Monitoring large and complex wildlife aggregations with drones

Mitchell B Lyons*, Kate J Brandis, Nick J Murray, John H Wilshire, Justin A McCann, Richard T Kingsford & Corey T Callaghan

Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, UNSW Australia, Sydney Australia 2052

* corresponding author: mitchell.lyons@unsw.edu.au

Running title: Monitoring wildlife aggregations with drones

Key words: aerial vehicle, bird, breeding, colonial, waterbird, automated detection, machine learning, ecology

NOTE: This is a preprint, still under consideration for publication, and this version has not yet undergone peer review

Abstract

- Recent advances in drone technology have rapidly led to their use for monitoring and managing wildlife populations but a broad and generalised framework for their application to complex wildlife aggregations is still lacking
- We present a generalised semi-automated approach where machine learning can map targets of interest in drone imagery, supported by predictive modelling for estimating wildlife aggregation populations. We demonstrated this application on four large spatially complex breeding waterbird colonies on floodplains, ranging from ~20,000 to ~250,000 birds, providing estimates of bird nests
- Our mapping and modelling approach was applicable to all four colonies, without any modification, effectively dealing with variation in nest size, shape, colour and density and considerable background variation (vegetation, water, sand, soil etc.). Our semi-automated approach was between 3 to 8 times faster than manually counting nests from imagery at the same level of accuracy
- This approach is a significant improvement for monitoring large and complex aggregations of wildlife, offering an innovative solution for monitoring large and complex aggregations where ground counts are costly, difficult or not possible. Our framework requires minimal technical ability, is open-source (e.g., Google Earth Engine and R), and generalisable to other surveys

1 Introduction

Recent advances in technology offer the potential to improve field methods for rapidly and effectively monitoring biodiversity (Pimm *et al.* 2015). Among these advances is the use of aerial vehicles, or drones, that can carry remote sensing instruments to capture extremely high spatial resolution imagery with temporal flexibility (Anderson & Gaston 2013). Drones are relatively easy to use and their increasing 'off the shelf' application to wildlife research has been innovative and exciting (Chabot & Bird 2012; Chabot & Bird 2015). There are increasing novel applications for monitoring both populations and behaviours of different fauna, including birds (Chabot & Francis 2016; Hodgson *et al.* 2018), elephants (Vermeulen *et al.* 2013), crocodiles (Evans *et al.* 2016) and marine mammals (Seymour *et al.* 2017).

Given the ability of drones to collect high quality data near large aggregations of wildlife, they offer an attractive opportunity for improved methods and increased cost effectiveness of monitoring wildlife population. The relative advantages of aerial counting for wildlife monitoring is long established, including reduced detection error, increased precision, reduced observer effects and retrospective analysis of data and. For example, aerial counting was more accurate and precise than ground counting using aerial images of penguin colonies (Fraser *et al.* 1999) and geese (Boyd 2000). Similar advantages of image-based counts over ground-based counts have been demonstrated for drone acquired imagery too (Hodgson *et al.* 2018).

At large spatial scales (km) and for large aggregations (e.g. >5,000-10,000 individuals), aerial surveys provide cost effective information on counts of individuals, breeding-pairs and nests (Caughley 1977; Kingsford & Porter 2009), although sometimes suffering high variability and imprecision (Kingsford 1999). High altitude imagery from aeroplanes allows large areas, if not whole aggregations, to be captured in single images (e.g. in Boyd (2000) ~30 photos captured flocks of ~10,000 geese). Owing to the fact that similar areas require many thousands of drone images and to the extra complexity from increased spatial resolution, drone use for monitoring wildlife aggregations continues to be limited to monitoring relatively small aggregations (i.e. < 5-10,000 individuals), though there are some notable exceptions (Chabot & Bird 2012; Chabot, Craik & Bird 2015; Afán, Máñez & Díaz-Delgado 2018).

