subset after assessing a repeatable threshold (Afán *et al.* 2018). Finally, we created a 50 cm buffer around the potential candidates, which facilitated a later manual check (Rush *et al*. 2018). We used the model builder tool in ArcGIS to reduce the user time spent in the process.

The shapefile derived from this processing was considered the automatic counting method (AT) output. With this shapefile, we then manually identified the different objects to confirm that they were birds and correctly assigned to species level. We then validated the AT method output by testing for accuracy (the proportion of birds detected correctly by the processing), commission errors (proportion of birds counted automatically but not confirmed in the manual process) and omission errors (proportion of birds present that were not detected automatically) as in Afán *et al.* (2018). After this process, we manually edited the AT output shapefile to delete the false positives, from which we obtained the semi-automatic output (SA). The process to obtain the SA from the AT output was fast to complete, and did not require expertise on the subject.

In order to compare the efficiency of both the AT and SA method, we first manually counted each of the test areas to identify the exact number of gulls and the species. Thus, the success and accuracy of the classification methods was determined by comparison between these manual counts and the AT and SA outputs. It took us approximately 2 hours to process each orthomosaic, and most of the time the method did not require human intervention. Whereas larger orthomosaics would require more detailed training samples due to an increase in variability within the image, they could be potentially processed over a similar time span, depending on the processing capacity of the computer used, highly reducing the human effort required for the process when compared to a throughout manual count of the same image.

2.4.2 Supervised classification (RGB+TIR)

We tested for accuracy improvement in counting and species recognition by conducting an additional classification of the TIR imagery. We first converted the TIR imagery to a 3-band color layer, where warmer objects were represented with white colors, and colder objects with blue-purple colors (Fig. 2). The classification was similar to the supervised classification explained above, but adding an extra step: after obtaining all "gull-like" candidates prior to generating the initial AT output from the RGB image, we created an extra training sample exclusively from the TIR layer. In this layer, we ran a supervised classification to detect the heat signatures in the thermal imagery. We then selected the candidates that were located within a given distance of a heat signature, and subtracted unrealistic values, in order to reduce the identification noise produced by already warmed stones or sand banks. Even though the imagery matched fairly well, the matching of both images was not 100% accurate due to the use of a dual sensor (with the spatial and spectral collection differences that it involves). Thus, we selected candidates that were located in a buffer radius of 50 cm from a heat signature (less than the normal distance between individuals). Because of the non-problematic detection of HG by the RGB

