Inter-Research Science Publisher

Endangered Species Research
Volume 56:187–199

Published April 4, 2025
https://doi.org/10.3354/esr01396

OPEN ACCESS

Feasibility of using very high-resolution satellite imagery to monitor Tristan albatrosses *Diomedea dabbenena* on Gough Island

Marie R. G. Attard[1,*], Richard A. Phillips[1], Steffen Oppel[2,3], Ellen Bowler[1], Peter T. Fretwell[1]

[1]British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge, CB3 0ET, UK
[2]RSPB Centre for Conservation Science, Royal Society for the Protection of Birds, The Lodge, Sandy, SG19 2DL, UK
[3]Swiss Ornithological Institute, Seerose 1, 6204 Sempach, Switzerland

ABSTRACT: The Tristan albatross *Diomedea dabbenena* is a Critically Endangered species that breeds exclusively on remote islands in the South Atlantic Ocean. Although the population continues to decline, regular on-the-ground monitoring of Tristan albatross populations is logistically challenging and costly. If this monitoring is reduced in the future, then alternative methods would be necessary to track long-term population trends. Here, we assessed the viability of using 31 cm resolution satellite imagery to count Tristan albatrosses during the breeding season on Gough Island. Counts of birds in a satellite image by 2 wildlife remote-sensing specialists were compared with GPS coordinates of active nests recorded in the field. Birds were detected at 103 (67.8%) of the 152 active nests in the cloud-free regions of the satellite image. Acquiring suitable imagery is challenging because upland nesting sites are prone to low-lying orographic cloud, with only 1 cloud-free image obtained across 8 seasons of archived and 1 yr of tasked imagery. Our research demonstrates that due to incomplete detection, and the limited availability of suitable imagery resulting from persistent cloud cover over the island, Tristan albatrosses cannot be reliably counted or monitored with available satellite imagery. Differences in detection probability were not explained by nest attributes or bird plumage colouration. More commercial satellites in orbit may improve chances of obtaining cloud-free imagery across the island in the future, but until then, on-the-ground monitoring is required if we are to obtain accurate population counts and for the UK to meet its commitments to monitor this species.

KEY WORDS: Albatross · *Diomedea dabbenena* · Satellite remote sensing · Population monitoring · Gough Island · Procellariiformes · Endangered species · Conservation

1. INTRODUCTION

Many seabirds of conservation concern nest on remote islands where monitoring is hindered by limited accessibility, lack of infrastructure, or unfavourable weather conditions. However, long-term monitoring data are critical for informing and evaluating conservation actions (McClelland et al. 2016, Brooke et al. 2018, Bird et al. 2021, Edney & Wood 2021). Very high-resolution (<1 m per pixel ground sample distance) satellite imagery can offer a cost-effective solution to monitor populations when ground counts are unavailable or infrequent (Edney & Wood 2021). This technology has been used to estimate the abundance of large seabirds such as penguins, albatrosses and boobies, either indirectly (occupied area × assumed densities) or directly (counts of individuals) (Barber-Meyer et al. 2007, Hughes et al. 2011, Fretwell et al. 2017, Weimerskirch et al. 2018, Dolliver 2019, Foley et al. 2020). As even higher resolution satellite imagery becomes available, it should be possible to detect remote colonies and assess population sizes of smaller bird species (Lynch et al. 2012).

*Corresponding author: maratt@bas.ac.uk

Publisher: Inter-Research · www.int-res.com

The Tristan albatross *Diomedea dabbenena* is the third rarest albatross species, categorised as Critically Endangered by the IUCN (BirdLife International 2018) and listed on Annex 1 by the multi-lateral Agreement on the Conservation of Albatrosses and Petrels (ACAP), which coordinates international activity to mitigate known threats to their populations (Phillips et al. 2016). It breeds only on Gough Island (approximately 1500 pairs) and Inaccessible Island (2–3 pairs) in the Tristan da Cunha group (Cuthbert et al. 2004). Modelling shows that the population has decreased over the past 2 decades, although the number of breeding pairs obtained from ground surveys each year is broadly stable (Oppel et al. 2022). This long-term decline was initially due to low adult survival caused by high incidental mortality (bycatch) in longline fisheries (Cuthbert et al. 2004, Jiménez et al. 2014). However, recent models suggest that low fledging success due to predation of chicks by invasive house mice *Mus musculus* L. on Gough Island is currently the key limiting factor (Wanless et al. 2009, Davies et al. 2015, Oppel et al. 2022). A baiting campaign to eradicate the mice in 2021 was unsuccessful (Samaniego et al. 2022).

Given the slow recovery potential of long-lived albatrosses, even if a second attempt at mouse eradication is successful, multiple censuses would be required in the following few decades to determine if the population is recovering, particularly in light of other threats such as bycatch in fisheries (Oppel et al. 2022). The Tristan albatross breeds biennially if successful in raising a chick, but some pairs that fail early are likely to breed again the following season (Tickell 1968, Croxall et al. 1990). Monitoring over successive years is therefore necessary to adequately determine population status. Ground surveys of Tristan albatrosses at Gough Island have been carried out annually since 2000/2001, involving a team of 2 researchers on station for an entire year, contributing to a total cost of ~100 000 GBP for a program that also includes monitoring of other species on the island (Caravaggi et al. 2019). If the ground surveys were discontinued due to financial or logistical challenges, alternative methods would be necessary to maintain monitoring. Satellite remote sensing offers a potential solution, providing the ability to cover large areas instantaneously without disturbance, potentially at a lower cost than ground surveys. However, applying satellite technology would still require financial investment in image acquisition and analysis, as well as periodic ground-truthing to ensure that satellite assessments remain accurate.

Among the commercial providers of high-resolution imagery, Maxar Technologies was the first to launch 31 cm resolution sensors, with WorldView-3 (active 2014–present) and WorldView-4 (active 2016–2019). These offer the longest archival record of 31 cm optical imagery. For an animal to be observed directly in 31 cm resolution imagery, it must meet 3 primary criteria: (1) be large enough to cover several pixels, (2) contrast with the background colour or other characteristics of the local habitat, and (3) occupy open areas (LaRue et al. 2017). The wandering albatross *Diomedea exulans*, northern royal albatross *D. sanfordi*, and short-tailed albatross *Phoebastria albatrus* meet these criteria and have been counted in WorldView-3 imagery (Fretwell et al. 2017, Weimerskirch et al. 2018, Brothers et al. 2022). Although these species were discernible in the imagery, the accuracy of the estimates vary, e.g. for wandering albatrosses, the mean count in satellite imagery of 894.5 birds was 18.6% higher than the original number of nests with eggs at Bird Island (Fretwell et al. 2017), and counts in satellite imagery were lower by 17 and 49% than those from aerial photographs at Kerguelen (Weimerskirch et al. 2018). This variability in counts may be attributed, in part, to fluctuations in attendance throughout the day as non-breeding or failed birds arrive and depart. Detectability can also be affected by factors such as small body size or dark plumage (Dolliver 2019), or differences in terrain between colonies, which influences the contrast between the bird and their immediate surroundings (Bowler et al. 2020). Thus, satellite imagery is promising as a monitoring tool but requires validation for new species and sites.