Manually counting targets of interest (e.g. individual animals, breeding-pairs, nests) from aerial images, regardless of capture platform, is laborious. This has driven the development of automated or semi-automated counting approaches (Chabot & Francis 2016; Hollings *et al.* 2018), aided by the widespread availability of increased computing power, growing computer literacy and new methods. Current approaches typically involve spectral thresholding (Chabot & Bird 2012; Seymour *et al.* 2017), point process algorithms (Descamps *et al.* 2011) or combinations of spectral properties and predictive modelling (Hodgson *et al.* 2018). These methods rely on high contrast (i.e. dark animals on light backgrounds or light animals on dark backgrounds) and consistency in the shape and colour of the targets (Hollings *et al.* 2018). They are generally only applicable when the spectral and structural characteristics of the animals (in the images) are unique compared to the rest of the image (Chabot & Francis 2016). More recently, remote sensing-based methods have been used to overcome challenges with low contrast and high variation among target objects (Groom *et al.* 2011; Drever *et al.* 2015; Liu, Chen & Wen 2015; Afán, Máñez & Díaz-Delgado 2018; Chabot, Dillon & Francis 2018).

Despite the interest in automated methods for counting aggregations of birds, their use by ecologists and managers for monitoring complex wildlife aggregations remains limited (Chabot & Francis 2016), with manual approaches still dominating (Buckland *et al.* 2012; Drever *et al.* 2015). There are three key reasons that have been highlighted for the disconnect between new methods and their ecological application: 1) most methods have only been demonstrated at small spatial scales relative to real-world applications (even if the number of individuals is very large) and in homogenous areas with little environmental complexity (Hollings *et al.* 2018); 2) ecological complexity and outcomes are not appropriately considered with respect to the mobility of individuals and variation in the types of target features of interest (Baxter & Hamilton 2018); and 3) there is a high technical threshold for implementing most methods (Chabot & Francis 2016).

In this paper, we develop a semi-automated framework for monitoring large complex wildlife aggregations using drone-acquired imagery. We use the case study of colonial waterbird breeding colonies because they present the key challenges currently inhibiting uptake of drone-based methods; the colonies cover large spatial extents and can have range of density of animals across these extents; there are many thousands of highly mobile individuals that cannot be contained to single drone images; the target features of interest are nests, which can exhibit significant differences in structure and colour across space and time (e.g. empty nests, adult/juvenile/chick/egg occupied nest, variable nest material, variable nest shape and arrangement); and considerable variation in background environment (mud, sand, water, live/dead vegetation). We developed a set of generalised methods, that transferred directly between colonies without modification, and required relatively little technical ability to apply. We captured imagery over four breeding waterbird colonies in New South Wales, Australia, ranging in size from ~20,000 to >200,000 birds, including the largest ever waterbird colonies to be surveyed by drone. We detail flight planning, image acquisition and processing, manual and automated methods for mapping and accurately counting nests. We include the Google Earth Engine and R code required for our analyses, along with a web-app to explore drone data, intermediate machine learning predictor and nest map layers.

2 Materials & methods

Our primary motivation was mapping and counting nests for breeding colonial waterbirds, with wide applicability. The methodology needed to work on both small (~10,000 – 20,000 birds) and large (200,000+ birds) colonies and be transferable across different environments and applications, requiring limited technical modification or ability. We developed a modular approach that included: 1) drone image surveys of four large breeding colonies; 2) manual counting of nests for training and validation; 3) a machine learning mapping method to map nests from drone imagery; and 4) a predictive modelling method to estimate total nest numbers.