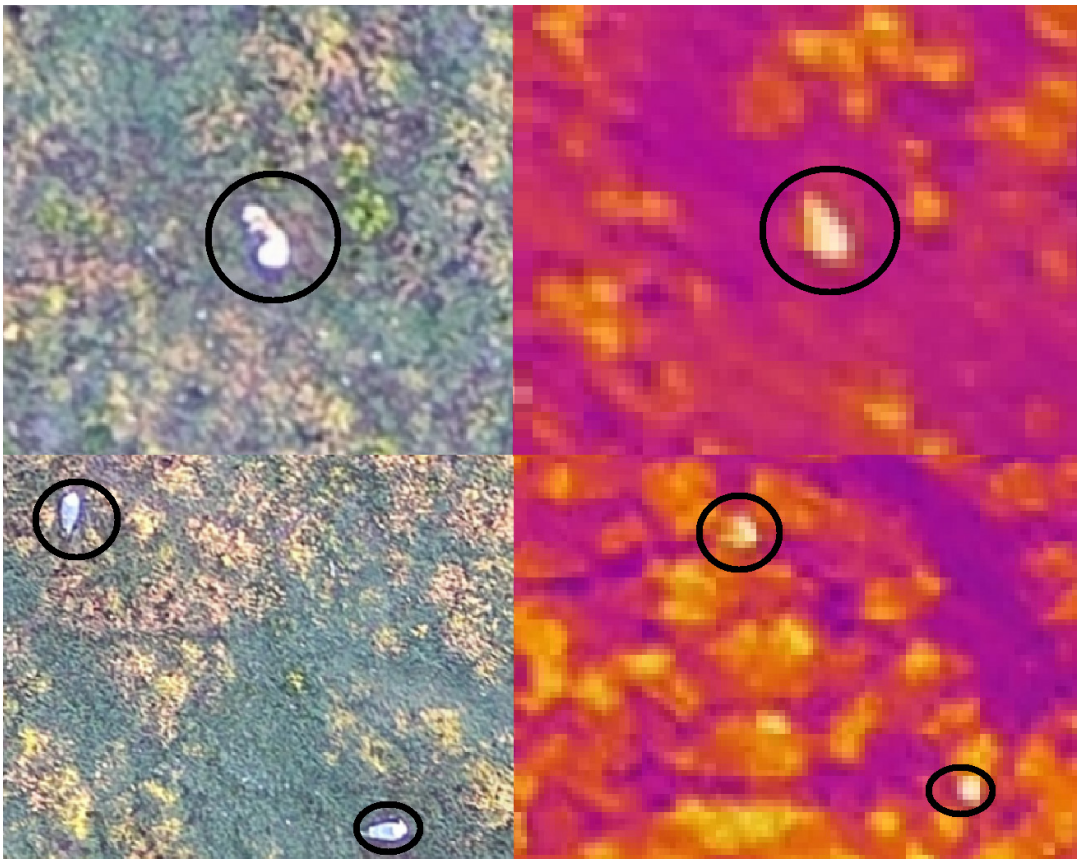

Fig. 2. Example of the combined imagery for both the RGB camera (left), and the transformed TIR imagery (right), where temperature values are interpreted as a 3-band layer (warmer objects are represented in white, colder colors in purple-blue). Individual gulls are located inside the black circles for easier interpretation.

cameras, we focused the semi-automatic detection effort of RGB + TIR on the more conspicuous LBBG.

2.5 Methodology, factors affecting the classification

Battery time and thus flying time is often a constraint in UAV surveys (Grenzdörffer 2013). We therefore tested the effect of flying at higher altitudes (and thereby reducing total flying time) on the precision of the analyses of the images. To do so, we used the orthomosaics obtained from the described flights in altitudes of 15, 20 and 50 m and conducted the AT and SA classifications (see above) and compared their agreement for both species with the manual counts.

2.6 Statistical analysis

We used a linear regression to compare the manual counts with the birds counted by the different methods on the aerial photos. We tested our variables for normality using the Shapiro–Wilk test ($p > 0.05$). We examined not normally distributed variables under nonparametric tests, and used nonparametric correlations (Spearman's rank) between AT/SA outputs and manual counts when necessary. As the independent variable, we used the number of birds counted with the method (AT/SA), whereas we used the number of birds counted manually as the dependent variable. To test for differences in the mean agreement with the manual counts of the semi-automatic counting method between the RGB and the RGB + TIR, we used a paired *t*-test.

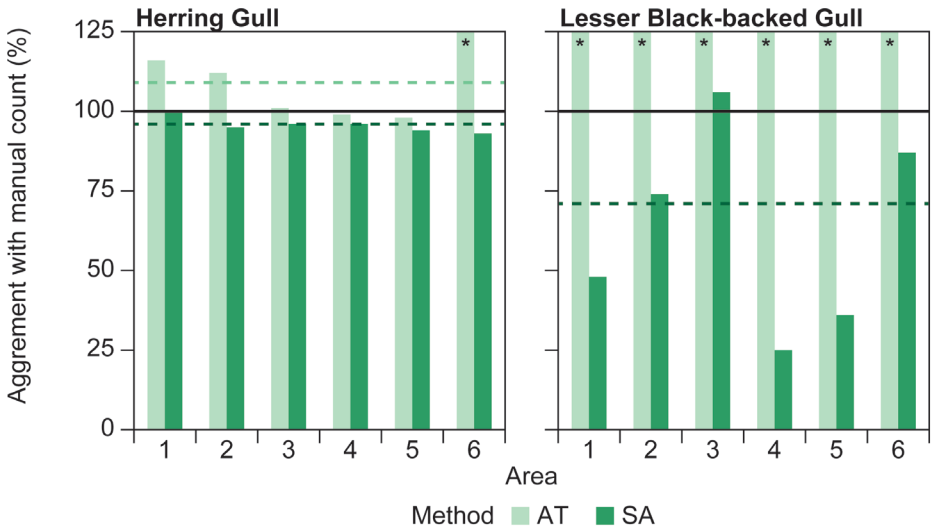

Fig. 3. Comparison of counts of gulls with the automatic method (AT, light green) and the semi-automatic method (SA, dark green) with the manual counts for Herring Gull (left) and Lesser Black-backed Gull (right). The black line represents a 100% agreement with the manual count, whereas the dark green dashed line represents the mean agreement for the SA method (96% for the Herring Gull; 71% for the Lesser Black-backed Gull), and the light green dashed line represents the mean agreement for the AT method (109% for Herring Gull; Lesser Black-backed Gull not represented). Bars presented with an asterisk (*) represent values higher than 125% of the manual counts. The data for this figure was obtained from the fixed wing UAV flight in 2018 at 50 m altitude.

We carried out the statistical analyses using the open-source program R v3.5.1 (R Core Team 2018), using the RStudio integrated development environment (IDE, v.0.99.903).

