Check for updates

RESEARCH ARTICLE

Walruses from space: walrus counts in simultaneous remotely piloted aircraft system versus very high-resolution satellite imagery

Hannah C. Cubaynes[1] (iD), Jaume Forcada[1], Kit M. Kovacs[2], Christian Lydersen[2], Rod Downie[3] & Peter T. Fretwell[1] (iD)

[1]British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK
[2]Norwegian Polar Institute, Fram Centre, 9296, Tromsø, Norway
[3]WWF-UK, Living Planet Centre, Rufford House, Brewery Road, Woking GU21 4LL, UK

Abstract

Regular counts of walruses (*Odobenus rosmarus*) across their pan-Arctic range are necessary to determine accurate population trends and in turn understand how current rapid changes in their habitat, such as sea ice loss, are impacting them. However, surveying a region as vast and remote as the Arctic with vessels or aircraft is a formidable logistical challenge, limiting the frequency and spatial coverage of field surveys. An alternative methodology involving very high-resolution (VHR) satellite imagery has proven to be a useful tool to detect walruses, but the feasibility of accurately counting individuals has not been addressed. Here, we compare walrus counts obtained from a VHR WorldView-3 satellite image, with a simultaneous ground count obtained using a remotely piloted aircraft system (RPAS). We estimated the accuracy of the walrus counts depending on (1) the spatial resolution of the VHR satellite imagery, providing the same WorldView-3 image to assessors at three different spatial resolutions (i.e., 50, 30 and 15 cm per pixel) and (2) the level of expertise of the assessors (experts vs. a mixed level of experience – representative of citizen scientists). This latter aspect of the study is important to the efficiency and outcomes of the global assessment programme because there are citizen science campaigns inviting the public to count walruses in VHR satellite imagery. There were 73 walruses in our RPAS 'control' image. Our results show that walruses were under-counted in VHR satellite imagery at all spatial resolutions and across all levels of assessor expertise. Counts from the VHR satellite imagery with 30 cm spatial resolution were the most accurate and least variable across levels of expertise. This was a successful first attempt at validating VHR counts with near-simultaneous, in situ, data but further assessments are required for walrus aggregations with different densities and configurations, on different substrates.

Introduction

Walruses (*Odobenus rosmarus*) live throughout much of the circumpolar Arctic, where they use sea ice to meet various biological needs such as giving birth and resting (Fay, 1982; Born et al., 1995). However, sea ice in the Arctic is declining at a rapid rate in terms of extent, thickness and seasonal presence (Meredith et al., 2019). Understanding how these important habitat modifications are affecting walruses is essential for management authorities to safeguard their existence. Ideally, walrus monitoring would be performed on a regular basis to capture impacts of the rapid and highly dynamic changes in sea ice conditions.

During late summer and early autumn, when sea ice is at a seasonal minimum, walruses tend to return to terrestrial haul-out sites that they have previously occupied. Therefore, counting them at their terrestrial haul-out sites during this period is the preferred abundance assessment method (e.g., Fischbach et al., 2022; Hammill et al., 2016; Kovacs et al., 2014; Lydersen et al., 2008). Regional population surveys are conducted in some areas for walruses, but the logistical challenges (including financial costs) limit the frequency and extent of such surveys (Fischbach et al., 2016; Kovacs et al., 2014). Satellite imagery has the potential to be a non-invasive solution to monitor large areas more frequently. However, further work is required to better understand what level of monitoring (e.g., haul-out occupation, herd size, abundance) is feasible using space-borne technologies. Earth-observing satellites that orbit the Earth are capable of imaging the whole walrus distribution range within one season. For instance, the Maxar constellation of very high-resolution (VHR) multispectral satellites captures images below 50 cm spatial resolution (i.e., GeoEye-1, WorldView-2 and WorldView-3) and could capture images of all known walrus haul-out sites in a 3 month period in ideal conditions, with the possibility for regional populations to be imaged within much shorter time frames (pers. comm. Maxar, 2022). Other VHR satellites, such as those operated by Airbus and Planet, may offer similar capabilities (Airbus, 2022; Planet, 2018). As more VHR satellites are launched into space and become operational (Clarke et al., 2021), the time coverage is expected to improve, which would facilitate synoptic surveys on shorter time scales. If VHR imagery is not required for a particular study (e.g., assessing haul-out site occupation), satellites with lower spatial resolution can be used, such as Sentinel-2 (10 m resolution; ESA, 2015) operated by the European Space Agency, or PlanetScope (3 m resolution; Planet, 2023) operated by Planet, both of which provide daily imagery for Arctic areas.

Very high-resolution satellites have already proved to be a useful complementary tool for monitoring wildlife such as penguins (Fretwell et al., 2012; Lynch & LaRue, 2014) and elephant seals (*Mirounga* spp. Laborie et al., 2023), and for finding walruses (Boltunov et al., 2012; Burn & Cody, 2005; Fischbach & Douglas, 2021; Matthews et al., 2022; Zinglersen et al., 2020). This new tool could provide a non-invasive mean to assess walruses, which can be disturbed easily by noise from boats or aircrafts traditionally used in surveys (Born et al., 2021). However, work is needed regarding the feasibility of using VHR satellite imagery to count the number of individuals in haul-out groups accurately. In particular, it is necessary to validate the counts in satellite imagery

with near-simultaneous aerial imagery that can provide an accurate baseline count (Fischbach & Douglas, 2021; Sherbo et al., 2023). Some previous attempts to count walruses from space (Boltunov et al., 2012; Zinglersen et al., 2020) could not provide validation of their counts due to the difficulty of acquiring near-simultaneous aerial imagery. Other studies (Fischbach & Douglas, 2021; Sherbo et al., 2023) were able to pair their satellite imagery with aerial imagery acquired 1.5–9 h apart from each other, but uncertainty in the accuracy of the satellite imagery herd size estimate remained due to the known dynamic nature of walrus aggregations and the likelihood of different numbers of walruses being present between the satellite and aerial imagery (Fischbach & Douglas, 2021).

Further considerations for counting walruses across a whole population, sub-species, or their entire pan-Arctic distribution using VHR include the feasibility of efficiently reviewing vast numbers of images. Two potential solutions are to use automated systems, or the power of the crowd, through citizen science projects. No automated systems currently exist, due to the requirements for large amount of training data, particularly for Convolutional Neural Networks, which are proving successful for the detection of wildlife in VHR satellite imagery (Borowicz et al., 2019; Bowler et al., 2020; Green et al., 2023). The other solution resides with citizen science campaigns. Recently, Weddell seals (*Leptonychotes weddellii*) were censused across the ice shelves surrounding the entire Antarctica continent through a crowdsourcing campaign (LaRue et al., 2020). A similar campaign, Walrus from Space (WWF, 2023; https://www.wwf.org.uk/learn/walrus-from-space), is currently inviting citizen scientists to search for and count Atlantic walruses (*O. rosmarus rosmarus*) in VHR satellite imagery. However, walrus counts made by the public need to be validated.

In this study, we aimed to validate walrus counts made by experts and citizen scientists in satellite imagery, using a 'control count' obtained using a remotely piloted aircraft system (RPAS) captured simultaneously with the satellite imagery. We assessed the accuracy of the walrus counts depending on the spatial resolution of the VHR satellite imagery (i.e., 50, 30 and 15 cm), and the experience level of the assessors (experts with or without field knowledge and non-expert/citizen scientists).

Materials and Methods

Study location

Walruses regularly haul out during summer at Sarstangen (78.72° N, 11.44° E), a narrow band of light beige and grey sand and gravel on the western shores of Spitsbergen

in the Svalbard Archipelago, Norway, extending into Forlandsundet (Fig. 1).