2.1 Study location and bird colony details

Straw-necked Ibis (*Threskiornis spinicollis*) are an Australian nomadic waterbird species which form very large breeding colonies, sometimes mixed with other waterbird species, when ecological conditions are favourable (Brandis *et al.* 2014). We surveyed four colonies: *Merrimajeel, Zoo Paddock, Eulimbah and Block Bank* (Table 1). We surveyed the colonies at around their maximum size, determined by progression of breeding (Brandis *et al.* 2011). Straw-necked Ibis typically make their nests in flooded wetlands and floodplains, using inundated vegetation as nesting material raised above ground/water level. Their nests can be isolated nests or 'clumped' (10-200 nests). The nests are generally round or oblong in shape, but are often irregular in large clumps, with trampled vegetation, forming a dark green to brown colour, which increasingly whitens with guano (Fig. 1), until nests are abandoned either when offspring are lost or chicks fledge; at the latter stages, nests begin to lose structure and colour. At any point, nests may be empty, occupied by adults, eggs or juveniles, or a combination depending on parental foraging and care and chick mortality and fledging. A colony usually has a mixture of nest and juvenile ages. Most (>90 – 95%) of birds in the colonies were Straw-necked Ibis, a dark glossy blue-black bird on their back and wings, and with a white underside (black when viewed from above). There were also small numbers (<500-1000) of Australian White Ibis (*T. Molucca*), Glossy Ibis (*Plegadis falcinellus*) and Spoonbills (*Platalea spp.*) (Fig. 1).

2.2 Drone data and processing

Drone image data were collected during Spring and Summer of 2016 and 2017 (Table 1), using a DJI Phantom 3 Professional quad-copter, with the stock single sensor red/green/blue (RGB) camera. Colonies were within large flooded extents (km's wide), so multi-rotor drones were the only option, with no landing area for fixed-wing drones. We launched a drone from an amphibious vehicle or canoe used to enter the colonies. Flights were conducted using parallel flight lines, at ~100 m and speed of 5-10 ms^{-1} (Lyons *et al.* 2018a; Lyons *et al.* 2019). We aimed to acquire imagery with ~70% forward and lateral overlap to ensure adequate coverage for post-processing. Depending on weather and environmental conditions, we surveyed 5 – 40 hectares per flight, requiring multiple flights to survey each colony. There were no obvious negative interactions with the waterbirds; further animal ethics considerations can be found in Lyons *et al.* (2018a), and a more detailed protocol for drone-based monitoring of waterbird colonies in Lyons *et al.* (2019).

The drone imagery was processed using the commercial software Pix4DMapper (v4.19, Pix4D SA), which uses a photogrammetry technique called ‘structure from motion’ to identify points in overlapping images, ultimately generating a 3D point cloud reconstruction of the landscape. The 3D information is then used to generate a digital surface model and an orthorectified image mosaic. Only standard accuracy GPS (5 – 10 m accuracy) was used for georeferencing. This resulted in some error in absolute geographic location, but was not important, given our focus on identification and relative position of nests in the image mosaics.

2.3 Semi-automated approach for nest counting

Once the imagery was acquired, we needed to effectively identify nests which were highly variable in shape and colour, and sometimes had low contrast to the surrounding environment (Fig. 1). We initially tested a point process algorithm (Descamps *et al.* 2011) but it could not handle large data sizes; an object-based image analysis routine (*sensu* Chabot, Dillon and Francis (2018) but it had difficultly identifying >3,000-5,000 nests with one ruleset; and a machine learning/modelling approach (Hodgson *et al.* 2018) but it could not identify >1,000 nests with one parameterisation (see Data accessibility for modified Matlab routine). No particular technique worked effectively within or between the colonies, supporting similar findings on the limitations of automated and semi-automated methods (Hollings *et al.* 2018). So, we developed a modular approach, adaptable to variable target properties and scalable to large spatial extents, applicable to multiple colonies. This involved first mapping the area of nests using a remote sensing approach, and then estimating the number of nests using a predictive modelling approach.

2.3.1 Manual counts for training and validation data

A comprehensive training and validation data set was critical for developing counting methods. So, we first manually and systematically counted all the nests in the imagery over all colonies. We imposed a 50 x 50 m grid of quadrats on each colony, and digitally annotated every visible nest. We used this gridded method for two reasons: 1) it enabled an observer to sequentially work through the whole colony, while reducing distraction (and computer memory overhead) from surrounding areas; and 2) it reflected real-world practices when users choose only a limited number of training quadrats to manually count nests. During the field work, we also counted nests (in situ) for GPS-tagged clumps at each colony which we used to test the accuracy of the drone-based manual counting.