3. Results

3.1 Supervised classification (RGB)

The results of the supervised classification differed between the two species of gulls. For the HG, the automatic process (AT) provided an average agreement of 109 ± 4.8% (98–127%, Fig. 3) with a positive relationship between manual and automatic counts ($F_{1,4}$ = 25.14, p < 0.007, R^2 = 0.83). The semi-automatic process (SA) provided a mean agreement of 95.7 ± 0.9% (93–100%, Fig. 3), and a strong positive relationship with the SA for counting HG ($F_{1,4}$ = 255.8, p < 0.001, R^2 = 0.98). For the LBBG, the automatic process (AT) provided an average agreement of 2208 ± 785.7% (691–5,863%, Fig. 3) with no relationship between manual and automatic counts ($F_{1,4}$ = 0.05, p = 0.834, R^2 = –0.23). In LBBG the semi-automatic process (SA) provided an average agreement of 70.8 ± 4.8% (48–81%, Fig. 3), and a positive relationship between manual and automatic counts ($F_{1,4}$ = 25.13, p < 0.007, R^2 = 0.83) .

3.2 Supervised classification (RGB+TIR)

The addition of a TIR camera to the semi-automatic counting method reduced the total number of candidates chosen by the supervised classification by 39.4 ± 8.8% (t_5 = 2.69, p = 0.04; Fig. 4A). However, the addition of the thermal band led to a reduction in the mean agreement from 88.9 ± 3.6% in the RGB imagery to 82.2 ± 5.4% (t_5 = 3.22, p = 0.02; Fig. 4B).

3.3 Factors affecting the classification

The accuracy of the semi-automatic counting method and the relation between manual and

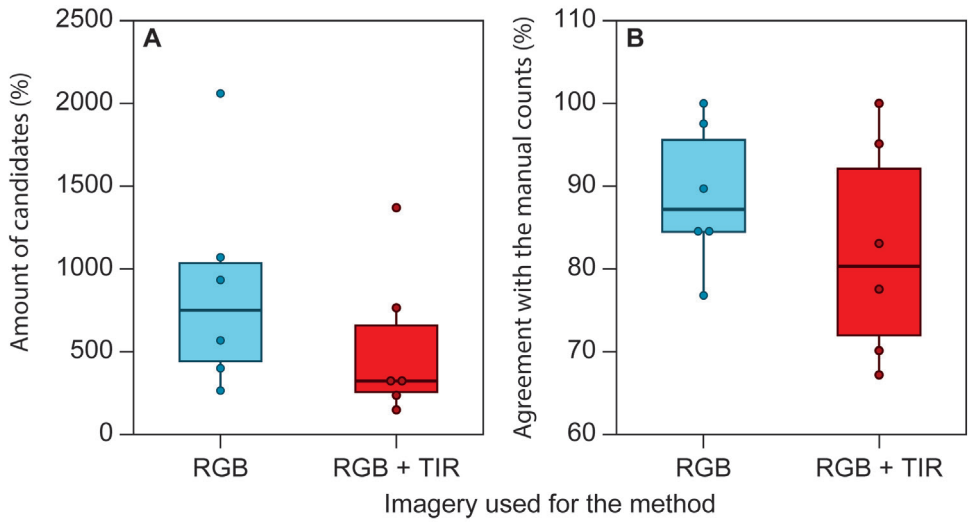

Fig. 4. A) Amount of candidates (e.g., noise) in% of number of individuals of LBBG recorded by the manual counts for the method using only RGB imagery (blue), and using a combination of RGB and TIR imagery (red). B) Semi-automatic counting mean agreement for the RGB imagery (blue), and for the combination of RGB and TIR imagery (red) with the manual counts of LBBG individuals. The data for this figure was obtained from the DJI Matrice 210 flight in 2019 at 50 m altitude.

semi-automatic counts varied with the height from which photos were taken (Table 1, Fig. 5). For the 50 m flight, the mean agreement of the HG counts was 93.0 ± 3.5% (75–104%), and the relationship between the manual counts and the semi-automatic counts was positive and highly significant ($F_{1,4}$ = 142.4, $p < 0.001$, R^2 = 0.96). The mean agreement for the LBBG was markedly lower, with a mean value of 76.9 ± 6.2% (59–98%), but the relationship with the manual counts was still positive and significant ($F_{1,4}$ = 28.5, p = 0.003, R^2 = 0.82). For the 20 m flight with a high quality camera, the mean agreement in detecting HG increased to an average of 97.7 ± 1.1% (92–100%; $F_{1,4}$ = 1,129, $p < 0.001$, R^2 = 0.99), and for the LBBG the mean agreement increased to 94.8 ± 1.8% (88–100%). For the 15 m flight with a low quality camera, the mean agreement in detecting HG remained unchanged (97.7 ± 1.3%, 91–101%; $F_{1,4}$ = 731, $p < 0.001$, R^2 = 0.99), whereas the detection of LBBG dropped significantly to an average of 74.8 ± 4.2% (57–88%), while still presenting a strong positive correlation with the manual counts ($F_{1,4}$ = 126.4, $p < 0.001$, R^2 = 0.95).