Tristan albatrosses are slightly smaller in size than wandering albatrosses (Cuthbert et al. 2003) and have a darker back, necessitating a rigorous assessment to evaluate whether satellite-based counts could replace ground surveys. Here, we present the first assessment of whether Tristan albatrosses can be observed and accurately counted in 31 cm resolution satellite imagery. Our aims were to test (1) whether the locations of manually labelled Tristan albatrosses in an orthorectified 31 cm resolution satellite image from the incubation period corresponded with GPS coordinates of active nests from a concurrent ground survey, and (2) whether nest attributes (slope and aspect) or bird plumage (reflecting sex and age) of the incubating bird influenced detectability in satellite imagery. We hypothesised that males and older birds would be more readily seen in satellite imagery than females and younger adults because their lighter plumage would show higher contrast against the vegetation. The results are discussed in terms of the potential of using very-high-resolution (VHR) satellite imagery as a tool for long-term monitoring of this population.

2. MATERIALS AND METHODS

2.1. Satellite imagery acquisition and processing

Satellite imagery of Gough Island (40° 21′ S, 9° 53′ W) was tasked for the 31 cm resolution WorldView-3 satellite system for the period of 17 January to 31 March 2022. However, as all imagery of the island captured within this period was obscured by cloud or thick haze, we searched the archives (Maxar Archive Search and Discovery; https://discover.maxar.com/) for cloud-free images from 2014 to 2022 between 25 December and 31 March (corresponding to the incubation or brood-guard stages, when a parent is always present on the nest). No images of the island in the archive were completely cloud free, but one image had a portion (2.62 km^2) that was cloud free and which corresponded with an area where there was a good sample of nests that were being monitored regularly at the time of the image acquisition. This WorldView-4 image (33 cm ground sampling distance resampled to 30 cm resolution) was taken at 10:46 h on 1 February 2018 (see Table S1 in Supplement 1 for details; www.int-res.com/articles/suppl/n056p187_supp1.pdf, for all supplementary tables and figures).

The satellite image consisted of multi-spectral and panchromatic files that were processed in ArcGIS Pro (version 3.3.0; ESRI 2024). Surface topography and the tilt of the satellite can distort features in complex landscapes, such as mountain slopes. As such, the image was orthorectified using a Digital Terrain Model (DTM) from Airbus (product WorldDEM™) to remove distortion and assign more accurate coordinates. The DTM represented bare earth terrain without obstruction features above ground and had an absolute vertical accuracy of <10 m and absolute horizontal accuracy of <6 m. The Gram-Schmidt pansharpening method was then applied to produce a better visualisation of the multiband image using the high-resolution (30 cm) panchromatic image (ESRI 2024).

2.2. On-the-ground monitoring

The number of breeding pairs (nests with eggs) on Gough Island has been counted annually since 2001 from vantage points in late January or early February in 8 areas delineated by ridgelines and valleys (Ryan et al. 2001). In addition, GPS locations of all nests in 3 intensively monitored colonies (Gonydale, Hummocks and Tafelkop) are recorded each year using a handheld GPS (Garmin GPSmap 60CS, accurate to <10 m), and the nests are monitored on a weekly basis to determine breeding success.

Tristan albatrosses are loosely colonial, with nests spaced more than 3 m apart (Ryan et al. 2001). Parents take turns incubating their egg, with each incubation shift normally lasting 1–3 wk. Adult males typically have mostly dark upper wings with a white patch on the elbow, and others have more extensive white on the upper wings, whereas most adult females retain brown feathers on the crown, back, breast and flanks (Fig. S1 in Supplement 1) (Ryan 2023). The oldest males have pale-white backs (Ryan 2000, del Hoyo et al. 2023). The plumage of females is browner than males of the same age. Annual ringing of breeding adults, and of chicks in Gonydale has taken place since 1976 and 1980, respectively (Ryan et al. 2001), and hence, a large proportion of birds in the study areas are individually marked and of known age. The identity of the partner attending the marked nests detectable in the satellite image was recorded shortly before (26 January) and after (5 February) the satellite image was taken in 2018. As this interval is shorter than the minimum duration of an incubation shift, the same individuals present on both occasions were assumed to have also been present on the day the satellite image was taken. Nests that failed between the 2 visits were excluded from further analyses.

2.3. Expert annotation of Tristan albatrosses in satellite imagery

In the cloud-free area within the nest boundary of the processed satellite image, 2 independent expert reviewers, (M. Attard and an independent expert; hereafter referred to as Expert 1 and Expert 2) — both experienced in satellite-image analyses of wildlife — conducted an initial identification of presumed albatrosses using ArcGIS Pro. Their methodology was broadly based on methods in Stapleton et al. (2014). GPS coordinates of active nests from the ground survey were overlaid on the satellite image as a guide, with a 10 m buffer around each GPS coordinate serving as a general search area. The reviewers also examined areas outside of this buffer to identify potential non-breeding albatrosses, which included pre-breeders (birds that had never bred), deferring breeders and failed breeders (Fretwell et al. 2017).

Each buffered area was reviewed initially at a fixed scale of 1:200 to 1:400 to identify features that could be albatrosses. If the feature was not immediately clear, the reviewers examined the area at multiple scales (up to approximately 1:800) to confirm the identification. This mixed-scale approach ensured that the feature was distinct and comparable in size to other features clearly identifiable as birds. After completing independent reviews, the experts convened to jointly review their annotations and resolve any uncertainties by examining each annotation together at multiple scales. Only features that both experts confidently classified as a presumed albatross were retained, resulting in a single shapefile representing a 'gold standard' distribution of detectable birds. This final annotation set was then used to calculate the proportion of active nests observed in the satellite image. If multiple satellite annotations fell within a 10 m radius of a nest recorded in the ground survey, only the closest annotation was considered a match, and the other annotations were not considered to be a bird on the nest. Any known nest not identified by the gold standard is assumed to be not visible in the satellite image, or insufficiently distinct to be confidently identified as an albatross.

Known nests were determined from GPS data during the ground survey, but these may not have corresponded to a detectable bird in the satellite image as reviewers may have been uncertain in identification or unable to discern the presence of a nesting albatross in more pixelated or blurred areas. As such, the 2 experts also annotated a pansharpened, non-orthorectified image of Gough Island to assess whether pixel smearing resulting from the orthorectification process influenced bird detection.

2.4. Factors that may influence detectability of nesting albatrosses

All statistical analyses were performed using python (v3.9.11) within a Jupyter notebook environment (Kluyver et al. 2016). We tested whether the detectability of nesting albatrosses was influenced by characteristics of the incubating parent (sex and minimum age from ringing), and nest position (slope and aspect, both extracted from the DTM). Aspect was measured from 0° to 360°, where 0° (or equivalently 360°) is north. These factors were included in generalised linear models (GLMs) with a logit link function for the binomial dependent variable indicating whether an albatross was observed in the satellite image. The GLM estimates reflect how the likelihood (or odds) of detection change with each predictor variable. Given the binary nature of the response variable (observed vs. unobserved), we used odds ratios — defined as the ratio of the odds of detection for one group compared to another — to quantify the strength and direction of the relationships between predictors (bird and nest characteristics) and detectability.