RPAS and satellite imagery acquisition

On 15 July 2022, we visited Sarstangen and flew a DJI Mavic 3 (Hasselblad camera sensor: 4/3 CMOS, effective pixels: 20 MP, focal length: 12 mm, image width: 5280 pixels, image height: 3956 pixels; for further detailed specifications, see: https://www.dji.com/uk/mavic-3/specs) under

Permit No. 22/00507-2 from the Governor of Svalbard (RiS number 11906). Palomino-González et al. (2021) tested the level of disturbance in response to RPAS flights undertaken above walruses in Svalbard and observed no disturbance when RPAS were flown at altitudes above 50 m. We added a precautionary 5 m and flew at 55 m above the walruses in our study. We observed no disturbance (e.g., no head lifting) during our RPAS flights within view of the remote pilot and observers on the ground. We piloted the RPAS to take off and land downwind, at
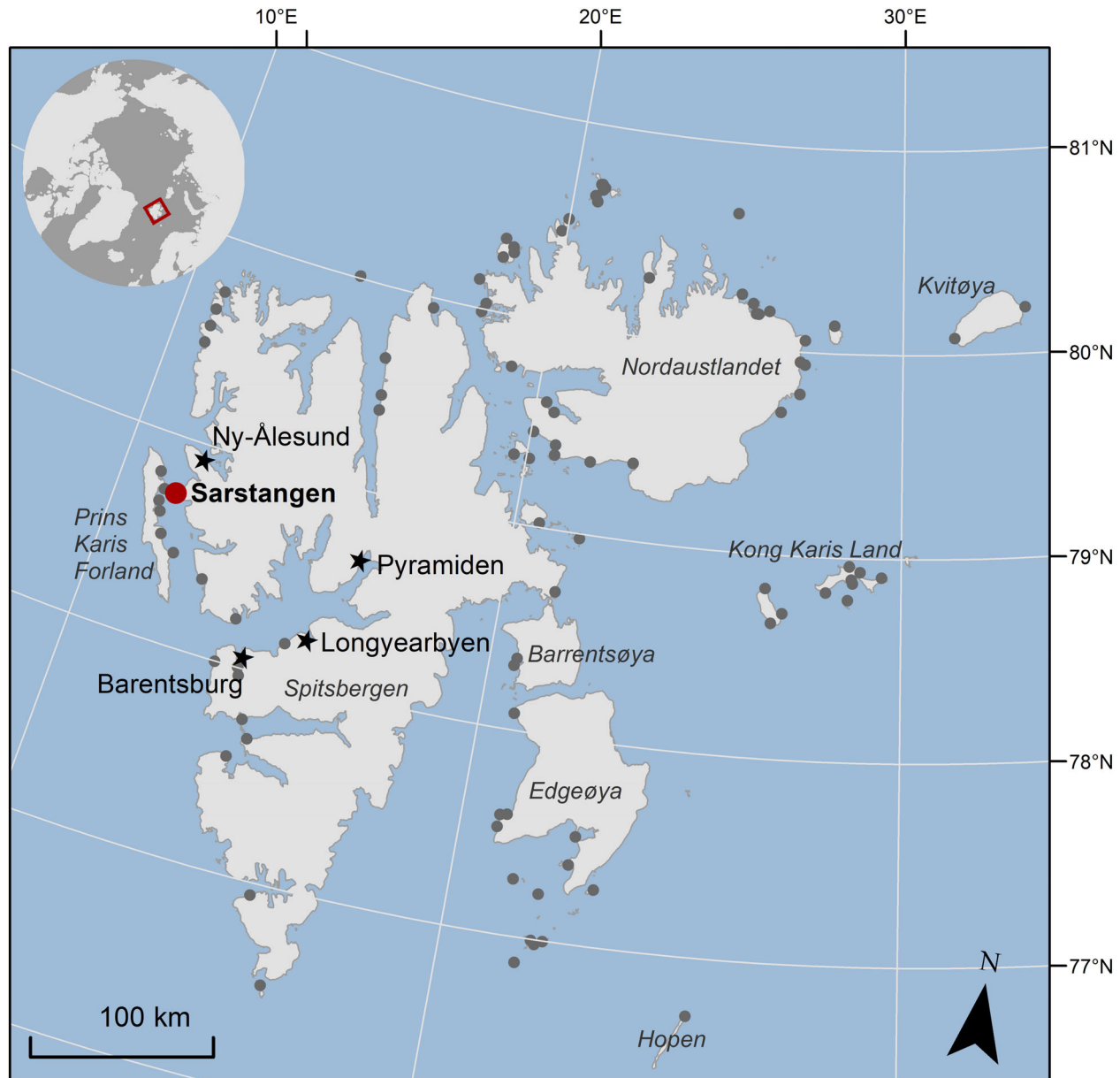

Figure 1. Map showing the Sarstangen walrus haul-out site, where RPAS and VHR satellite imagery were acquired (large red dot). Other known walrus haul-out sites are shown as small grey dots and human settlements are indicated by black stars. Basemap and walrus haul-out sites © Norwegian Polar Institute (2014, 2024).

distances greater than 300 m from the walruses. Herein, we used an image of the haul-out site that was captured at 13:10 UTC on 15 July (2022-07-15T13:10Z; Table 1) because it was closest in time to the satellite image acquisition. The image was captured at nadir.

A WorldView-3 image of the Sarstangen walrus haul-out site was captured on 15 July 2022, at 13:25 UTC (2022–07-15T13:25Z), 15 min after the RPAS image (catalogue ID: 104001007777EA00; product type: ortho-ready standard 2A). We were on-site at Sarstangen maintaining continuous visual observations of the walruses during and somewhat after capturing the satellite imagery – and group composition and number had not changed from the time of the RPAS-captured image we used. We obtained the same satellite image at three different spatial resolutions: 50 cm (using Maxar Technologies' downsampling algorithm), 30 cm (the raw resolution) and 15 cm (using Maxar Technologies' high-definition algorithm; Maxar Technologies, 2023; Yalcin et al., 2021; Table 1). For all satellite images, the mean sun azimuth angle was 214.3°, the mean sun elevation angle was 31° and the off-nadir angle was 14°. All three satellite images were pansharped using the Brovey algorithm and standard deviations stretch, as it rendered the clearest image to discern individual walruses.

Counting walruses

Walruses hauled out at Sarstangen were counted on the RPAS image and the three satellite images (15, 30 and 50 cm spatial resolution; Table 1) by four different groups of assessors:

- Group 1: five assessors with a high level of expertise in counting wildlife in aerial and, or VHR satellite imagery, with field knowledge (including the field researchers and RPAS operators of the present study);
- Group 2: five assessors with experience in photographic counts of wildlife but with no field knowledge;
- Group 3: nine assessors, without any experience in counting walruses and, or other wildlife in aerial and, or VHR satellite imagery;
- Group 4: 36 assessors, representative of the crowd with mixed levels of experience in counting walruses and, or wildlife in aerial and, or in VHR satellite imagery.

Assessors from Groups 1, 2 and 3 counted walruses in the four images, sequentially and independently, from lowest (50 cm satellite image) to highest spatial resolution (RPAS image). Assessors from Group 4 reviewed one of the four images, assigned to them at random, to replicate what is most likely to happen in citizen science projects, where assessors review some imagery but not all. Overall, each of the four images was reviewed by nine assessors from Group 4 selected at random. All assessors counted walruses following the same protocol (see Data S1), using the open-source software VGG Image Annotator (Dutta & Zisserman, 2019). Only the walruses hauled out were included in the count. A walrus was considered hauled out, when it was completely on land or grounded in the shallow waters (Kovacs et al., 2014). All counts per assessor and per image type are available in Data S2.