2.3.2 Machine learning mapping

Subsequently, we applied a supervised machine learning approach to map nests at each colony. We defined nests as any material or bird that constituted a nest or nest clump, based on our experience in the field. Motivated by its robustness to redundant predictor variables, we used a random forest classifier (Breiman 2001). Random forests are a machine learning algorithm that uses information from a training set and a suite of relevant predictor variables to predict class membership of all the image pixels in the study area. Random forests are particularly robust to redundant predictors, which is an important feature given that all data came from the one sensor. This allowed us to include many different image-based predictor variables without altering the approach for different colonies.

With all nests manually identified, we sampled a subset to train the random forest classifier. We randomly placed points across the colony, at least 30 m apart, and randomly chose a number of those points as a classifier training location. To approximate the 50 x 50 m quadrats, a 30 m buffer was placed around each chosen training locations, within which all manually counted nests were selected for training. We trialled between 5 and 20 training locations for each of the colonies. The classifier also requires non-target features (non-nest) randomly spread across the colonies: 1000 points for the smaller colonies (*Eulimbah* and *Block Bank*) and 10,000 points for the larger colonies (*Merrimajeel* and *Zoo Paddock*).

We derived arithmetic and textural metrics from the red, green and blue channels (r, g, b respectively) in the drone data to use as predictor variables in the random forest classification. These included: a 'white' index $\frac{b+g}{r}$; a Laplacian-8 edge-detection kernel on the 'white' metric; an RGB vegetation index $\frac{g-r}{g+r}$ (Bendig *et al.* 2015); a 'green brightness' index $\frac{g}{b+g+r}$; the 'contrast', 'variance', 'inverse difference moment' and 'shade' texture metrics from the Gray Level Co-occurrence Matrix (Haralick 1979), applied to each of the 'white' index and blue band; the standard deviation within a 2 m and 7 m radius of each pixel applied to the 'shade' metric and vegetation index; and a 1st and 2nd order difference of gaussians (Polakowski *et al.* 1997) on the 'shade' metric.

The training data set was compiled by extracting the pixel values for each image metric layer within a 10 cm buffer, around each training nest and non-nest point, so the random forest classifier was a binary nest and non-nest classification. The algorithm was parameterised with 500 trees and a minimum leaf population of 10. We implemented the classification in the Google Earth Engine (Gorelick *et al.* 2017), allowing seamless prototyping, visualisation and production environment for processing the large high resolution image data sets. Any contiguous areas less than 0.03 m^2 were removed (classification noise was unlikely to be bird nests) and exported from the Earth Engine. The Google Earth Engine is freely available to anyone, and we provide the code required to run the classifications, along with an interactive web-app to explore some drone data, predictor layers and nest classification interactively (link in Data accessibility section).

2.3.2 Predictive model estimation

To estimate the number of nests as a function of the mapped nest area for each colony, we used predictive modelling. We first summarised the number of manually counted nests and the mapped nest area within each 50 x 50 m quadrat. We then predicted the number of nests in each quadrat, with the whole colony count being the sum of the quadrat estimates. We used two simple approaches: 1) an assumption that the number of nests was directly proportional to the mapped nest area (linear area:count ratio); and 2) a generalised linear model (GLM; Poisson error distribution) of nest count as a function of nest area and local nest density. We expected that the local density of nests would have a relationship to the number of nests. Density was calculated as the percentage of the 50 x 50 m quadrat mapped as nests. Using a GLM with a negative binomial error distribution or a generalised additive model with smoothers for nest area and density provided no appreciable gains, so neither was pursued.