To address the positive skew in values for slope, we applied a cube root transformation to better meet the assumptions of the GLM. This transformation compresses the scale of the slope values, resulting in lower estimates compared to using the original values. Consequently, the coefficients from the GLM reflect the change in the log-odds of detecting a nesting albatross for a unit increase in the transformed slope. Therefore, a unit increase in the cube root of the slope corresponds to a larger change in the original slope value. This only affects estimates in models that include slope.

Slope and aspect were available for 152 nests that were active when the satellite image was taken, including 52 nests for which the sex and age of the incubating bird was known. We first fitted a full GLM for these 52 nests, incorporating all potential predictors: slope, aspect, sex, and minimum age. To identify the most parsimonious model, we used the 'pdredge' function from the pyDredge package in python (version 3.9.11), which evaluates all possible subsets of the predictor variable. We used Akaike's information criterion adjusted for low sample size (AICc) to identify the top models, considering models within 2 AICc units of the top model as influential. For each top model, we report the coefficient estimates, p-values and 95% confidence intervals to assess the relative importance of predictor variables (Muff et al. 2022).

To assess potential multicollinearity between the 4 predictors, we calculated the variance inflation factor (VIF) for all variables included in the GLMs (Table S2 in Supplement 1). This step helped ensure that no predictors were highly correlated, which could lead to unreliable estimates of the model coefficients. A separate GLM was conducted to test the influence of slope and aspect on albatross detectability, using the larger sample of 152 nests. Differences in detectability based on minimum age were analysed separately for males (n = 26) and females (n = 26) using a GLM; note that this 1:1 ratio was coincidental and not a result of a deliberate sampling design. A chi-squared test was used to determine whether the proportion of nesting males and females detectable in the satellite image were equal.

2.5. Assessing inter-observer reliability of satellite-based Tristan albatross counts

To further evaluate the reliability of satellite imagery for detection and counting of Tristan albatrosses, 9 independent volunteers were recruited to complete blind annotations by marking presumed albatrosses in the imagery. These observers were categorised into 3 groups based on experience: (1) those with extensive experience in using satellite imagery or unoccupied aircraft systems (UASs) for wildlife detection and counting (Observers 1–4), (2) those with albatross fieldwork experience (Observers 4–6), and (3) those with no prior experience in annotating satellite imagery or albatross fieldwork (Observers 7–9). To avoid bias, the expert reviewers who created the reference annotations were not included in the volunteer pool.

The orthorectified satellite image was divided into 100 × 100 m tiles, resulting in 303 tiles covering a total area of 3.03 km^2. A random number generator was used to select 24 tiles where at least 1 nest was present, and 6 tiles where no nests were present based on GPS coordinates. These tiles were uploaded to an online open-source annotation software, VGG image annotator (VIA) (Dutta & Zisserman 2019), and each volunteer was provided with instructions on how to locate and label albatrosses (see Supplement 2; www.int-res.com/articles/suppl/n056p187_supp2.pdf).

The misclassification rate of each observer was assessed in relation to one another, the gold standard and coordinates of known nests. The gold standard encompassed all birds that were detected with confidence by the 2 experts, whereas the ground survey only identified birds on nests. Consequently, the definition of true positive, false positive, and false negative varied. In general, we defined albatrosses which were correctly predicted to be present as true positive (TP), albatrosses incorrectly predicted to be present as false positive (FP), and albatrosses incorrectly predicted to be absent as false negative (FN). Annotations classified as true positives were those located within 3 m of the gold standard or another observer, as Tristan albatross nests are usually >3 m apart (Ryan et al. 2001), or within 10 m of the nest GPS coordinates, based on the accuracy of the handheld GPS device. A cross-tabulation table was created for each metric. To assess consistency, metrics were averaged for each observer relative to all other observers. The response of each observer was compared separately to the gold standard and known nests to assess their overall performance. F1-scores were used to measure the accuracy of each observer's performance based on a trade-off between recall and precision, as applied in other studies (Bowler et al. 2020). These metrics were calculated as follows:

$$\text{recall} = \frac{\text{TPs}}{\text{TPs + FNs}},\ \text{precision} = \frac{\text{TPs}}{\text{TPs + FPs}},$$
$$\text{F1-score} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall + precision}} \qquad (1)$$

Recall identifies how well the observer identifies true positives (i.e. out of all albatrosses in the 'actual' data, what proportion is predicted to be present by the 'predictor') while precision measures the quality of positive predictions by the observer (i.e. out of all the albatrosses predicted by the observer, what percentage is truly positive according to the 'actual' data) (Daskalaki et al. 2006). A high recall value reflects a better counting performance, and a high precision value indicates higher quality. F1-scores range from 0–1, with a higher value indicating a better overall performance.

2.6. Spatial clustering for multiple observers

Observer annotations were aggregated using a hierarchical spatial clustering approach to group labels based on geographic proximity, implemented in a python script to generate consensus-labelled data. This approach is broadly based on the methodology outlined by Jones et al. (2018). Pairwise Euclidean distances between observer labels were computed, and hierarchical clustering was performed using the Ward linkage method with a 3 m threshold. To ensure unique observer representation within each cluster, only the closest label from each observer was retained. Remaining unmatched labels were then assessed to determine whether they were within 3 m of any other unmatched labels. If such labels were found, a new cluster was formed. The median geographical position for each cluster was calculated, and the bird count was determined for each agreement threshold. For example, an agreement threshold of 3 meant that only clusters with at least 3 observers in agreement were counted. A linear regression was performed to assess whether the proportion of nests correctly identified by the observer-clustered data (using an agreement threshold of 2) varied with nest density.

3. RESULTS

3.1. Nest detection in orthorectified vs. un-orthorectified satellite images

In the orthorectified satellite image, 103 out of 152 nesting birds (67.8%) recorded on the ground were detected successfully by the gold standard (Table 1, Fig. 1). In comparison, only 91 of the 152 nesting birds (59.9%) were detected in the un-orthorectified image using the same consensus-based approach (Table 1).

Each Tristan albatross typically appears as 2–4 pixels in shades ranging from white to greenish-grey in