Analysis 1: effect of image spatial resolution and level of experience

We tested the effects of spatial resolution of the image and the experience level of assessors on count accuracy by evaluating the variation and bias between assessors. We used the

Table 1. RPAS image and WorldView-3 image processed at three different spatial resolutions (i.e., 15, 30 and 50 cm) analysed in this study.
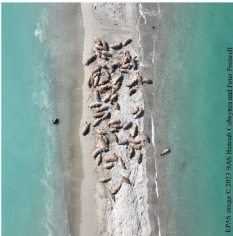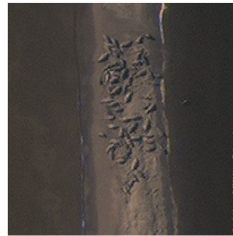
Date and Time (UTC, YYYY-MM-DDTHH:mmZ)	2022-07-15T13:10Z	2022-07-15T13:25Z	2022-07-15T13:25Z	2022-07-15T13:25Z
Platform type	RPAS	VHR satellite	VHR satellite	VHR satellite
Platform model	Mavic 3	WorldView-3 (HD uplift algorithm)	WorldView-3	WorldView-3 (downsampling algorithm)
Spatial resolution	1.5 cm	15 cm	30 cm	50 cm
Image				

Image sources: RPAS image © 2023 BAS Hannah Cubaynes and Peter Fretwell (Data S4), Satellite images © 2023 Maxar Technologies.
RPAS, remotely piloted aircraft system; VHR, very high-resolution.

counts from Group 1, Group 2 and Group 3, which included 74 counts in total. We used a factorial design with two variables: 'experience' with three levels (Group 1, Group 2, Group 3) and 'image spatial resolution' with four levels (RPAS, VHR 15, VHR 30, VHR 50) to assess impacts of image resolution and experience. The experimental unit was 'assessor', nested in 'experience' with an unbalanced design since Group 1 and Group 2 had five assessors each and Group 3 comprised nine assessors. Group 1 acted as a control group with accurate counts to model the effects of 'experience', as the assessors from this group had previously obtained a ground-truthed count in the field (the knowledge of the total number of walruses in the RPAS image), which had a feedback effect when all the subsequent images were counted. The consensus walrus count for the RPAS imagery provided by Group 1 acted as the ground-truthed count.

Based on inspection of the data distribution, we modelled walrus counts in images as log-normal random variables $Y \sim \mathrm{LN}(\alpha, \sigma^2)$ with mean $E(Y) = \mu = e^{\alpha + \sigma^2/2}$ and variance $\mathrm{var}(Y) = e^{2\alpha + \sigma^2}(e^{\sigma^2} - 1)$. In a linear mixed-effects modelling formulation, we expressed the count of 'assessor' k from 'experience' group i of 'image spatial resolution' j as:

$$\log\left(\mu_{i,j,k}\right) = \beta_0 + \beta_{1,i} + \beta_{2,j} + \beta_{3,i,j} + \gamma_{kj} + \alpha_{k,i} + \epsilon_{i,j,k}$$

$$\gamma_{kj} \sim N\left(0, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}\right)$$

$$\alpha_{k,i} \sim N(0, \boldsymbol{\Psi})$$

$$\epsilon_{i,j,k} \sim N\left(0, \sigma^2 \Lambda_{ki}\right)$$

where β_0 is the overall mean; $\beta_{1,i}$ is a fixed-effect for 'experience', with $i = 1, \ldots, 3$; $\beta_{2,j}$ is a fixed-effect for 'image spatial resolution', with $j = 1, \ldots, 4$; $\beta_{3,i,j}$ is a fixed-effect interaction term; γ_{kj} is a random intercept for 'image spatial resolution' j within I 'assessor' k, with $k = 1, \ldots, K$; $\alpha_{k,i}$ is a random effect for 'assessor' k nested in 'experience' group i; $\epsilon_{i,j,k}$ is a within-group error term; Λ_{ki} are positive-definite matrices to model heteroscedasticity (see below); $\sigma_{\boldsymbol{\gamma}}^2$ is the 'assessor'-specific effects variance; $\boldsymbol{\Psi}$ is a positive-definite symmetric variance–covariance matrix of grouped between-assessor residuals.

To evaluate differences between experience levels and image spatial resolution levels, we set Group 1 and RPAS as reference terms in the corresponding fixed-effects contrasts matrices,

$$X_i = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } X_j = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We tested for heterogeneity in between-assessor variance using specific positive-definite matrix structures of the random effects. We built models with a general symmetric matrix and with a simplified (diagonal only) matrix, with 'experience' and 'image spatial resolution' as grouping variables ($g_{i,j}$):

$$\boldsymbol{\Psi}^S = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_{1,2}} & \sigma_{g_{1,3}} \\ \sigma_{g_{2,1}} & \sigma_{g_2}^2 & \sigma_{g_{2,3}} \\ \sigma_{g_{3,1}} & \sigma_{g_{3,2}} & \sigma_{g_3}^2 \end{bmatrix}, \quad \boldsymbol{\Psi}^D = \begin{bmatrix} \sigma_{g_1}^2 & 0 & 0 \\ 0 & \sigma_{g_2}^2 & 0 \\ 0 & 0 & \sigma_{g_3}^2 \end{bmatrix}.$$

For each grouping variable, we compared models with each of these structures and a model with an intercept-only random effect for 'assessor'. Models with $\boldsymbol{\Psi}^S$ evaluated the correlation in residuals between levels of grouping variables, and models with $\boldsymbol{\Psi}^D$ tested the hypothesis that $\sigma_{g_1}^2 \neq \sigma_{g_2}^2 \neq \sigma_{g_3}^2$ (i.e., uncorrelated between-assessor heterogeneity).

To assess heterogeneity in within-assessor variance (heteroscedasticity), we selected competing variance function models (Λ_{ki}) after inspection of patterns in the residuals of the mixed-effects models with better fit. We considered:

- $\mathrm{var}(\epsilon_{i,j,k}) = \sigma^2 \delta_{g_{k,i}}^2 \rightarrow$ different variances by factor levels g_i
- $\mathrm{var}(\epsilon_{i,j,k}) = \sigma^2 |v_{k,i,j}|^{2\delta} \rightarrow$ variance as power of a covariate ($v_{k,i,j}$)
- $\mathrm{var}(\epsilon_{i,j,k}) = \sigma^2 e^{2\delta v_{k,i,j}} \rightarrow$ variance exponential of a covariate ($v_{k,i,j}$)

where δ is a vector of variance parameters of the grouping variable, and $v_{k,i}$ is a variance covariate derived from the fitted values of a model which was updated during the fitting process.

We fitted models using maximum-likelihood methods in the nlme package (Pinheiro & Bates, 2000; Pinheiro, Bates, & R Core Team, 2023) in R (R Core Team, 2023). Given the relatively low sample sizes, we used a set of models with reduced parameters and the small sample size Akaike information criterion ($\mathrm{AIC_c}$) for multi-model selection and inference (Burnham & Anderson, 2002). For a given set of models, we obtained Δ_i as $\mathrm{AIC_{c}}_i - \mathrm{minAIC_c}$, where $\mathrm{minAIC_c}$ is the minimum $\mathrm{AIC_c}$ from the model set, to calculate 'Akaike weights' for each model i. These were $w_i = \exp\left(-\frac{1}{2}\Delta_i\right) / \sum_{r=1}^{R} \exp\left(-\frac{1}{2}\Delta_r\right)$. For inference, we obtained averaged estimates of parameters of interest as $\widehat{\overline{\theta}} = \sum_{r=1}^{R} w_i \widehat{\theta}_r / \sum_{r=1}^{R} w_i$, where $\widehat{\theta}_r$ is the parameter value estimated with model r. We also assessed model fit using residual plots for heteroscedasticity, normal probability plots and numerical summaries based on approximate confidence intervals of parameters (see Data S3).

To evaluate bias in walrus counts derived from 'image spatial resolution' and 'experience', we used predictions from the best-fit models for each analysis. Predictions were expressed at the counting scale, $E(Y)$, by exponentiation and using an appropriate bias correction, $e^{\widehat{\alpha} + \widehat{\sigma^2}/2}$. This method was expected to perform well because counts were relatively large, without zero values and they had relatively small dispersion (O'Hara & Kotze, 2010).