We used a resampling procedure to examine the number of manually counted 50 x 50 m quadrats needed to accurately estimate the number of nests for a whole colony. This involved repeated random sampling of n quadrats, estimating the number of nests using the area ratio and GLM approaches described above. We used 800 iterations without replacement (i.e. Monte Carlo resampling, not a bootstrap) for each of 1, 2, …, n_{max} quadrats. This resulted in a sampling distribution of 800 whole-colony nest count estimates at each n.

To simulate the scenario of limited resources for manual counting, we implemented another resampling approach to determine whether a given sample of the manually counted quadrats provided an accurate estimate (plus a confidence interval) of nest count for a whole colony. This involved a random draw of *n* quadrats (i.e. scenario of choosing a set of quadrats for training), and applying a repeated *k*-fold cross-validation using the area ratio and GLM estimation approaches. Each random draw of quadrats was stratified by mapped nest area density, to simulate choosing a range of nest density quadrats to count. We used $k = 10$ and 10 repeats for the cross-validation, and varied *n* from ~10-40% of the total number of manually counted quadrats. This resulted in a sampling distribution of 100 nest count estimates for each random draw of quadrats, and we took the mean as the resampling estimate and 2.5 and 97.5 percentiles as a 95% confidence interval. We decided on *k*-fold resampling as a good approach to reduce bias for the small sample sizes, but a range of resampling options are available (Lyons *et al.* 2018b). All statistical analysis was performed in R version 3.5.1 ((Team 2018); see Data accessibility section).

3 Results

3.1 Manual training and validation nest counts

The four study colonies varied widely in size, number of nests and bird density (Table 1). The flying height of ~100 m generated orthomosaic imagery with a pixel size between 3 – 4 cm. It took 5 – 15 minutes to manually count the nests in a 50 x 50 m quadrat, with higher nest density on the upper end of that time. Ibis nests and the flooded colony environment were so variably complex that it was often not possible to accurately manually count nests, even from 3 – 4 cm pixel drone imagery. Occasionally, artefacts from drone imagery processing also obstructed counting. The accuracy of the manual counting was estimated using the on-ground field counts, which ranged from ±6% to ±12% (Table 1). The smallest colony had a manual count of 7717 nests and the largest colony had 96,989 nests, and with an estimated population of over 200,000 birds at the time (Lyons *et al.* 2018a).

3.2 Semi-automated approach

The same Google Earth Engine code was applied to each colony, showing that the nest area mapping routine was robust to differing background environments and nest characteristics within and among each of the colonies. Around 10 of the 30 m training buffer locations were required for consistent classification of the large extent colonies (*Merrimajeel*, *Zoo Paddock*; ~5% total area), and around 5 for the smaller extent colonies (*Eulimbah, Block Bank*; ~10% total area). Our assessment of consistent was relatively ad hoc, using a visual assessment of whether nests and background were well separated. We left a quantitative assessment of accuracy for estimation of the total numbers. The chosen predictor variables did a good job at extracting the salient features of the bird colonies (Fig. 2) and the machine learning classification appeared to identify nests and birds appropriately (Fig. 3).

The first resampling routine demonstrated that considerable variation in nest estimates was likely given any random draw of quadrats, but only a small subset of the quadrats was required to capture most of the variation and provide estimates within the manual count error range (Supplementary Fig. 1). There was no noticeable improvement in using the GLM estimation method over the straight area ratio method. Comparing the results of the nest count estimates for individual quadrats showed that there was a large amount of variation among estimates for individual quadrats, the primary motivation for the use of a resampling-based estimate (Supplementary Fig. 2).