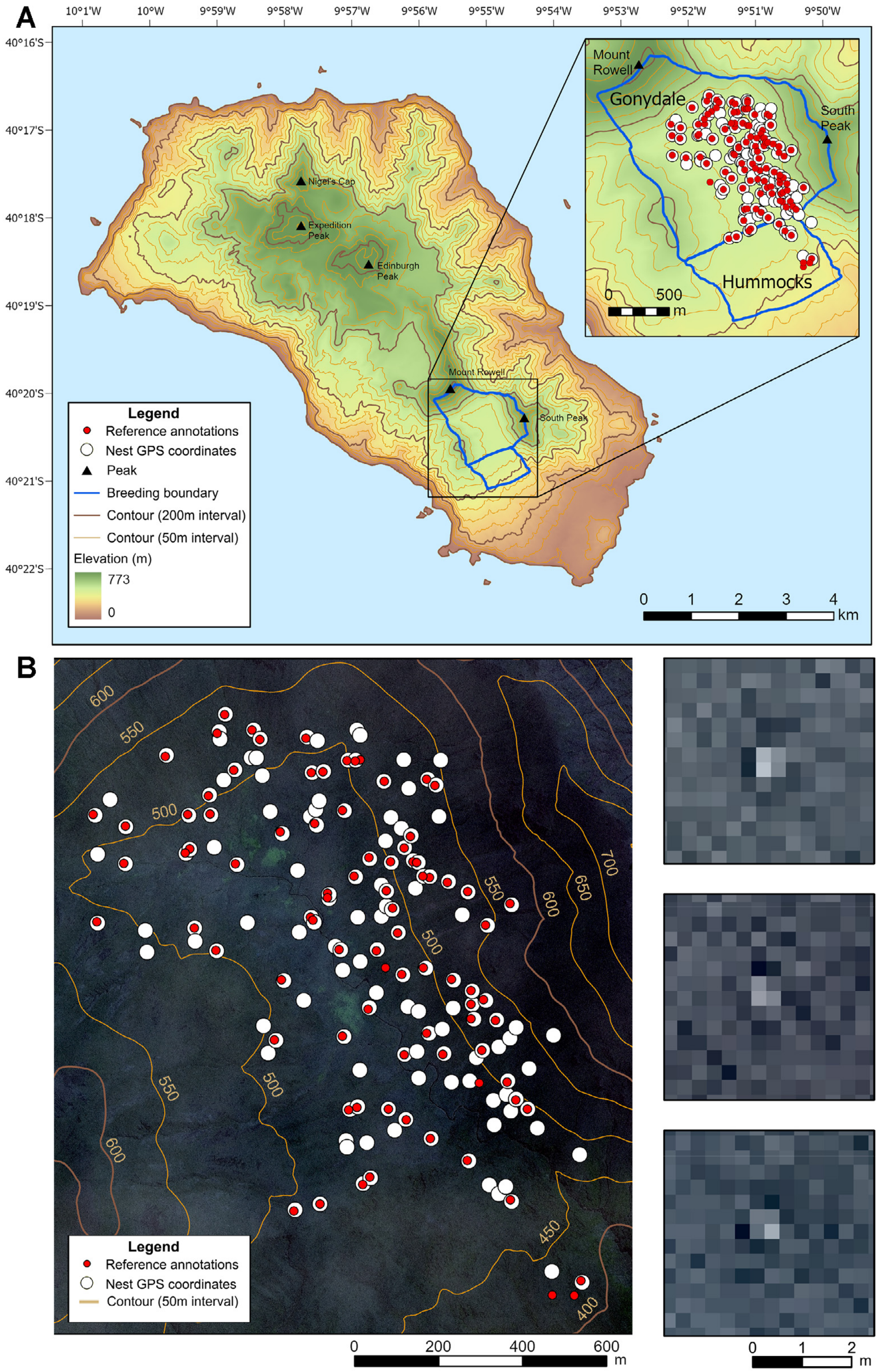

Fig. 1. (A) Map of Gough Island and relevant features, with an inset map of the 2 study areas (Gonydale and Hummocks) that were cloud-free in the satellite image. (B) Satellite image (33 cm resolution downsampled to 30 cm resolution) showing GPS coordinates of Tristan albatross nests from ground surveys in 2018, and nests and presumed non-breeders observed in imagery. Three random examples from the satellite image shows an individual nesting Tristan albatross as light pixels in the centre of the panel. Base map and breeding boundaries from the Royal Society for the Protection of Birds (RSPB). Credit for satellite image © 2024 Maxar Technologies

Table 1. Summary of detection metrics for guided annotations by 2 experts (Expert 1 and Expert 2) and resulting 'gold standard' (the consensus of both experts) relative to known nests. The metrics provide insights into the performance of expert observers in detecting nesting albatrosses based on satellite imagery. The metrics are accuracy (precision), sensitivity (recall), F1-score, total known nests, identified albatrosses matched with nests, presumed albatrosses without nest matches, and total reference annotations for 2 reference observers and the consensus gold standard. Presumed albatrosses that do not match a nest are either non-breeders or misidentifications

Image correction	Annotator	Accuracy (precision)	Sensitivity (recall)	F1 score	Total known nests	Identified albatrosses (nests matched)	Presumed albatrosses (no nest match)	Total reference annotations
Orthorectified	Expert 1	0.92	0.54	0.68	152	82	7	89
	Expert 2	0.89	0.70	0.79	152	107	13	120
	Gold standard	0.89	0.68	0.77	152	103	13	116
Un-orthorectified	Expert 1	0.82	0.55	0.66	152	84	19	103
	Expert 2	0.82	0.67	0.74	152	102	22	124
	Gold standard	0.81	0.60	0.69	152	91	22	113

the 31 cm resolution satellite imagery. Notably, a dark ring was observed around some nesting birds which may be the result of trampling or vegetation removal. Differences were observed between the annotations of the 2 experts. Expert 1 was more conservative, recording 89 birds (including 82 on nests), whereas Expert 2 identified 120 birds (including 107 on nests). After discussion, several annotations from Expert 2 were excluded from the final count as they could not be identified confidently without reference to the GPS coordinates of known nests. Testing different raster brightness settings revealed that the dark rings around nests were more pronounced in the darker images, aiding nest identification. Differences in detection rates across different parts of the image may be partly due to variable pixel stretching during orthorectification. These distortions, which depend on the local slope, aspect, and sensor tilt, can alter the contrast between birds or nests and their surroundings.

3.2. Detection of presumed non-breeding albatrosses

Presumed non-breeding birds, as well as nesting birds, were detected in both image types. Totals of 13 and 22 presumed non-breeding birds were observed more than 10 m from any active nest location in the orthorectified and un-orthorectified images, respectively.

3.3. Factors influencing nest detectability in satellite imagery

The minimum age of incubating males or females did not significantly influence nest detectability in satellite imagery using the gold standard. Based on the logistic regression model, there was no statistically significant association between age and nest detectability in males (odds ratio: 1.05 ± 0.07; 95% CI: 0.93, 1.19; $p = 0.44$; Table S3) and females (odds ratio: 1.00 ± 0.09; 95% CI: 0.84, 1.20; $p = 0.99$; Table S4).

We found no significant difference in detectability between nesting males and females; among individuals identified as the gold standard, 84.2% (n = 22) of nesting females and 69.2% (n = 18) of nesting males were detected in the satellite imagery (Fig. S2; $\chi^2 = 0.98$, df = 1, $p = 0.32$).

None of the GLM models (ΔAICc < 2) explained >5% of the variance in detectability (Table S5). All 4 predictors (aspect, slope, minimum age, and sex) were included in at least 2 of the 7 most parsimonious GLM models for the 52 nests where bird age and sex were also known. However, the effect of each predictor was insignificant, as the 95% confidence intervals included zero.