For each level in 'experience' and 'image spatial resolution', we obtained $\text{bias}\left(\widehat{Y}\right) = E_y\left(\widehat{Y}\right) - Y$ and mean squared error (MSE) as $E_y\left[\left(\widehat{Y} - Y\right)^2\right] = \text{var}_y\left(\widehat{Y}\right) + \text{bias}^2\left(\widehat{Y}\right)$, where Y is the RPAS ground-truthed count. To obtain a robust estimate of $\text{var}_y\left(\widehat{Y}\right)$ we used stratified non-parametric balanced bootstrap resampling (Davison et al., 1986), using 'assessor' within 'experience' group as resampling unit. In each of $B = 2000$ bootstrap samples, each assessor in each experience group appeared exactly B times in the union of the bootstrap samples, which maintained the natural hierarchy of the data. This method had a slightly better nominal coverage of model parameters by the confidence intervals than a nonparametric residual bootstrap method (Carpenter et al., 2003) and performed better in simulations (data not presented) due to the simplified random effects structure of our models. At each simulation, model predictions of $\log\left(Y_{i,j}\right)$ were obtained from model-averaged parameter estimates to preserve uncertainty in model selection.

Analysis 2: effect of the crowd on walrus count accuracy at different spatial resolutions

To generate a skill pool equivalent to crowdsourcing campaigns, where citizen scientists review a portion of the imagery, we combined the counts from Group 2, Group 3 and Group 4 into a group named Crowd. Because this group had a mixture of assessors counting walruses in all four images (14 assessors) and assessors counting walruses in one of the four images (36 assessors), it was an unbalanced design with 50 assessors and 90 counts in total (excluding the same two outliers mentioned in the section above). To model data dependence in the predictors, we retained 'assessor' as a grouping variable.

We expressed the count of 'assessor' k per 'image spatial resolution' j as:

$$\log\left(\mu_{j,k}\right) = \beta_0 + \beta_{1,j} + \gamma_{kj} + \alpha_{k,j} + \epsilon_{j,k}$$

$$\gamma_{kj} \sim N\left(0, \sigma_\gamma^2 \mathbf{I}\right)$$

$$\alpha_{k,j} \sim N\left(0, \sigma_\alpha^2 \mathbf{I}\right)$$

$$\epsilon_{j,k} \sim N\left(0, \sigma^2 \Lambda_k\right)$$

where β_0 is the overall mean; $\beta_{1,j}$ is a fixed-effect for 'image spatial resolution', with $j = 1, \ldots, 4$; γ_{kj} is a random intercept for 'image spatial resolution' j within 'assessor' k, with $k = 1, \ldots, K$; $\alpha_{k,j}$ is a random intercept for the interaction 'assessor' k and 'image spatial resolution'; $\epsilon_{j,k}$ is a within-group error term; Λ_{kj} are positive-definite matrices to model heteroscedasticity (see below); σ_γ^2 is the assessor-specific effects variance; σ_α^2 is the variance of the interaction 'assessor' and 'image spatial resolution'.

We tested for heterogeneity in between-assessor and within-assessor variances, similar to Analysis 1.

We fitted, selected and evaluated the models tested here, in the same manner as in Analysis 1. Inference and bias assessment were also evaluated using the same methods as used for Analysis 1, with the difference that 'assessor' was the main resampling unit for the stratified non-parametric balanced bootstrap resampling, though we retained the grouping by 'experience', despite it not being explicitly modelled, to preserve the original data structure in each bootstrap simulation.

Walrus herd density

We used a semi-automated method to determine the herd density of walruses in the RPAS image using ESRI ArcMap 10.8 (ESRI, 2023). First, the RPAS image was georeferenced to the satellite image (Georeferencing tool in ArcMap). This step was required due to the inherent inaccuracies in the spatial referencing of satellite and RPAS images. One expert placed points in the middle of each individual walrus present in the RPAS image. Then, we constructed a convex hull around each point to draw an outline around the group of walruses (Minimum Bounding Geometry tool with the option of Convex Hull in ArcMap). As this outline included only half of the body of some walruses, we then used the Buffer tool to ensure that all walruses were included in the outline, using a buffer of 1.5 m, as the average size of an adult walrus is 3 m (2.7 m for females and 3.2 m for males; Lydersen, 2018).

The herd density was estimated as the quotient of the number of walruses within the outline divided by the area in m^2 of the outline of the group of walruses.

Results

Model selection

For Analysis 1, we tested 11 models to investigate the effect of 'image spatial resolution' and 'experience'

Table 2. Selection based on AIC_c, ΔAIC_c and AIC_c weight (w) statistics of models with effects of 'experience' and 'image spatial resolution' on walrus counts.

Model	Structure	df	log $\mathscr{L}$	AIC_c	ΔAIC_c	w
11	$\beta_0 + \beta_{2,j} + \beta_{3,i,j}; \mathrm{var}(\epsilon_{i,j,k}) = \sigma^2 \delta_{g_{k,j}}^2$	16	100.6	−159.7	0	0.845
8	$\beta_0 + \beta_{2,j} + \beta_{3,i,j} + \gamma_{kj}; \mathrm{var}(\epsilon_{i,j,k}) = \sigma^2 \delta_{g_{k,j}}^2$	17	100.6	−156.3	3.38	0.155
10	$\beta_0 + \beta_{2,j} + \beta_{3,i,j} + \gamma_{kj}; \mathrm{var}(\epsilon_{i,j,k}) = \sigma^2 e^{2\delta v_{i,j,k}}$	15	80.99	−123.7	35.96	0
9	$\beta_0 + \beta_{2,j} + \beta_{3,i,j} + \gamma_{kj}; \mathrm{var}(\epsilon_{i,j,k}) = \sigma^2 |v_{k,i,j}|^{2\delta}$	15	80.46	−122.6	37.03	0
6	$\beta_0 + \beta_{2,j} + \beta_{3,i,j} + \gamma_{kj}$	14	65.83	−96.5	63.12	0
5	$\beta_0 + \beta_{2,j} + \beta_{3,i,j} + \alpha_{k,i}; \sigma_{g_1}^2 \neq \sigma_{g_2}^2 \neq \sigma_{g_3}^2$	16	67.47	−93.4	66.26	0
1	$\beta_0 + \beta_{2,j} + \beta_{3,i,j} + \gamma_{kj} + \alpha_{k,i}$	15	65.83	−93.4	66.28	0
7	$\beta_0 + \beta_{2,j} + \beta_{3,i,j} + \alpha_{k,i}; \sigma_{g_1}^2 \neq \sigma_{g_2}^2 \neq \sigma_{g_3}^2; \mathrm{var}(\epsilon_{i,j,k}) = \sigma^2 \delta_{g_{k,j}}^2$	18	70.13	−91.8	67.84	0
4	$\beta_0 + \beta_{2,j} + \beta_{3,i,j} + \alpha_{k,i}$	19	67.47	−82.9	76.79	0
2	$\beta_0 + \beta_{1,i} + \beta_{2,j} + \gamma_{kj} + \alpha_{k,i}$	9	50.01	−79.2	80.45	0
3	$\beta_0 + \beta_{2,j} + \gamma_{kj} + \alpha_{k,i}$	7	44.68	−73.7	86.01	0

Between-group and within-group ($\epsilon_{i,j,k}$) variances are indicated next to the main model equation, when selected. (See text for parameter structure and definitions.) df are model degrees of freedom and log$\mathscr{L}$ is the log-likelihood value. Models are ranked according to best fit based on lowest AIC_c, from top to bottom.

Table 3. Selection based on AIC_c, ΔAIC_c and AIC_c weight (w) statistics of models with effects of 'image spatial resolution' on walrus counts.