4. Discussion

The semi-automatic method based on Rush *et al.* (2018) for counting individuals proved to be efficient for one of our studied species, the Herring Gull (HG). For this species, we obtained similar mean agreement levels (~ 96%) as Rush *et al.* (2018) (Fig. 3). The precision of detection of the other study species, the Lesser Black-backed Gull (LBBG), corresponded to 71% of those present. One of the reasons for this low accuracy is the black back combined with the species' preference for nesting in more densely vegetated areas (Coulson 2019) and the low spectral resolution of the camera at 50 m height. Pixels corresponding to the black backs of the gulls were frequently not differentiated from other dark areas in the picture (e.g., shadows or dark patches) making its detection by the non-specificity of the training sample more difficult (cf. Burke *et al.* 2019).

In order to improve the detection of inconspicuous species in our collected imagery, and as a proof-of-concept, we added a thermal band to our sensors. We predicted a reduction in the selection of false polygons resulting from the addition of an independent extra layer exclusively

Table 1. Percentage of agreement of the semi-automatic method relative to the manual counts at the different heights for Herring Gull (HG) and Lesser Black-backed Gull (LBBG). The Adj R^2 shows the correlation between these two counts. Number of transects = 7. The drone DJI Matrice 210 conducted flights at 50 and 20 m, whereas DJI Phantom 4 Pro flew at 15 m.

	50 m			20 m			15 m		
Species	Agreement	R^2	*p*	Agreement	R^2	*p*	Agreement	R^2	*p*
HG	93.0 ± 3.5%	0.96	< 0.001	97.7 ± 1.1%	0.99	< 0.001	97.7 ± 1.3%	0.99	< 0.001
LBBG	76.9 ± 6.2%	0.82	0.003	94.8 ± 1.8%	0.99	< 0.001	74.8 ± 4.2%	0.95	< 0.001

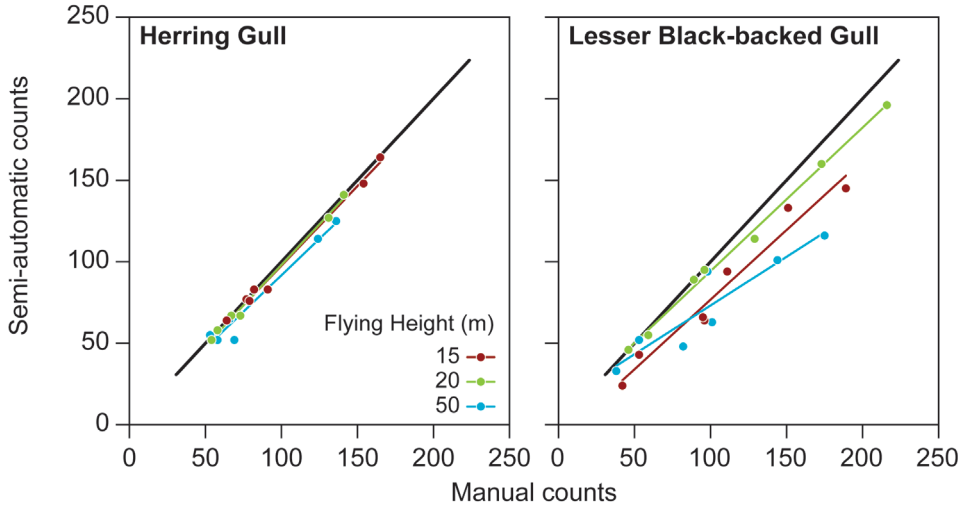

Fig. 5. Linear regression between the manual counts of the test areas from the UAV orthomosaic and counts conducted by the semi-automatic (SA) methods in the 2019 flights for Herring Gull and Lesser Black-backed Gull. The black line represents a 1:1 relation between counts, whereas the coloured lines represent the adjusted R-square of the model for each of the flying heights: 15 m (red), 20 m (green) and 50 m (blue). The colored dots represent each of the test areas per flying height (n=7). The data for this figure was obtained from the DJI Matrice 210 and DJI Phantom 4 Pro flights in 2019 at 15, 20 and 50 m altitude.