3.4. Inter-observer variability

Observer reliability was variable; 6 of the 9 observers showed a high number of false positives, and therefore lower precision scores (<0.75) relative to the gold standard (Fig. 2A, Table S6). Observers 1 and 6 had the highest precision (0.94 and 0.85, respectfully) but the lowest recall (0.46 and 0.49), indicating fewer detections but greater reliability (Tables S6 & S7). Conversely, Observer 9 had the lowest precision (0.42) and highest recall (0.69) compared to the gold standard, suggesting that this observer labelled many albatrosses, many of which were likely incorrect. Consequently, Observer 9 had the lowest F1-score compared to the gold standard

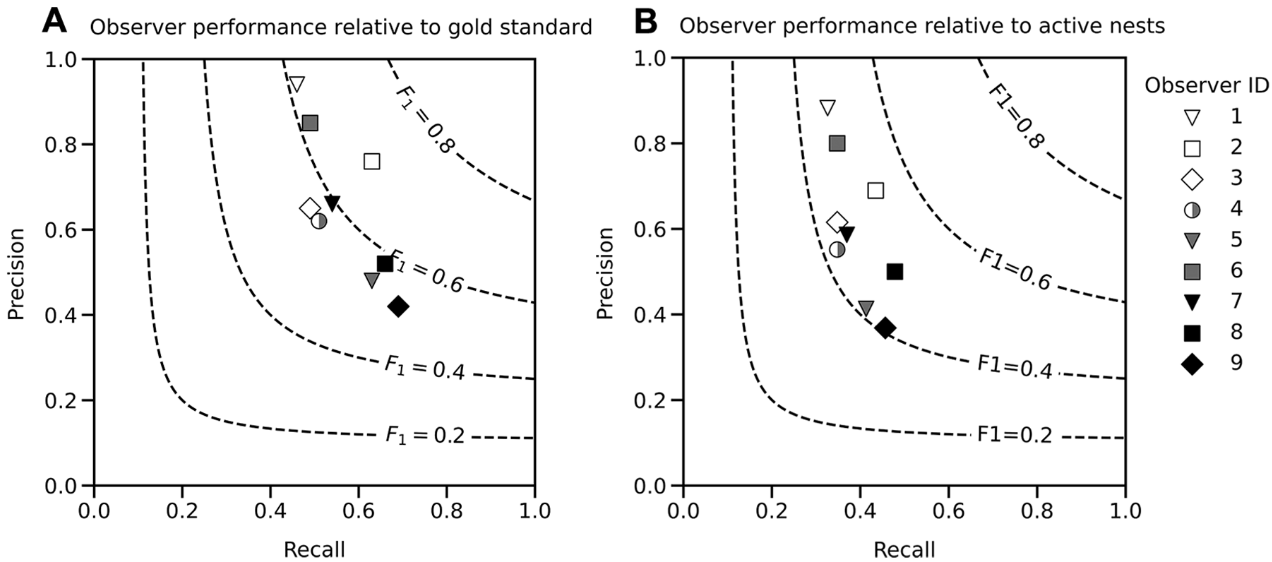

Fig. 2. Scatterplot of average precision–recall values for presumed albatrosses annotated by 9 observers (predicted) across 30 tiles, relative to (A) the gold standard from the satellite imagery (actual) and (B) the active nests (actual) based on GPS coordinates. The volunteers, numbered 1–9, included those with experience in satellite imagery and/or unoccupied aircraft systems (UAS) to detect/count wildlife (Observers 1–4; white symbols), those with albatross fieldwork experience (Observers 4–6; grey symbols) and those with no prior satellite imagery or albatross fieldwork experience (Observers 7–9; black symbols). Black dashed lines show the position of possible recall–precision values corresponding to F1-scores of 0.2, 0.4, 0.6 and 0.8. Annotations classified as true positives were those located within 3 m of the gold standard, as Tristan albatross nests are usually >3 m apart (Ryan et al. 2001), or within 10 m of the nest GPS coordinates, given the accuracy of the handheld GPS device used during annual monitoring (Garmin GPSmap 60CS, accurate to <10 m)

(0.52) (Tables S8 & S9). No albatross was identified by any observer in 13 of the 30 tiles.

As expected, observers with experience of annotating satellite or UAV imagery (Group 1; F1-score range: 0.56–0.69) and those with experience of albatross fieldwork (Group 2; F1-score range 0.54–0.62) generally outperformed those with no prior experience (Group 3: range 0.52–0.59) (Fig. 2A and Table S9). Notably, 2 volunteers from Group 3 achieved the highest recall relative to the gold standard, whereas 3 observers from Group 1 and 2 had the highest precision.

The mean F1-scores in Table S8 illustrate how consistently the observers classify Tristan albatross nests relative to each other. The mean F1-score per observer ranged from 0.56 to 0.68, indicating a moderate level of agreement among observers. The 9 observers each correctly identified between 15 to 22 of the 46 known nests (Table S10) presented to them.

3.5. Bird counts from spatially clustered data

In the clustered dataset, the number of presumed albatrosses increased exponentially as the minimum number of observers labelling the same bird decreased (Fig. 3A). Setting too low an agreement threshold led to overestimation, as labels were included that had a low probability of being an albatross, reducing precision. On the other hand, a high threshold risked the exclusion of valid detections, lowering accuracy. The optimal threshold, therefore, balances accuracy (including more observers) and precision of counts. The number of nests where 2 or more observers agreed increased with the number of known nests, but there was large variation around this relationship ($R^2 = 0.43$, $p < 0.001$; Fig. 3B), suggesting that other factors may affect detection probability. All labels agreed on by all 9 observers appeared as distinct white dots and matched the gold standard (Fig. 3C), whereas labels with no agreement (including with the expert reviewers) were false positives. Labels agreed upon by 4 observers showed a mix of true positives and false positives relative to the gold standard.

4. DISCUSSION

In this study, we demonstrate that satellite remote sensing is unlikely to be a reliable method for long-term monitoring of population size or trends in the Tristan albatross for 2 reasons: (1) obtaining cloud-free images during breeding stages (incubation and brood-guard), when an adult is present at each nest, is challenging, and (2) even in cloud-free images, only