Model	Structure	df	log $\mathscr{L}$	AIC_c	ΔAIC_c	w
8c	$\beta_0 + \beta_{1,j}; \mathrm{var}(\epsilon_{j,k}) = \sigma^2 \delta_{g_{k,j}}^2$	8	99.56	−181.3	0.00	0.775
3c	$\beta_0 + \beta_{1,j} + \gamma_{kj}; \mathrm{var}(\epsilon_{j,k}) = \sigma^2 \delta_{g_{k,j}}^2$	9	99.56	−178.9	2.47	0.225
4c	$\beta_0 + \beta_{1,j} + \gamma_{kj}; \mathrm{var}(\epsilon_{j,k}) = \sigma^2 e^{2\delta v_{j,k}}$	7	86.31	−157.3	24.10	0
5c	$\beta_0 + \beta_{1,j} + \gamma_{kj}; \mathrm{var}(\epsilon_{j,k}) = \sigma^2 |v_{k,j}|^{2\delta}$	7	85.50	−155.6	25.71	0
6c	$\beta_0 + \beta_{1,j} + \gamma_{kj} + \alpha_{k,j}; \mathrm{var}(\epsilon_{j,k}) = \sigma^2 e^{2\delta v_{j,k}}$	8	86.31	−154.8	26.51	0
1c	$\beta_0 + \beta_{1,j} + \gamma_{kj}$	6	45.81	−93.4	102.75	0
2c	$\beta_0 + \beta_{1,j} + \gamma_{kj} + \alpha_{k,j}$	7	45.81	−76.2	105.10	0
7c	$\beta_0 + \beta_{1,j} + \gamma_{kj} + \alpha_{k,j}; \mathrm{var}(\epsilon_{j,k}) = \sigma^2 |v_{k,j}|^{2\delta}$	8	45.81	−73.8	107.51	0

Within-group ($\epsilon_{j,k}$) variances are indicated next to the main model equation, when selected. (See text for parameter structure and definitions.) df are model degrees of freedom and log$\mathscr{L}$ is the log-likelihood value. Models are ranked according to best fit based on lowest AICc, from top to bottom.

(Table 2) on walrus count accuracy. Models 8 and 11 (Table 2) provided the best fit. For Analysis 2, we tested eight models to assess the effect of the crowd on walrus count accuracy (Table 3). Models 3c and 8c had the best fit. These models that provided the best fit for Analyses 1 and 2 were used to estimate parameters of interest in Table 4. More details on model selection are provided in Data S3.

Count accuracy

Results for both Analyses 1 and 2 are presented together to facilitate comparison, even though they were obtained independently. Because the variance in counts was generally low, the (root) mean square error was very close to the estimates of bias and both indices showed similar patterns of change across experience levels (Table 4). As expected, the lowest bias was for the walrus counts in the RPAS image, across all groups of assessors (Table 4). For

Group 1, estimates of bias and MSE for the RPAS image were 0, which means that the modelled parameter estimates ($\hat{\theta}$) predicted the observed counts with perfect accuracy. This was expected from the experimental design and feedback effects in counts and also denoted the very small individual measurement errors in this group. As the feedback effects carried over for the counts in VHR imagery, Group 1 also produced the most accurate counts for the other image types. Group 1 served as the ground-truthed count ($Y = 73$ walruses).

When looking at Group 2, Group 3 and the Crowd, all assessors across these three groups undercounted walruses across all levels of spatial resolution (Fig. 2). There was less bias for the counts in VHR 30 (Table 4). The highest variance was observed for VHR 15 across all groups (Table 4).

For the observed counts, Group 2 (experts) showed less variation than Group 3 (mixed level of experience), but variations in counts were similar between both groups,

Table 4. Observed $\left(\text{bias}(\overline{Y})\right)$ and estimated $\left(\text{bias}(\widehat{Y})\right)$ bias, mean squared error (MSE) and root mean square error (RMSE) of walrus counts from images from different platforms and resolutions obtained by assessors with different experience levels (first analysis); and by a crowd (second analysis).

Image type	Experience	bias($\overline{Y}$)	bias($\widehat{Y}$)	%bias($\widehat{Y}$)	var$_y$($\widehat{Y}$)	MSE	$\sqrt{\text{MSE}}$
RPAS	Group 1	0.00 (0.00)	0.00 [0.00; 0.01]	0.00	0.00	0.00	0.00
VHR 15	Group 1	−8.00 (5.96)	−7.94 [−12.55; −3.99]	−10.88	4.98	68.08	8.25
VHR 30	Group 1	−5.60 (5.18)	−5.23 [−9.53; −1.72]	−7.17	3.80	31.18	5.58
VHR 50	Group 1	−5.60 (4.51)	−5.29 [−8.95; −2.49]	−7.25	2.77	30.79	5.55
RPAS	Group 2	0.20 (0.45)	0.20 [0.00; 0.60]	0.28	0.03	0.07	0.26
VHR 15	Group 2	−26.00 (2.12)	−25.88 [−27.33; −24.26]	−35.45	0.61	670.22	25.89
VHR 30	Group 2	−16.20 (8.23)	−16.16 [−20.79; −9.11]	−22.14	8.94	270.07	16.43
VHR 50	Group 2	−27.80 (7.92)	−28.06 [−34.20; −21.61]	−38.44	9.57	796.87	28.23
RPAS	Group 3	0.33 (1.12)	0.31 [−0.34; 0.89]	0.42	0.12	0.22	0.47
VHR 15	Group 3	−22.14 (6.01)	−22.04 [−25.91; −17.87]	−30.19	4.38	490.13	22.14
VHR 30	Group 3	−19.00 (8.72)	−19.16 [−24.48; −14.18]	−26.25	7.17	374.30	19.35
VHR 50	Group 3	−27.00 (6.84)	−27.16 [−31.14; −23.10]	−37.21	4.15	742.05	27.24
RPAS	Crowd	0.13 (0.76)	0.12 [−0.15; 0.46]	0.17	0.02	0.04	0.2
VHR 15	Crowd	−25.24 (5.66)	−24.92 [−26.98; −22.85]	−34.13	1.19	622.1	24.94
VHR 30	Crowd	−18.83 (8.62)	−18.62 [−22.28; −15.20]	−25.50	3.10	349.69	18.7
VHR 50	Crowd	−23.04 (12.47)	−23.96 [−28.34; −19.61]	−32.83	5.10	579.41	24.07

In squared brackets are estimated 95% bootstrap confidence intervals, and in parentheses are standard deviations for estimated and observed bias respectively.

MSE, mean square error; RPAS, remotely piloted aircraft system; VHR, very high-resolution.

when looking at the predicted counts (Fig. 2). Model predictions performed well in terms of precision with appropriate sample size (i.e., Group 3; $n = 9$) and underperformed otherwise (Group 2; $n = 5$ assessors; Fig. 2). The Crowd showed more variation in the observed counts than any other groups, due in part to its reasonably high sample size ($n = 50$), which was the highest among all groups (Fig. 2).

When comparing the experience level of the assessors (Group 2 and Group 3), there was less bias and variance for experts (Group 2) for the satellite image with 30 cm resolution. For the satellite images with 50 cm resolution, both groups showed similar estimated bias and variance, and for the satellite image with 15 cm resolution, there was less bias and variance for Group 3 (mixed level of experiences).

Walrus herd density

The semi-automated perimeter drawn around the walruses (blue full line in Fig. 3) is 839.40 m^2 and includes 73 individual walruses. Herd density was thus 0.09 walrus/m^2.

Discussion

To the best of our knowledge, this study is the first to successfully obtain near-simultaneous aerial- and space-borne sensor imagery of a walrus herd resting on shore. The time difference of 15 min between the satellite and RPAS imagery and the presence of observers on the ground providing direct observations of the walrus herds during these 15 min supported the assumption that no changes in the number of walruses or composition occurred between the image captures. Other studies have tried to obtain satellite imagery captured during aerial surveys and were able to pair them within 1.5 h (Sherbo et al., 2023), or 5–9 h (Fischbach & Douglas, 2021). However, the dynamic fluxes in herd attendance limited their inference in validating their interpretation of herd size.