For the *k*-fold nest count estimation, we decided that an adequate number of quadrats (*n*) to use would be signified by most of the estimates from each *k*-fold cross-validation falling within the error margin of the manual nest counting (Fig. 4, Table 2). For the two largest colonies, *Merrimajeel* and *Zoo Paddock*, we used 30 quadrats (~12% of all 50 x 50 m quadrats) to provide accurate estimates. For the two smaller colonies, *Eulimbah* and *Block Bank*, we used 15 and 10 quadrats respectively (~20% and ~30% of total quadrats respectively). The manual effort time saving was best for the larger colonies – the nest counts were eight times faster for the two larger colonies (*Merrimajeel* and *Zoo Paddock*), but only five and three times faster for *Eulimbah* and *Block Bank* respectively (Table 2). The estimation was most accurate for the smaller two colonies, and there was some over-estimation for the larger colonies, particularly *Zoo Paddock* (Fig. 4), that could not be rectified with more training data. Again, there was no noticeable gain in using the GLM estimation method over the straight area ratio method; the gain from stratifying the random draw by mapped nest density was far more appreciable.

4 Discussion

We developed a generalised approach for monitoring complex wildlife aggregations, demonstrated through a semi-automated analysis providing estimates of numbers of nests in four large and complex waterbird colonies, using remotely sensed data captured via drones. The method was effective and provided accurate estimates at significant time savings compared to manual counts from the imagery. In our study, we obtained credible and useful estimates for one of Australia's more extensive breeding of colonial waterbirds. Our methodology is simple and robust enough to be applied in multiple environments, and works for both simple and complex target features. Continued development will see drone-based monitoring become integrated into waterbird monitoring (Lyons *et al.* 2019), and used to help quantify salient biological features like nesting success (Sarda-Palomera *et al.* 2017). There are potential benefits for monitoring some of the many other species of birds that form complex aggregations, as well as other animals such as marine animals in coral reef and rocky shore environments, migrating ungulates across different vegetation types or even unorthodox applications for counting spatially and spectrally complex target features such as coral bommies.

Our approach is modular, and the nest mapping and counting is implemented on free open source platforms, allowing users to readily change parameters or substitute their own or more appropriate methods. The semi-automated approach applied a machine learning classifier to high-resolution drone imagery to identify nests (Figs. 2 & 3), supported by modelling to estimate nest counts (Fig. 4). The methods were effectively applied across four different waterbird colonies, that contained highly variable target features on variable backgrounds. The colonies ranged in size from around 7000 nests to almost 100,000 nests (Table 1), and our semi-automated method required only relatively small amount of training data to produce comparable accuracy to manually counting from the drone imagery (Fig. 4, Table 2). Here we further discuss the cost-benefit aspects, opportunities for wider uptake, current challenges, and finish with some recommendations moving forward.

4.1 Cost-benefit of the semi-automated approach

The two key motivators for drone-based automated methods are reducing (on-ground) human observer bias and reducing cost (Chabot & Bird 2015; Baxter & Hamilton 2018; Hodgson *et al.* 2018; Hollings *et al.* 2018). For large and complex wildlife aggregations, such as our waterbird colonies, it is rarely possible to perform comprehensive on-ground counts and so drone-use provides an attractive option, and coupled with semi-automated methods, presents significant time savings too. In our case the large colonies were eight times quicker to survey (Table 2), representing a cost (time or money) saving of almost 90% compared to full manual counts. Cost-benefit analysis will continue to vary with user ability and conditions, including data acquisition in the field, drone image processing, modelling and programming, and even the level of detail and accuracy required for monitoring outcomes. Benefits will also vary with the nature of application with one-off monitoring perhaps better achieved using manual methods, until the technology and processes become routine. The efficiency increases in Table 2 are probably more likely to represent the potential time savings of further application of the methods. For new applications in novel environments, large manual counts are probably still required to understand the potential sources of error.