containing heat signatures. Although results from 50 m altitude showed that combining RGB with TIR imagery reduced the final mean agreement by ~ 6%, these results should not be interpreted as demonstrating the inefficiency of using multispectral cameras. The drop in accuracy was most likely linked to a reduction in the quality of the automatic output (particularly from the false positives that generated up to 10 times more polygons than there were real gulls in the image, i.e., generating excess noise). In other words, at this height, where the accuracy of the counting method is low due to reduced imagery resolution, a reduction in quality of the automatic output may reduce true positives, reducing overall accuracy. Nevertheless, we obtained a reduction of 40% among the false positives, which in itself represents a major achievement at this level of image resolution, and opens the door for further testing at lower heights and increased thermal sensor quality where most of the false positives errors could be eliminated (Chabot *et al.* 2019; Lee *et al*. 2019). Equally, thermal infrared cameras continue to be expensive, the sensors have low resolution, and data processing creates software and hardware challenges (Kays *et al.* 2019, Scholten *et al.* 2019). In our case, we obtained a final resolution of 1.20 cm/px for the RGB imagery compared to 7.10 cm/px for TIR imagery from the same drone platform. Furthermore, the

RGB/TIR imagery did not spatially match 100% when overlaid, necessitating more sophisticated image processing to improve accurate overlay matching as well as improved spectral resolution to improve thermal utility.

Drone survey is more efficient, cheaper and allows access to areas otherwise impossible (Chabot & Francis 2016, Brisson-Curadeau *et al.* 2017, Han *et al.* 2017, Afán *et al.* 2018). However, such advantages also come with some limitations, such as battery life and the need for high quality imagery. These constraints present the researcher with a trade-off between the extent of the area to be covered and the quality of the images: flying higher allows coverage of greater areas at the same battery power, but costs in terms of the spatial resolution of the images. We found that flying at different heights (50 m, 20 m, and 15 m) and with different cameras had a huge impact on the quality of the pictures and thus the precision of the semi-automatic counting method (Fig. 5).

Flying at 50 m was more energy efficient than flying at 20 m, provided the same camera is used, but was far less accurate in detecting breeding gulls. Even with a faster processing time of the imagery by flying with the sensor at 50 m, the 20 m flight showed an increase of ~ 4% in the agreement with the manual counts for semi-automatic HG detection, and of ~ 20% for LBBG detection. Flying lower demonstrably increased accuracy, especially when detecting less conspicuous birds (such as the LBBG). System requirements will vary depending on focal species and background characteristics. High altitude drone monitoring of conspicuous animals (e.g., Mute Swans *Cygnus olor* (Clausen *et al*. 2020) against a dark water background) is likely to reduce image processing time and increase area coverage per battery cycle without loss of accuracy compared to more cryptic species with complex backdrop environments. In addition, we could observe an effect of the size of the sample of the colony on the semi-automatic output when flying at 50 m for the HG, but not when flying at 15 and 20 m (Fig. 5). A similar trend was observed for the LBBG, especially when the camera quality was lower, and the flying height was higher. This underlines the importance of using adequate quality cameras and specific flying heights regarding the target species, to avoid flaws in the semi-automatic counting methods and achieve their maximum efficiency. Under all circumstances, we recommend thorough literature reviews before conducting flights under given conditions, and not to fly below 15 m to prevent unnecessary disturbance to breeding birds (cf. Grenzdörffer 2013, Vas *et al.* 2015). Birds unattached to nests will usually show far longer escape distances to drones than breeding birds, so drone monitoring at higher altitudes is recommended for studies undertaken outside the breeding season (Holm *et al.* 2018).

Regarding photographic image quality, we were unable to perceive detection probability differences for HG when comparing the DJI Matrice 210 (Zenmuse X5S camera, considered the better quality camera) with a DJI Phantom 4 Pro. Both averaged an agreement of 97%, with a correlation between the number of manual and semi-automatic counts of 0.99 (Table 1). However, the mean agreement when identifying LBBG dropped significantly from ~ 95% when using the high quality camera (Matrice 210) to ~ 75% when using the lower quality camera (Phantom Pro 4). This drop in accuracy confirms the value of using better quality cameras, especially for more cryptic species, because the Phantom 4 Pro generated more noise for the less conspicuous species, despite lower flight height and equivalent spatial resolution (0.5 cm/px). High megapixel ratings on smaller cameras (e.g., the standard camera on the DJI Phantom Pro) often result in greater noise and inferior image quality. We therefore recommend surveyors to study the habitat distribution and breeding behaviour of their focal species, as this may permit the use of cheaper equipment while achieving similar levels of accuracy.