This study demonstrated that VHR satellite imagery can be used to estimate abundance of hauled-out walruses with reasonable accuracy when compared to total counts from on-site RPAS-facilitated images (i.e., total ground truthing). Satellite imagery with a 30 cm spatial resolution provided the most accurate results (compared with 15 cm and 50 cm resolution). This is intuitive with respect to the 50 cm resolution because it provides less detail, making it more difficult to be confident in the detection (see also Cubaynes et al., 2019, 2023; Fretwell et al., 2014). It was a bit more surprising that the 15 cm imagery had a lower count accuracy. This is likely because the 15 cm imagery is actually only a modified version of the 30 cm image using a high-definition algorithm, which may create artefacts that do not match the reality on the ground. Further work exploring the value of such algorithms is needed. A satellite image with a raw resolution of 15 cm would likely provide more accurate results. Currently, no
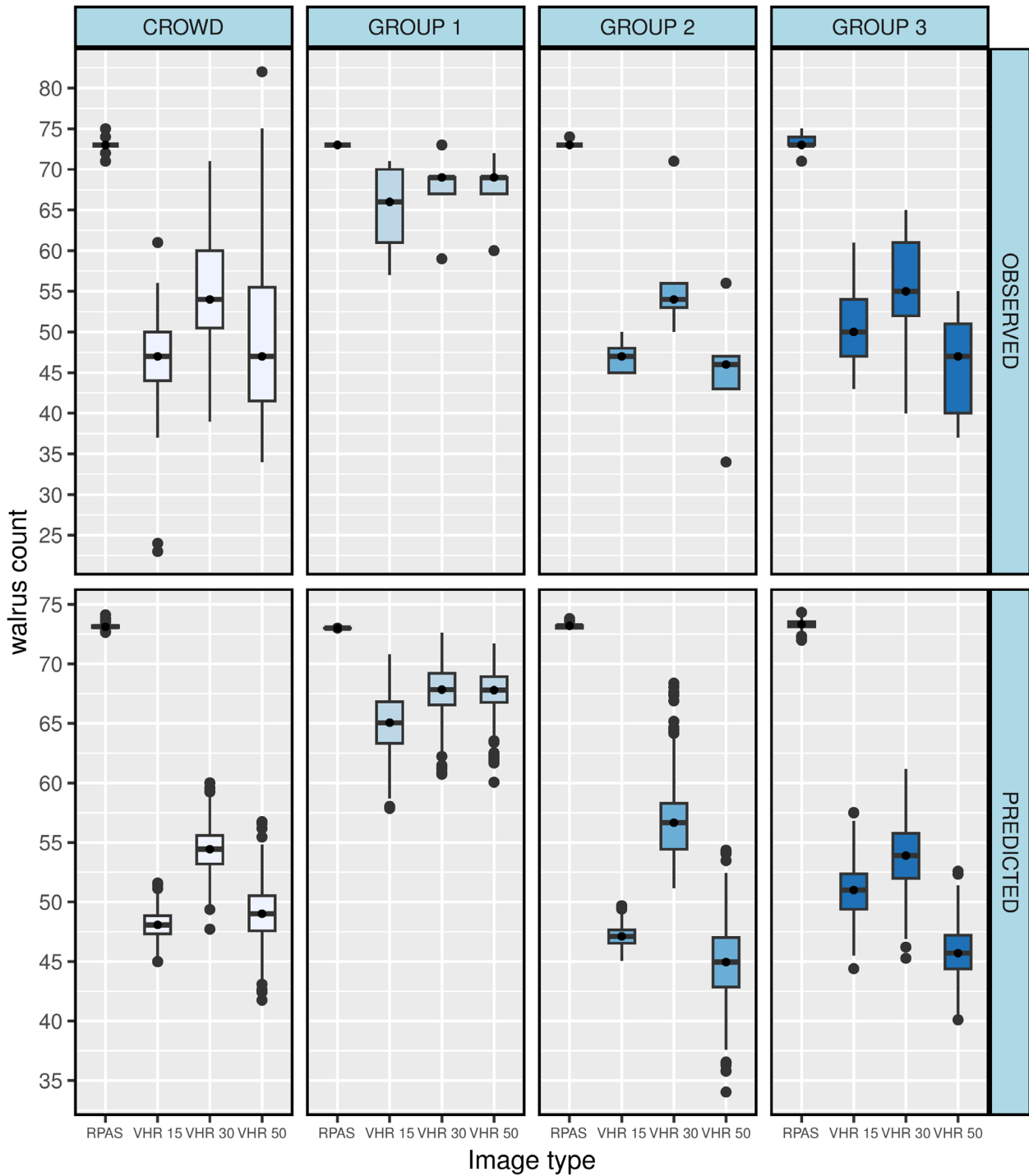
Figure 2. Boxplots of walrus counts by image spatial resolution and experience level of the assessors (Analysis 1) and by image spatial resolution and for the Crowd (Analysis 2). The top row provides observations, and the bottom row provides bootstrap predictions from log-linear models in Analyses 1 (GROUP 1–3) and 2 (CROWD). Boxes represent 50% of the central data, with limits defined by first (25%) and third (75%) quartiles. The solid lines are median values. Vertical lines are maximum and minimum values without outliers, which are depicted by solid points.

Figure 3. Georeferenced RPAS image with the point annotations (red dot) used to draw a convex hull bounding geometry (black dotted line) and buffered to 1.5 m on the outside to include all walruses (blue full line).

VHR satellites offer such raw spatial resolution, but Albedo is currently developing a constellation of VHR satellites that are designed to capture optical images with a 10 cm resolution. Launch of these satellites is planned for 2024 (Albedo, 2023).

All image assessor groups tested in this study under-counted walruses in the satellite images, across all spatial resolutions. This was a somewhat surprising result because the images used in this study represent ideal conditions for counting walruses individually. The animals were spread out and the background was light and not structurally complex, which allowed individual walruses to be distinguished easily. There were also not many animals in the images, reducing the risk of observer fatigue and subsequent errors in counting. Furthermore, light conditions were also ideal. Thus, our expectation is that undercounting will be even more pronounced when larger herds are assessed. Walruses are known to gather tightly in densities of 0.63–1.56 walruses/m^2 and in groups of hundreds of animals, or even 1000s or 10s of 1000s in some areas in the Pacific, covering areas of 100 000 m^2 (Battaile et al., 2017; Fischbach et al., 2016, 2022; Fischbach & Douglas, 2022).

The crowd was less accurate at counting walruses in VHR satellite imagery than experts, as would be expected, and their counts were more variable due to the higher number of counters and the more varied experience levels. Other factors may have played a role in the reduced accuracy, such as the potential for higher level of distraction for members of the public doing counts. However, the higher variation in the observed counts made by the crowd could be modelled, providing more precise mean predictions and a measure of uncertainty representative of citizen science projects. Thus, the crowd counts could be calibrated if some ground-truthed images are included in the model. For larger haul-out groups, a greater increase in bias and variation can be expected, lowering the accuracy. LaRue et al. (2020) also saw a reduced accuracy in the crowd counts, with the crowd over-counting animals by misidentifying rocks as Weddell seals.

The density estimated in this study is low compared to many reported densities for walrus haul-out groups (Battaile et al., 2017; Lydersen et al., 2012; Stewart et al., 2014). It is likely that group size, temperature, pre-dation risk, topography of the haul-out site, disturbance levels and other factors result in variable densities at haul-out sites. This is an important issue to address when assessing aerial imagery for abundance determination. Models should, where feasible, include a density measure for each group being counted, so that uncertainty can be estimated.

An alternative to citizen scientists' walrus counts in satellite imagery is automated detection. Several studies have been successful in detecting various types of wildlife in VHR satellite imagery (Borowicz et al., 2019; Bowler et al., 2020; Duporge et al., 2020; Green et al., 2023), but more work is needed to provide accurate counts. Convolutional neural networks (CNNs), a type of machine learning, is proving to be promising for detecting albatrosses (Bowler et al., 2020), whales (Borowicz et al., 2019; Green et al., 2023) and seals (Hoekendijk et al., 2021). So, it is likely going to be useful for walruses too. CNNs require large training datasets (Lecun et al., 2015), which are yet to be produced for walruses.