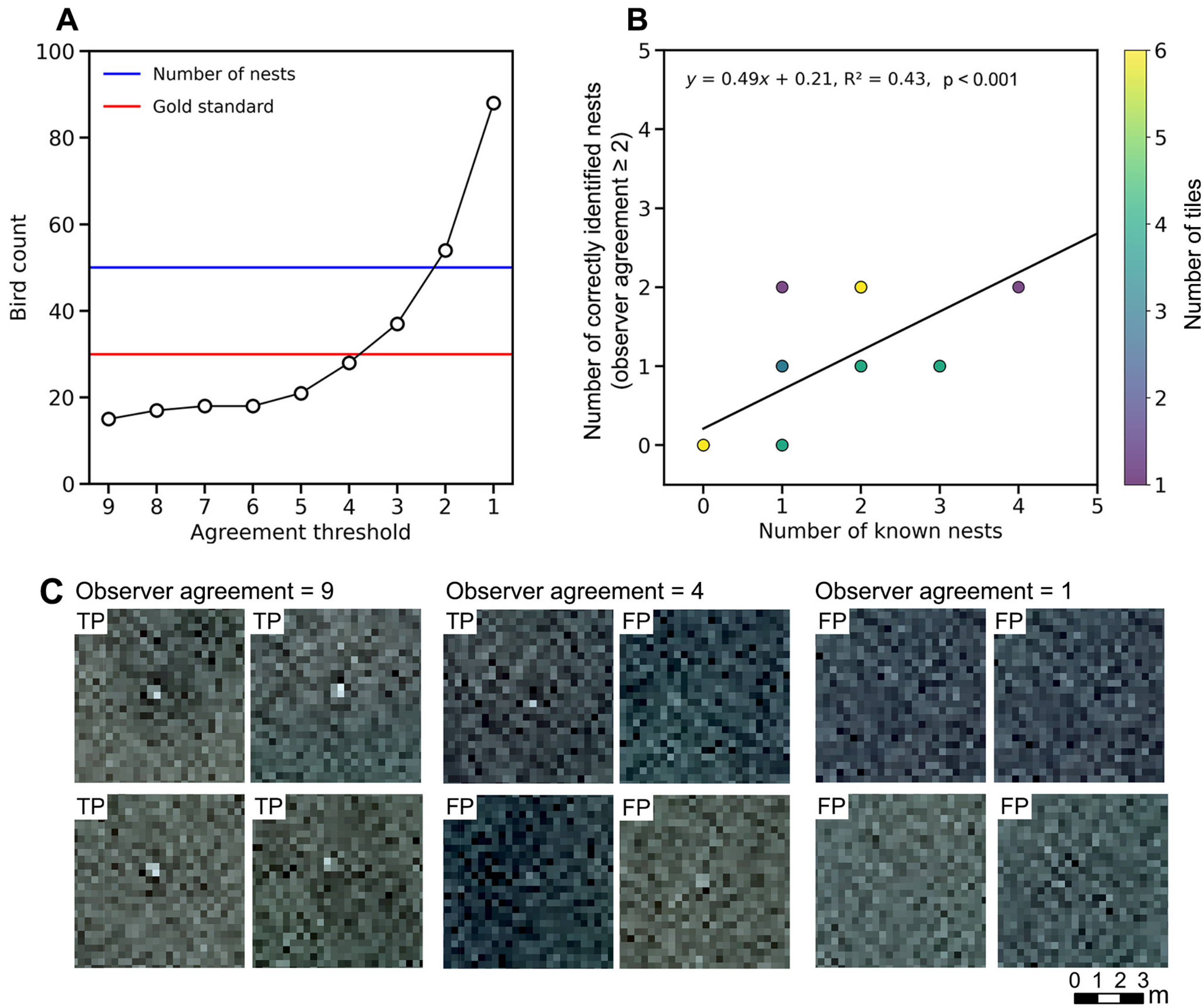

Fig. 3. Tristan albatross detection and counts in 31 cm satellite imagery by 9 observers. (A) Line graph showing number of presumed Tristan albatrosses labelled by 9 observers based on different agreement thresholds for a subset of 30 tiles. Raw labels from different observers were clustered if they were within 3 m of each other, with each cluster containing no more than 1 label per observer. Median geographical position for each cluster was calculated, and bird count was determined for each agreement threshold. For example, an agreement threshold of 3 includes all clusters where 3 or more observers agree, while an agreement threshold of 9 only includes clusters where all 9 observers agree. (B) Linear regression between the number of known nests and the number of correctly identified nests. Each point represents a tile, with the colour indicating the number of tiles with the same combination of correctly identified nests and total nests. Number of correctly identified nests is based on the clustered dataset, where the agreement threshold is set to 2 observers. (C) Satellite image examples of albatrosses labelled by all 9, exactly 4, and only 1 observer. Centre of each image is the location of the label based on 3 m clustered data for the corresponding agreement score. Labels by all 9 observers were true positives (TPs), labels agreed by 4 observers included TPs and false positives (FPs) and all labels labelled by only one of the 9 observers were FPs relative to the gold standard. Credit for satellite images © 2024 Maxar Technologies

67.8% of known nests identified from ground surveys were detected with confidence. This was considerably lower than the detection rates of other species of albatrosses in previous studies (Fretwell et al. 2017, Weimerskirch et al. 2018, Dolliver, 2019).

Although counts of birds in satellite imagery may not provide an accurate estimate of the number of albatross nests, they could in theory be useful as an index of population size if the probability of detection remains relatively constant. However, the correlation between the number of nests that were identified correctly and the total number of nests in our observer-clustered dataset was weak, and the probability of detection varied considerably across space. As such, it seems that detectability is too variable for the satellite counts to provide a robust index of population size for our study species. Both the irregular availability of cloud-free images, and the variable detection probability render currently available satellite imagery an unlikely method to derive a reliable index of population changes in Tristan albatrosses.

Several factors can contribute to variability in detection, including body size, background complexity, contrast between species and surrounding habitat, and shadowing from upslope terrain (Duporge et al. 2021, Attard et al. 2024). In Tristan albatrosses, sexual size dimorphism is minimal, with males only larger in wing, bill, and tarsus lengths (Cuthbert et al. 2003). These minor size differences are unlikely to affect detection probability. Tristan albatrosses have darker plumage than wandering albatrosses — particularly the females — and were detected at much lower rates in our study compared to wandering albatrosses in previous studies (Fretwell et al. 2017, Weimerskirch et al. 2018). However, a high proportion of both male and female Tristan albatrosses were missed in the satellite images from Gough Island, suggesting that plumage colour per se was not the main problem.

The open, tree-less heath and grassland surrounding nests on Gough Island generally provide a relatively uniform, high-contrast dark green background, but variations in grass length and coloration may obscure incubating birds, especially on flat ground where the grass is longer, or sunny, north-facing slopes with thicker vegetation. Many Tristan albatrosses on Gough Island prefer to nest on north or west-facing slopes and will avoid wet peat bogs (Ryan et al. 2001). While no nests would be obscured from above by vegetation, some may have been hidden behind small terrain features, especially if the satellite image was captured at an oblique angle. Notably, we found no effect of sex or minimum age of the incubating bird on detectability. Instead, we suspect that the surrounding vegetation or other aspects of the terrain that obscure incubating birds are more influential in determining detection probability. The effect may depend on the angle from which a satellite image is taken, but whatever the reason, it is likely to make long-term remote monitoring extremely challenging.

The detection of individual nesting birds in satellite imagery can be improved through orthorectification and by combining counts from multiple observers (Stapleton et al. 2014). In this study, orthorectification improved detection rates for active nests from 59.9 to 67.8% in the consensus annotations, because orthorectification enhanced detectability by making some birds appear larger or by increasing contrast due to pixel smearing. In contrast, the increase in presumed non-breeders in the un-orthorectified image may reflect the impact of geometric distortions caused by factors such as camera angle, terrain relief, and sensor geometry. These findings suggest that intentionally distorting or stretching images during processing may increase detectability for certain species. Similarly, the orthorectification process stretches or compresses pixels based on the satellite capture angle and topography, which could prove valuable for detecting small-bodied species in certain environments.

We used 9 independent observers to evaluate the potential benefits of crowdsourcing. These observers identified only a proportion of the known nests while flagging multiple additional labels more than 10 m from known nests, representing either false positives or non-breeding individuals. Comparisons between each observer and the gold standard revealed that experience in remote sensing or fieldwork enhances proficiency in detecting Tristan albatrosses in satellite imagery. However, experienced observers were more likely to overlook some birds, as they tended to only select features that they were most confident in identifying. Conversely, the lower F1-scores for observers with no prior experience highlight the need for additional training to improve the accuracy of novice volunteers. While aggregating detections from multiple observers increased the detection probability, this approach could not overcome the fundamental limitation that many nesting Tristan albatrosses were undetectable in the available imagery.