In summary, the semi-automated counting method proposed by Rush *et al.* (2018) applied in this study proved an accurate and fast solution to reduce the manual handling of drone imagery in the case of identifying HG. However, we found major differences in accuracy between survey platforms, especially for the less conspicuous LBBG, which lead us to recommend forward planning and testing of equipment in advance in relation to the specific features of a proposed case study with regard to securing effective data collection. In this paper, we underline the importance of considering the study species and test the UAV equipment for the specific purpose

before conducting a full survey, as many factors can influence the success and efficiency of monitoring using drones, including associated costs of the equipment and impacts on individuals in terms of disturbance, especially when outside the breeding season.

Acknowledgments. We thank our colleagues who participated in the field and with logistics. We thank Geoffrey Groom for providing advice regarding the initial set up of the ArcGIS software. The Danish Nature Agency is thanked for giving permission to conduct the study in the nature reserve. The study was ethically approved because of the extensive tolerance of the breeding gulls towards the flights with the UAV, and other species present on the island were apparently not disturbed by the study. Thanks are also due to Anthony D. Fox for improving the English of the manuscript. We thank the two anonymous reviewers and the editors for constructive comments to an earlier draft.

Conflict of interests. None declared.

Author contributions. ACC and TB conceived the study and designed methodology; TH collected the UAV data; ACC analyzed the UAV data and performed the semi-automated classification; and ACC led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

UAV Regulations. Before the fieldwork began all necessary permissions for the BL 7-15 areas (areas especially sensitive to noise) were granted from the relevant Danish authorities ("The Danish Transport, Construction and Housing Authority" and "The Danish Environmental Protection Agency"). The UAVs were controlled by two certified UAV pilots.

Pesivien lokkien laskenta lennokien avulla: kameran laatu ja lentokorkeus vaikuttavat puoliautomaattisen laskentamenetelmän tarkkuuteen

Kun laajojen merilintukolonioiden laskemiseen käytetään lennokkeja, voidaan välttää monia perinteiseen laskentaan liittyviä ongelmia, mutta kuvien manuaalinen käsittely on kallista ja haastavaa. Puoliautomaattisten kuva-analyysien käyttäminen edellyttää kuvalta hyvää spatiaalista resoluutiota: silloin täytyy valita optimaalinen tasapaino suhteellisen pienen alueen (matala lentokorkeus)/hyvän resoluution ja laajemman alueen (korkea lentokorkeus)/huonomman resoluution välillä. Yksittäisen linnun havaitsemisen todennäköisyyttä lisää se, että kontrasti linnun ja taustan välillä on suuri. Tätä voidaan parantaa esimerkiksi lämpösensoreita käyttämällä. Tutkimuksessa käytettiin puoliautomaattista analyysimenetelmää lennokkien ottamien monispektrikuvien analyysiin harmaa- ja selkälokkikolonian laskemiseen. Tietokone ohjelmoitiin havaitsemaan eri korkeuksista ja eri kameroilla otettuja kuvia niiden spektrien perusteella. Puoliautomaattisen laskennan tulokset olivat lähimpänä manuaalista laskentatulosta kun lentokorkeus oli matalin (20 m) ja kameran resoluution suurin (97.7 ± 1.1 harmaalokilla, 94.8 ± 1.8% selkälokille). Tarkkuus vaihteli myös eri ajojen välillä: paras tulos saatiin matalalla lennettäessä, laadukkaalla kameralla, ja lämpösensorin käyttäminen vähensi taustamelua noin 40%.

References

Afán, I., Máñez, M. & Díaz-Delgado, R. 2018: Drone Monitoring of Breeding Waterbird Populations: The Case of the Glossy Ibis. — Drones 2 (4): 42.

Ardupilot. 2019. Available at https://ardupilot.org/planner/

Barr, J.R., Green, M.C., DeMaso, S.J. & Hardy, T.B. 2018: Detectability and visibility biases associated with using a consumer-grade unmanned aircraft to survey nesting colonial waterbirds. — Journal of Field Ornithology 89: 242–257.

Bregnballe, T. & Lyngs, P. 2014: Udviklingen i ynglebestanden af Sølvmåger i Danmark 1920-2012. — Dansk Ornitologisk Forening 108 (3): 187–198.

Burke, C., Rashman, M., Wich, S., Symons, A., Theron, C. & Longmore, S. 2019: Optimizing observing strategies for monitoring animals using drone mounted thermal infrared cameras. — International Journal of Remote Sensing 40 (2): 439–467.

Brisson-Curadeau, E., Bird, D., Burke, C., Fifield, D.A., Pace, P., Sherley, R.B. & Elliot, K.H. 2017: Seabird species vary in behavioural response to drone census. — Scientific Reports 7: 17884.