The results discussed herein are based on one satellite image of a walrus group hauled out on a sandy beach and with low density of walruses. Successful detection and counting of walruses are expected to be affected by the contrast between the walrus and the background environment, which may be influenced by a number of parameters, including geological substrate, light levels, haze and colour of the walruses. Therefore, we recommend repeated studies of haul-out groups with different conditions, including less ideal conditions – such as variable densities of walruses and various substrates (e.g., rocky shores, sandy beaches with large scattered boulders). Here we used RPAS imagery captured nearly simultaneously with the satellite imagery, which allowed us to ground truth estimates from the satellite imagery, but remote cameras installed at known haul-out sites at

heights that allow total counts of groups could work in a similar manner (Øren et al., 2018), although we anticipate that for larger and tighter haul-out groups, individual walruses will be difficult to distinguish. Therefore, applying the same method as used for aerial surveys, where the outline of the group is digitized and the density of portions of this tight group is estimated from satellite-captured images will be needed to provide counts that are accurate enough to be useful for management purposes (Fischbach et al., 2022). Additionally, future satellites are expected to have increasingly good resolution.

Conclusion

Walruses can be counted individually in VHR optical satellite imagery, at least when they gather on shores at relatively low densities. In this study, the highest currently available raw spatial resolution of 30 cm provided the best accuracy; therefore, we recommend that this resolution be selected for assessing abundance until the raw resolution improves for optical satellite imagery. In our study, citizen scientists provided less accurate counts, but the data created in this very cost-effective manner could be calibrated provided some appropriate ground-truth data. Our methods need to be tested further with larger walrus groups on variable substrates.

Author Contributions

HCC, PTF, KMK, CL, JF, RD conceived, planned and conducted the fieldwork; JF and HCC curated and analysed the data; HCC and JF drafted the manuscript. All authors provided critical feedback and helped shape the research, analysis and the manuscript.

Acknowledgements

Conflict of Interest Statement

The authors declare no conflicts of interest.

Data Availability Statement

All the counts per assessor and per image type are provided in a csv format (Data S2).

References

Airbus. (2022) Pleiades imagery user guide. Available from: https://www.intelligence-airbusds.com/en/8718-user-guides [Accessed 18th August 2023].

Albedo. (2023) Technical specifications. Available from: https://albedo.com/#capability [Accessed 18th August 2023].

Battaile, B.C., Jay, C.V., Udevitz, M.S. & Fischbach, A.S. (2017) Evaluation of a method using survey counts and tag data to estimate the number of Pacific walruses (*Odobenus rosmarus divergens*) using a coastal haulout in northwestern Alaska. *Polar Biology*, **40**, 1359–1369. Available from: https://doi.org/10.1007/s00300-016-2060-5

Boltunov, A., Evtushenko, H., Knijnikov, A., Puhova, M. & Semenova, V. (2012) *Space technology for the marine mammal research and conservation in the Arctic: results of the pilot project to develop methods of finding walruses on satellite images*. Murmansk: WWF-Russia.

Born, E.W., Gjertz, I. & Reeves, R.A. (1995) *Population assessment of Atlantic walrus (Odobenus rosmarus rosmarus L.)*. Oslo: Norsk Polarinstitutt.

Born, E.W., Wiig, Ø. & Tange Olsen, M. (2021) Chapter 12 – anthropogenic impacts on the Atlantic walrus. In: Keighley, X., Tange Olsen, M. & Jordan, P.D. (Eds.) *The Atlantic walrus: multidisciplinary insights into human-animal interactions*. London, UK: Academic Press, pp. 263–308.

Borowicz, A., Le, H., Humphries, G., Nehls, G., Höschle, C., Kosarev, V. et al. (2019) Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. *PLoS One*, **14**(10), e0212532. Available from: https://doi.org/10.1371/journal.pone.0212532

Bowler, E., Fretwell, P.T., French, G. & Mackiewicz, M. (2020) Using deep learning to count albatrosses from space:

assessing results in light of ground truth uncertainty. *Remote Sensing*, **12**(12), 2026. Available from: https://doi.org/10.3390/rs12122026

Burn, D.M. & Cody, M.B. (2005) Use of satellite imagery to estimate walrus abundance at Round Island, Alaska. *Proceedings of the 16th biennial conference on the biology of marine mammals*. San Diego, CA, USA, p. 46.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and inference: a practical information-theoretic approach*. New York: Springer.

Carpenter, J.R., Goldstein, H. & Rasbash, J. (2003) A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society: Series C (Apllied Statistics)*, **52**(4), 431–443. Available from: https://doi.org/10.1111/1467-9876.00415

Clarke, P.J., Cubaynes, H.C., Stockin, K.A., Olavarría, C., de Vos, A., Fretwell, P.T. et al. (2021) Cetacean strandings from space: challenges and opportunities of very high resolution satellites for the remote monitoring of cetacean mass strandings. *Frontiers in Marine Science*, **8**, 650735. Available from: https://doi.org/10.3389/fmars.2021.650735

Cubaynes, H.C., Clarke, P.J., Goetz, K.T., Aldrich, T., Fretwell, P.T., Leonard, K.E. et al. (2023) Annotating very high-resolution satellite imagery: a whale case study. *MethodsX*, **10**, 102040. Available from: https://doi.org/10.1016/j.mex.2023.102040

Cubaynes, H.C., Fretwell, P.T., Bamford, C., Gerrish, L. & Jackson, J.A. (2019) Whales from space: four mysticete species described using new VHR satellite imagery. *Marine Mammal Science*, **35**(2), 466–491. Available from: https://doi.org/10.1111/mms.12544

Davison, A.C., Hinkley, D.V. & Schechtman, E. (1986) Efficient bootstrap simulation. *Biometrika*, **73**(3), 555–566. Available from: https://doi.org/10.1093/biomet/73.3.555

Duporge, I., Isupova, O., Reece, S., Macdonald, D.W. & Wang, T. (2020) Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. *Remote Sensing in Ecology and Conservation*, **7**(3), 369–381. Available from: https://doi.org/10.1002/rse2.195

Dutta, A. & Zisserman, A. (2019) The VIA annotation software for images, audio and video. *Proceedings of the 27th ACM International Conference on Multimedia*. Nice, France, pp. 2276–2279.

ESA. (2015) Sentinel-2 user handbook. Available from: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook [Accessed 19th January 2024].

ESRI. (2023) *ArcGIS desktop: release 10.8*. Redlands, CA: Environmental System Research Institute.

Fay, F.H. (1982) Ecology and biology of the Pacific walrus, *Odobenus rosmarus divergens* Illiger. *North American Fauna*, **74**, 1–279.

Fischbach, A.S. & Douglas, D.C. (2021) Evaluation of satellite imagery for monitoring pacific walruses at a large coastal haulout. *Remote Sensing*, **13**(21), 4266. Available from: https://doi.org/10.3390/rs13214266

Fischbach, A.S. & Douglas, D.C. (2022) *Pacific walrus coastal haulout occurrences interpreted from satellite imagery* (ver. 2.0, December 2022). U.S. Geological Survey Data Release.

Fischbach, A.S., Kochnev, A.A., Garlich-Miller, J.L. & Jay, C.V. (2016) Pacific walrus coastal haulout database, 1852–2016 —background report. U.S. Geological Survey Open-File Report 2016-1108, p. 27.