Our study underlines some of the challenges in obtaining accurate counts of nesting albatrosses in satellite imagery, several of which also apply to aerial photographs. The presence of partners, failed breeders and non-breeders introduces uncertainty, particularly if they occupy empty nests (Johnstone et al. 1975). Although the number of birds on a nest pedestal can be counted more readily in aerial imagery because of the higher resolution, the contents cannot be determined. It is rarely possible to differentiate between birds on or off a nest using satellite imagery unless multiple images are taken within the same season to determine which birds remain in the same locations. The number of adults at the colony — aside from those at active nests — fluctuates daily and throughout the breeding season (Baker et al. 2023, Poncet et al. 2006). While such complications could all be addressed with advanced statistical models accounting for detection probabilities (Clement 2016, Clare et al. 2021, McKibben et al. 2023), the key requirement to have multiple repeat observations during a period of population closure is unlikely to be met given the sporadic availability of cloud-free satellite imagery on Gough Island.

Ground-based monitoring provides more granular data, including the percentage of birds sitting on nests that are incubating eggs, which ranges from 71 to 84% for Antipodean albatross *Diomedea antipodensis antipodensis* and 82 to 87% for Gibson's albatross *D. a. gibsoni* (Parker et al. 2023), and the proportion of birds at the colony that are on nests, which varied with time and date from 0.35 to 0.88 in Gibson's albatross (Elliott et al. 2024) and 0.39 to 0.75 in Antipodean albatross (Rexer-Huber et al. 2024). While future advancements in satellite technology may enhance monitoring efforts, particularly in tracking large-scale population trends, comprehensive monitoring of critically endangered species like the Tristan albatross requires not only counts but also a detailed understanding of population dynamics and drivers, which can only be achieved through long-term demographic monitoring.

5. CONCLUSION

Our findings indicate that 31 cm resolution satellite imagery is currently not a viable tool for long-term monitoring of Tristan albatross population trends, although it could potentially be used to detect a very large population decline. The persistent cloud cover at Gough Island is a major limitation, particularly at the high altitudes where Tristan albatrosses nest. Despite satellite tasking over a 2.5 mo period, obtaining a suitable image for a full census proved impossible.

The problems of limited spatial resolution and persistent cloud cover may be addressed in the near future by more and better resolution satellites. The Albedo Space Corporation (Albedo) aims to complete a constellation of 24 optical satellites with 10 cm image resolution by 2027 (Attard et al. 2024). This increased spatial resolution could make Tristan albatrosses easier to identify in the imagery. Albedo is one of several satellite providers aiming to launch more VHR earth observation satellites over the next 5 yr, with MAXAR, Airbus and others all planning a significant increase in the number of platforms. This increase in the number of satellites could alleviate the problem of high cloud cover, as more satellite capacity gives a greater chance of collecting imagery in the few times when the breeding areas are cloud free. However, until detectability improves markedly — likely requiring cross-referencing of bird locations across multiple satellite images and further validation — we recommend the continuation of current on-the-ground monitoring of Tristan albatrosses on Gough Island.

Data availability. The data and python analysis script are available on Zenodo (doi:10.5281/zenodo.14711473). This repository includes nest boundaries, satellite image tiles, satellite and ground survey data, and the python script used for data analysis. The data is also available through the Polar Data Centre at the British Antarctic Survey. The Digital Terrain Model of Gough Island, used in this study, is a WorldDEM^{TM} product from Airbus, acquired by RSPB. Due to licensing agreements, the DTM cannot be distributed outside of RSPB; however, it can be purchased directly from Airbus. The satellite image used to build the dataset is available for purchase directly from Maxar Technologies (formally DigitalGlobe).

Acknowledgements. This study is a contribution to the Ecosystems and Wildlife from Space programmes within the British Antarctic Survey *Polar Science for a Sustainable Planet* initiative, funded by the Natural Environment Research Programme. M.R.G.A. is funded by Darwin Plus (DPLUS132). We are grateful to the Tristan da Cunha Conservation Department and Island Administrators for granting permission to conduct this research. We acknowledge the Conservation Data Management Unit at the Royal Society for the Protection of Birds (RSPB), particularly Alex Whittle and Antje Steinfurth, for providing digital elevation models and geographic information, as well as RSPB for supplying breeding boundary data for Tristan albatrosses at Gough Island. Special thanks go to the Gough Island field team — Roelf Daling, Vonica Perold, Kim Stevens, Jaimie Cleeland, Kate Lawrence, and Fabrice LeBouard — for their dedication in collecting and sharing the ground counts. We sincerely thank Penny Clarke ('Expert 2') for her invaluable assistance as a reference annotator, and Cath Attard, Connor Bamford, Marcia Blyth, Hannah Cubaynes, Hana Merchant, Elizabeth Pearmain, Norman Ratcliffe, and Sally Thorpe for completing observer annotations. Finally, we thank Maxar technologies for granting permission to publish satellite image tiles and associated annotations derived from their imagery.

LITERATURE CITED

Attard MRG, Phillips RA, Bowler E, Clarke PJ, Cubaynes H, Johnston DW, Fretwell PT (2024) Review of satellite remote sensing and unoccupied aircraft systems for counting wildlife on land. Remote Sens 16:627

Baker GB, Jensz K, Cunningham R, Robertson G, Sager P, Thompson DR, Double MC (2023) Population assessment of white-capped albatrosses *Thalassarche steadi* in New Zealand. Emu 123:60–70

Barber-Meyer SM, Kooyman GL, Ponganis PJ (2007) Estimating the relative abundance of emperor penguins at inaccessible colonies using satellite imagery. Polar Biol 30:1565–1570

Bird JP, Woodworth BK, Fuller RA, Shaw JD (2021) Uncertainty in population estimates: a meta-analysis for petrels. Ecol Solut Evid 2:e12077

BirdLife International (2018) *Diomedea dabbenena*. The IUCN Red List of Threatened Species 2018:e.T22728364 A132657527. https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22728364A132657527.en

Bowler E, Fretwell PT, French G, Mackiewicz M (2020) Using deep learning to count albatrosses from space: assessing results in light of ground truth uncertainty. Remote Sens 12:2026

Brooke M de L, Bonnaud E, Dilley BJ, Flint EN and others (2018) Seabird population changes following mammal eradications on islands. Anim Conserv 21:3–12

Brothers N, Bone C, Wellbelove A (2022) Albatross population monitoring using satellite imagery, a case study: short-tailed albatross *Phoebastria albatrus* at the Senkaku Islands. Mar Ornithol 50:7–12

Caravaggi A, Cuthbert RJ, Ryan PG, Cooper J, Bond AL (2019) The impacts of introduced house mice on the breeding success of nesting seabirds on Gough Island. Ibis 161:648–661

Clare JDJ, Townsend PA, Zuckerberg B (2021) Generalized model-based solutions to false-positive error in species detection/nondetection data. Ecology 102:e03241

Clement MJ (2016) Designing occupancy studies when false-positive detections occur. Methods Ecol Evol 7: 1538–1547