Calladine, J. 1997: A comparison of Herring Gull *Larus argentatus* and Lesser Black-backed Gull *Larus fuscus* nest sites: their characteristics and relationships with breeding success. — Bird Study 44: 318–326.

Chabot, D., Craik, S.R. & Bird, D.M. 2015: Population census of a large Common Tern colony with a small unmanned aircraft. — PLoS ONE 10: e0122588.

Chabot, D. & Francis, C.M. 2016: Computer-automated bird detection and counts in high-resolution aerial images: a review. — Journal of Field Ornithology 87

(4): 343–359.
Chabot, D., Stapleton, S. & Francis, C.M. 2019: Measuring the spectral signature of polar bears from a drone to improve their detection from space. — Biological Conservation 237: 125–132.
Clausen, K.K., Holm, T.E., Pedersen, C.L., Jacobsen, E.M. & Bregnballe, T. 2020: Sharing waters: the impact of recreational kayaking on moulting mute swans *Cygnus olor*. — Journal of Ornithology 161: 469–479.
Coulson, J.C. (1st ed.) 2019: Gulls. – HarperCollins Publishers, London.
Descamps, S., Béchet, A., Descombes, X., Arnaud, A. & Zerubia, J. 2011: An automatic counter for aerial images of aggregations of large birds. — Bird Study 58 (3): 302–308.
Dulava, S., Bean, W.T. & Richmond, O.M. 2015: Environmental reviews and case studies: Applications of unmanned aircraft systems (UAS) for waterbird surveys. — Environmental Practice 17: 201–210.
Environmental Systems Research Institute (ESRI). 2019. ArcGIS Desktop Help 10.5 Geostatistical Analyst: https://www.esri.com/
Foody, G.M. & Mathur, A. 2006: The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. — Remote Sensing of Environment 103 (2): 179–189.
Fretwell, P.T., LaRue, M.A., Morin, P., Kooyman, G.L., Wienecke, B., Ratcliffe, N., Fox, A.J., Fleming, A.H., Porter, C. & Trathan, P.N. 2012: An Emperor Penguin Population Estimate: The First Global, Synoptic Survey of a Species from Space. — Plos ONE 7 (4).
Fuller, A.R., McChesney, G.J. & Golightly, R.T. 2018: Aircraft Disturbance to Common Murres (*Uria aalge*) at a Breeding Colony in Central California, USA. — Waterbirds 41(3): 257–267.
Goodenough, A.E., Carpenter, W.S., MacTavish, L., Theron, C., Delbridge, M. & Hart, A.G. 2018: Identification of African Antelope Species: Using Thermographic Videos to Test the Efficacy of Real-Time Thermography. — African Journal of Ecology 56 (4).
Grenzdörffer, G.J. 2013: UAS-based automatic bird count of a common gull colony. — International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 1: 169–174.
Han, Y., Hwa, S. & Kwon, O. 2017: Possibility of applying unmanned aerial vehicle (UAV) and mapping software for the monitoring of waterbirds and their habitats. — Journal of Ecology and Environment 41: 21.
Hodgson, J.C., Mott, R., Baylis, S.M., Pham, T.T., Wotherspooon, S., Kilpatrick, A.D., Segaran, R.R., Reid, I., Terauds, A. & Koh, L.P. 2018: Drones count wildlife more accurately and precisely than humans. — Methods in Ecology and Evolution 9: 1160–1167.
Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T. & Robinson, A. 2018: How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. — Methods in Ecology and Evolution 9 (4): 881 – 892.
Holm, T.E., Kanstrup, N., Riddervold, M., Jensen, L.Ø. & Bregnballe, T. 2018: Brug af droner til overvågning af ynglende vandfugles reaktioner på menneskelig færdsel. — Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 32 s. - Teknisk rapport nr. 129
Holm, T. E., Bregnballe, T., Rasmussen, L. M., Riddervold, M., Castro, A. C., Sterup, J. & Kotzerka, J. 2019. Overvågning af ynglefugle ved brug af droner. Videnskabelig rapport fra DCE - Nationalt Center for Miljø og Energi, Nr. 311 https://dce2.au.dk/pub/SR311.pdf
Inman, V.L., Kingsford, R.T., Chase, M.J. & Leggett, K.E.A. 2019: Drone-based effective counting and ageing of hippopotamus (*Hippopotamus amphibius*) in the Okavango Delta in Botswana. — PLoS ONE 14 (12): e0219652.
Kays, R., Sheppard, J., Mclean, K., Welch, C., Paunescu, C., Wang, V., Kravit, G. & Crofoot, M. 2019: Hot monkey, cold reality: surveying rainforest canopy mammals using drone-mounted thermal infrared sensors. — International Journal of Remote Sensing 40 (2): 407–419.
Lee, W.Y., Park, M. & Hyun, C. 2019: Detection of two Arctic birds in Greenland and an endangered bird in Korea using RGB and thermal cameras with an unmanned aerial vehicle (UAV). — PLoS ONE 14(9): e0222088.
López, J.J. & Mulero-Pázmány, M. 2019: Drones for Conservation in Protected Areas: Present and Future. — Drones 3 (10).
Lyons, M., Brandis, K., Wilshire, J., Murray, N., McCann, J., Kingsford, R. & Callaghan, C. 2019a: A protocol for using drones to assist monitoring of large breeding bird colonies. — EcoEvoRxiv, April 29.
Lyons, M.B., Brandis, K.J., Murray, N.J., Wishire, J.H., McCann, J.A., Kingsford, R.T. & Callaghan, C.T. 2019b: Monitoring large and complex wildlife aggregations with drones. — Methods in Ecology and Evolution 10: 1024–1035.
Magness, D.R., Eskelin, T., Laker, M. & Renner, H.M. 2019: Evaluation of small unmanned aerial systems as a census tool for Aleutian Tern *Onychoprion aleuticus* colonies. — Marine Ornithology 47: 11–16.
Møller, A.P. 1978. Mågnernes Larinae yngleudbreddelse, bestandsstørrelse og -ændringer i Danmark, med supplerende oplysninger om forholdene i det øvrige Europa. — DOF 72.
Pfeifer, C., Barbosa, A., Mustafa, O., Peter, H., Rümmler, M. & Brenning, A. 2019: Using Fixed-Wing UAV for Detecting and Mapping the Distribution and Abundance of Penguins on the South Shetlands Islands, Antarctica. — Drones, 3 (39).