Fischbach, A.S., Taylor, R.L. & Jay, C.V. (2022) Regional walrus abundance estimate in the United States Chukchi Sea in autumn. *The Journal of Wildlife Management*, **86**, e22256. Available from: https://doi.org/10.1002/jwmg.22256

Fretwell, P.T., LaRue, M.A., Morin, P., Kooyman, G.L., Wienecke, B., Ratcliffe, N. et al. (2012) An emperor penguin population estimate: the first global, synoptic survey of a species from space. *PLoS One*, **7**(4), e33751. Available from: https://doi.org/10.1371/journal.pone.0033751

Fretwell, P.T., Staniland, I.J. & Forcada, J. (2014) Whales from space: counting southern right whales by satellite. *PLoS One*, **9**(2), e88655. Available from: https://doi.org/10.1371/journal.pone.0088655

Green, K.M., Virdee, M.K., Cubaynes, H.C., Aviles-Rivero, A.I., Fretwell, P.T., Gray, P.C. et al. (2023) Gray whale detection in satellite imagery using deep learning. *Remote Sensing in Ecology and Conservation*, **9**, 829–840. Available from: https://doi.org/10.1002/rse2.352

Hammill, M.O., Mosnier, A., Gosselin, J.-F., Higdon, J.W., Stewart, D.B., Doniol-Valcroze, T. et al. (2016) Estimating abundance and total allowable removals for walrus in the Hudson Bay-Davis Strait and south and east Hudson Bay stocks during September 2014. DFO Canadian Science Advisory Secretariat Research Document 2016/036, p. v + 37.

Hoekendijk, J.P.A., Kellenberger, B., Aarts, G., Brasseur, S., Poiesz, S.S.H. & Tuia, D. (2021) Counting using deep learning regression gives value to ecological surveys. *Scientific Reports*, **11**, 23209. Available from: https://doi.org/10.1038/s41598-021-02387-9

Kovacs, K.M., Aars, J. & Lydersen, C. (2014) Walruses recovering after 60+ years of protection in Svalbard, Norway. *Polar Research*, **33**, 26034. Available from: https://doi.org/10.3402/polar.v33.26034

Laborie, J., Authier, M., Chaigne, A., Delord, K., Weimerskirch, H. & Guinet, C. (2023) Estimation of total population size of southern elephant seals (*Mirounga leonina*) on Kerguelen and Crozet Archipelagos using very high-resolution satellite imagery. *Frontiers in Marine Science*, **10**, 1149100. Available from: https://doi.org/10.3389/fmars.2023.1149100

LaRue, M.A., Ainley, D.G., Pennycook, J., Stamatiou, K., Salas, L., Nur, N. et al. (2020) Engaging "the crowd" in remote sensing to learn about habitat affinity of the Weddell seal in Antarctica. *Remote Sensing in Ecology and Conservation*, **6**(1), 70–78. Available from: https://doi.org/10.1002/rse2.124

Lecun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature*, **521**(7553), 436–444. Available from: https://doi.org/10.1038/nature14539

Lydersen, C. (2018) Walrus *Odobenus rosmarus*. In: Würsig, B., Thewissen, J.G.M. & Kovacs, K.M. (Eds.) *Encyclopedia of marine mammals*, 3rd edition. London, UK: Elsevier, pp. 1045–1048.

Lydersen, C., Aars, J. & Kovacs, K.M. (2008) Estimating the number of walruses in Svalbard from aerial surveys and behavioural data from satellite telemetry. *Arctic*, **61**(2), 119–128.

Lydersen, C., Chernook, V.I., Glazov, D.M., Trukhanova, I.S. & Kovacs, K.M. (2012) Aerial survey of Atlantic walruses (*Odobenus rosmarus rosmarus*) in the Pechora Sea, August 2011. *Polar Biology*, **35**, 1555–1562. Available from: https://doi.org/10.1007/s00300-012-1195-2

Lynch, H.J. & LaRue, M.A. (2014) First global census of the Adélie penguin. *The Auk*, **131**(4), 457–466. Available from: https://doi.org/10.1642/AUK-14-31.1

Matthews, C.J.D., Dispas, A. & Mosnier, A. (2022) Evaluating satellite imagery for Atlantic walrus (Odobenus rosmarus rosmarus) stock assessment-a pilot study. Canadian Technical Report of Fisheries and Aquatic Sciences 3492, p. v + 25.

Maxar Technologies. (2023) HD technology. Available from: https://explore.maxar.com/HD-Technology.html?utm_source=blog&utm_medium=organic&utm_campaign=hd-announcement [Accessed 19th January 2024].

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A. et. al. (2019) Polar regions. *IPCC special report on the ocean and cryosphere in a changing climate*. Cambridge, UK: Cambridge University Press, pp. 203–320. https://doi.org/10.1017/9781009157964.005

Norwegian Polar Institute. (2014) *Kartdata Svalbard 1:1000000 (S1000 Kartdata)*. Tromsø, Norway: Norwegian Polar Institute. Available from: https://doi.org/10.21334/npolar.2014.63730e2e

Norwegian Polar Institute. (2024) *Hvalross liggeplasser/Walrus haulout sites*. Tromsø, Norway: Norwegian Polar Institute. Available from: https://geodata.npolar.no/arcgis/rest/services/Temadata/F_Walrus_Haulout_Sites_Svalbard/MapServer [Accessed 18th August 2023].

O'Hara, R. & Kotze, J. (2010) Do not log-transform count data. *Nature Precedings*, **1**(2), 118–122. Available from: https://doi.org/10.1038/npre.2010.4136.1

Øren, K., Kovacs, K.M., Yoccoz, N.G. & Lydersen, C. (2018) Assessing site-use and sources of disturbance at walrus haul-outs using monitoring cameras. *Polar Biology*, **41**, 1737–1750. Available from: https://doi.org/10.1007/s00300-018-2313-6

Palomino-González, A., Kovacs, K.M., Lydersen, C., Ims, R.A. & Lowther, A.D. (2021) Drones and marine mammals in Svalbard, Norway. *Marine Mammal Science*, **37**, 1212–1229. Available from: https://doi.org/10.1111/mms.12802

Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-effects modeals in S and S-PLUS*. New York: Springer.

Pinheiro, J.C., Bates, D.M. & R Core Team. (2023) nlme: Linear and nonlinear mixed effects models. Available from: https://CRAN.R-project.org/package=nlme [Accessed 15th March 2023].

Planet. (2018) Skysat imagery product specification. Available from: https://assets.planet.com/marketing/PDF/SkySat-Ortho-Scene-Product-Spec-Sheet.pdf [Accessed 19th January 2024].

Planet. (2023) PlanetScope. Available from: https://developers.planet.com/docs/data/planetscope/ [Accessed 19th January 2024].

R Core Team. (2023) *R: a language and environment for statistical computing*. Vienna, Austria: Foundation for Statistical Computing. Available from: https://www.r-project.org/ [Accessed 15th March 2023].

Sherbo, B.A.H., Iacozza, J., O'Hara, A., Zhao, S.T., Ghazal, M. & Matthews, C.J.D. (2023) Validation of satellite imagery for quantitative population assessment and monitoring of Atlantic walrus. *Remote Sensing Applications: Society and Environment*, **32**, 101065. Available from: https://doi.org/10.1016/j.rsase.2023.101065

Stewart, R.E.A., Born, E.W., Dunn, J.B., Koski, W.R. & Ryan, A.K. (2014) Use of multiple methods to estimate walrus (*Odobenus rosmarus rosmarus*) abundance in the Penny-Strait-Lancaster sound and west Jones sound stocks, Canada. *NAMMCO Scientific Publications*, **9**, 95–122.

WWF. (2023) Walrus from space. Available from: https://www.wwf.org.uk/learn/walrus-from-space [Accessed 18th August 2023].

Yalcin, I., Kocaman, S., Saunier, S. & Albinet, C. (2021) Radiometric quality assessment for Maxar HD imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XXIV ISPRS Congress*, pp. 797–804. https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-797-2021

Zinglersen, K.B., Garde, E., Langley, K. & Mätzler, E. (2020) RemoteID: identification of Atlantic walrus at haul out sites in Greenland using high-resolution satellite images. Greenland Institute of Natural Resources in collaboration with Asiaq Greenland Survey. Technical Report 111. Greenland: Greenland Institute of Natural Resources, p. 62. Available from: https://natur.gl/ [Accessed 18th April 2023].

Supporting Information

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

Data S1. Protocol to count walruses in RPAS and VHR satellite imagery using VGG.
Data S2. Counts from all observers for the drone image and all three satellite images.
Data S3. Model selection.
Data S4. Drone image of walruses hauled out at Sarstangen, Norway, on 15th July 2022.