Croxall JP, Rothery P, Pickering SPC, Prince PA (1990) Reproductive performance, recruitment and survival of wandering albatrosses *Diomedea exulans* at Bird Island, South Georgia. J Anim Ecol 59:775–796

Cuthbert RJ, Phillips RA, Ryan PG (2003) Separating the Tristan albatross and the wandering albatross using morphometric measurements. Waterbirds 26:338–344

Cuthbert R, Sommer E, Ryan P, Cooper J, Hilton G (2004) Demography and conservation of the Tristan albatross *Diomedea* [*exulans*] *dabbenena.* Biol Conserv 117:471–481

Daskalaki S, Kopanas I, Avouris N (2006) Evaluation of classifiers for an uneven class distribution problem. Appl Artif Intell 20:381–417

Davies D, Dilley B, Bond A, Cuthbert R, Ryan P (2015) Trends and tactics of mouse predation on Tristan albatross *Diomedea dabbenena* chicks at Gough Island, South Atlantic Ocean. Avian Conserv Ecol 10:5

del Hoyo J, Carboneras C, Jutglar F, Collar N, Kirwan GM, Garcia EFJ (2023) Tristan albatross (*Diomedea dabbenena*), Vers 1.0. In: Medrano F, Keeney BK (eds) Birds of the world. Cornell Lab of Ornithology, Ithaca, NY

Dolliver JE (2019) Using satellite imagery to count nesting albatross from space. MSc thesis, Oregon State University, Corvallis, OR

Duporge I, Isupova O, Reece S, Macdonald DW, Wang T (2021) Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. Remote Sens Ecol Conserv 7: 369–381

Dutta A, Zisserman A (2019) The VIA annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia. Association for Computing Machinery, New York, NY, p 2276–2279

Edney AJ, Wood MJ (2021) Applications of digital imaging and analysis in seabird monitoring and research. Ibis 163: 317–337

Elliott G, Walker K, Rexer-Huber K, Tinnemans J and others (2024) Gibson's wandering albatross: demography, satellite tracking and census. Final Report prepared for New Zealand Department of Conservation, Wellington

ESRI (2024) Fundamentals of pan sharpening. https://web.archive.org/web/20240724145624/https://pro.arcgis.com/en/pro-app/latest/help/analysis/raster-functions/fundamentals-of-pan-sharpening-pro.htm (accessed 12 February 2024)

Foley CM, Fagan WF, Lynch HJ (2020) Correcting for within-season demographic turnover to estimate the island-wide population of king penguins (*Aptenodytes patagonicus*) on South Georgia. Polar Biol 43:251–262

Fretwell PT, Scofield P, Phillips RA (2017) Using super-high resolution satellite imagery to census threatened albatrosses. Ibis 159:481–490

Hughes BJ, Martin GR, Reynolds SJ (2011) The use of Google EarthTM satellite imagery to detect the nests of masked boobies *Sula dactylatra.* Wildl Biol 17: 210–216

Jiménez S, Phillips RA, Brazeiro A, Defeo O, Domingo A (2014) Bycatch of great albatrosses in pelagic longline fisheries in the southwest Atlantic: contributing factors and implications for management. Biol Conserv 171: 9–20

Johnstone GW, Milledge D, Dorward DF (1975) The white-capped albatross of Albatross Island: numbers and breeding behaviour. Emu 75:1–11

Jones FM, Allen C, Arteta C, Arthur J and others (2018) Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. Sci Data 5:180124

Kluyver T, Ragan-Kelley B, Fernando P, Granger B and others (2016) Jupyter notebooks — a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B (eds) ELPUB, IOS Press, Amsterdam, p 87–90

LaRue MA, Stapleton S, Anderson M (2017) Feasibility of using high-resolution satellite imagery to assess vertebrate wildlife populations. Conserv Biol 31:213–220

Lynch HJ, White R, Black AD, Naveen R (2012) Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. Polar Biol 35:963–968

McClelland GTW, Bond AL, Sardana A, Glass T (2016) Rapid population estimate of a surface-nesting seabird on a remote island using a low-cost unmanned aerial vehicle. Mar Ornithol 44:215–220

McKibben FE, Abadi F, Frey JK (2023) To model or not to model: false positve detection error in camera surveys. J Wildl Manag 87:e22365

Muff S, Nilsen EB, O'Hara RB, Nater CR (2022) Rewriting results sections in the language of evidence. Trends Ecol Evol 37:203–210

Oppel S, Clark BL, Risi MM, Horswill C and others (2022) Cryptic population decrease due to invasive species predation in a long-lived seabird supports need for eradication. J Appl Ecol 59:2059–2070

Parker GC, Rexer-Huber K, Walker K, Elliott G (2023) Antipodean wandering albatross population study 2023. Final report to the New Zealand Department of Conservation, Wellington. Parker Conservation, Dunedin

Phillips RA, Gales R, Baker GB, Double MC and others (2016) The conservation status and priorities for albatrosses and large petrels. Biol Conserv 201:169–183

Poncet S, Robertson G, Phillips RA, Lawton K, Phalan B, Trathan PN, Croxall JP (2006) Status and distribution of wandering, black-browed and grey-headed albatrosses breeding at South Georgia. Polar Biol 29:772–781

Rexer-Huber K, Whitehead E, Parker GC, Patterson E, Walker K, Welch J, Elliott G (2024) Antipodean wandering albatrosses and white-chinned petrels 2024. Draft final report to the New Zealand Department of Conservation. Parker Conservation, Dunedin

Ryan P (2000) Separating albatrosses: Tristan or wandering? Afr Birds Birding 5:35–39

Ryan P (2023) Guide to seabirds of southern Africa. Struik Naure, Cape Town

Ryan PG, Cooper J, Glass JP (2001) Population status, breeding biology and conservation of the Tristan albatross *Diomedea* [*exulans*] *dabbenena.* Bird Conserv Int 11:35−48

Samaniego A, Jolley W, McClelland P (2022) A lesson for planning rodent eradications: interference of invasive slugs during the Gough Island mouse eradication attempt in 2021. Wildl Res 50:344−355

Stapleton S, LaRue MA, Lecomte N, Atkinson S, Garshelis D, Porter C, Atwood T (2014) Polar bears from space: assessing satellite imagery as a tool to track Arctic wildlife. PLOS ONE 9:e101513

Tickell WLN (1968) The biology of the great albatrosses, *Diomedea exulans* and *Diomedea epomophora.* Antarct Res Ser 12:1−55

Wanless RM, Ryan PG, Altwegg R, Angel A, Cooper J, Cuthbert R, Hilton GM (2009) From both sides: dire demographic consequences of carnivorous mice and longlining for the critically endangered Tristan albatrosses on Gough Island. Biol Conserv 142:1710−1718

Weimerskirch H, Delord K, Barbraud C, Le Bouard F, Ryan PG, Fretwell PT, Marteau C (2018) Status and trends of albatrosses in the French southern territories, western Indian Ocean. Polar Biol 41:1963−1972

Editorial responsibility: Brendan Godley,
University of Exeter, Cornwall Campus, UK
Reviewed by: P. Frost and 2 anonymous referees
Submitted: August 21, 2024; Accepted: February 4, 2025
Proofs received from author(s): March 30, 2025