R Core Team. 2018: R: A language and environment for statistical computing. — R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richards, J.A. & Jia, X. (4th eds) 2006: Remote sensing digital image analysis: an introduction. — Springer, Berlin.

Rush, G.P., Clarke, L.E., Stone, M. & Wood, M.J. 2018: Can drones count gulls? Minimal disturbance and semiautomated image processing with an unmanned aerial vehicle for colony-nesting seabirds. — Ecology and Evolution 8 (24):12322–12334.

Sardà-Palomera, F., Bota, G., Viñolo, C., Pallarés, O., Sazatornil, V., Brotons, L., Gomáriz, S. & Sardà, F. 2011: Fine-scale bird monitoring from light unmanned aircraft systems. — Ibis, 154 (1): 177 – 183.

Sardà-Palomera, F., Bota, G., Padilla, N., Brotons, L. & Sardà, F. 2017: Unmanned aircraft systems to unravel spatial and temporal factors affecting dynamics of colony formation and nesting success in birds. — Journal of Avian Biology 48: 1273–1280.

Scholten, C.N., Kamphuis, A.J., Vredevoogd, K.J., Lee-Strydhorst, K.G., Atma, J.L., Shea, C.B., Lamberg, O.N. & Proppe, D.S. 2019: Real-time thermal imagery from an unmanned aerial vehicle can locate ground nests of a grassland songbird at rates similar to traditional methods. — Biological Conservation 233: 241–246.

Valle R.G. & Scarton F. 2019: Drones improve effectiveness and reduce disturbance of censusing Common Redshanks Tringa totanus breeding on salt marshes. — Ardea 107: 275–282.

Vas, E., Lescroel, A., Duriez, O., Boguszewski, G. & Gremillet, D. 2015: Approaching birds with drones: first experiments and ethical guidelines. — Biology Letters 11(2): 20140754.

Weissensteiner, M.H., Poelstra, J.W. & Wolf, J.B.W. 2015: Low-budget ready-to-fly unmanned aerial vehicles: an effective tool for evaluating the nesting status of canopy-breeding bird species. — Journal of Avian Biology 46: 425–430.

Appendix

Appendix I. Summary of sampling protocols denoting UAV type, camera type, flight date, flight time, flight height, image resolution and number of battery changes.

UAV type	Camera type	Flight date & time of day	Flight time (min.)	Flight height (m)	Image resolution	Battery changes
Event 38 model E384	Sony IL CE-QXI	30 May 2018	150	50	1.20 cm/px (RGB)	1
DJI Matrice 210	DJI Zenmuse X5S	14 May 2019, 17:33–18:47	74	20	0.53 cm/px	2
DJI Matrice 210	DJI Zenmuse XT2 TIR	15 May 2019, 04:56–06:42	106	50	1.20 cm/px (RGB), 7.10 cm/px (TIR)	4
DJI Phantom 4 Pro	DJI Phantom 4 Pro standard	15 May 2019, 07:14–09:52	158	15	0.53 cm/px	5