4.2 Opportunities for uptake of semi-automated methods

Transferability across environments and spatial scales prevents widespread adoption of semi- and fully- automated methods in wildlife monitoring (Chabot & Francis 2016; Hollings *et al.* 2018). We successfully implemented our semi-automated approach, using the same routine/code, on four different waterbird colonies, providing opportunities of transferability for other avian applications but also more broadly to a range of different large complex aggregations of wildlife. The key challenge we overcame was identification of target features with high spatial and spectral variation, on high variability backgrounds, across large spatial extents. Most current detection approaches rely on methods that require high consistency in the spatial and spectral organisation of target and background features. Our use of a random forest classifier efficiently handled redundant predictor data (Breiman 2001), allowing inclusion of many different colour, spatial and textural metrics as predictor layers. This helped capture more of the spatial and spectral variation in target features, compared to just using the image colour, as well as potential image blur and illumination artefacts (Fig. 3, top row).

Research is increasingly tackling this issue of consistency across target and background features, adapting methods from remote sensing mapping (Afán, Máñez & Díaz-Delgado 2018; Chabot, Dillon & Francis 2018). Application of a single consistent detection routine to many different applications will provide significant opportunity for broad uptake across scientific and management applications (Hollings *et al.* 2018). Although we found that existing methods (Descamps *et al.* 2011; Chabot, Dillon & Francis 2018; Hodgson *et al.* 2018) were not directly able to deal with the level of complexity in our case studies, our method ultimately required more training data, which reduces the overall cost saving. Thus continued development of a range of methods will provide opportunities for significant time and cost savings when applied over large spatial extents, over time.

Detection approaches from imagery, such as drone imagery, are increasingly benefiting from the remote sensing disciplines (Chabot, Dillon & Francis 2018), due to innovations in dealing with large volumes of data efficiently. Existing detection methods typically deal with image tiles in the order of 1-10 Mb. Our waterbird colonies involved 500 Mb to 5 Gb of data, requiring significantly improved data management and analysis. Use of the Google Earth Engine platform (or similar platforms) enables handling of large data, and will facilitate future expansion into web-based tools where users only supply imagery and training data, reducing local expertise and computing resource requirements.

We successfully identified both nests and individual birds when they were away from their nests (see Fig. 3, 3rd row). This demonstrates the opportunity to use our mapping driven approach to identify and count individual waterbirds. Indeed Chabot, Dillon and Francis (2018) used an object-based mapping approach for identifying and counting individual Snow Geese. If only individual birds were of interest, and they did not form complex spatial aggregations, the mapping process would be sufficient to identify and count individuals (i.e. the *k*-fold estimation process would be unnecessary). This represents an additional cost saving because it took less training data to train the machine learning mapping (e.g. for *Merrimajeel*, ~5% of the quadrats were needed to train the random forest, but 12% were needed to train the *k*-fold estimation). For small and simple tasks (e.g. counting just a few thousand birds or nests) our *k*-fold estimation process could also be replaced with simple thresholding or classification of the predictor metrics. For example, thresholding and vectorizing the predictor layers we used (e.g. Fig. 2, bottom row) produces accurate nest counts, but these thresholds become increasingly variable as spatial scale increases, making consistent application difficult.

4.3 Challenges for drone-based monitoring

Our main challenge was converting mapped nests to nest count estimates. Although comparable to manual counting, we were unable to rectify over estimation for the *Zoo Paddock* colony (Fig. 4, Table 2). This colony has a large spatial extent but was only sparsely populated, compared to the other large colony (*Merrimajeel*; Table 1). Improved modelling of density effects may reduce this problem. As it was, only five out of the 40 scenarios we ran would be considered a sizable overestimation (Fig. 4), and even then these numbers would be unlikely to affect management decisions, but this may vary depending on location and species of interest. We randomly selected quadrats, so a more judicious initial choice of quadrats for training may rectify this issue to some degree.

Another challenge is the potential impact of uncertainty (~6-12%, Table 1) in manual counting that can propagate through to the mapping and estimations. Moving semi-automated methods to increasing spatial scales or more complex environments requires dealing with more variation in image quality and limitations in the resolution able to be captured (Hollings *et al.* 2018). In our surveys, image resolution and quality was challenge, affected by our ability to access appropriate remote points for take-off and landing, along with environmental and ethics considerations that limited time available to collect imagery. This in turn resulting in varying incident sun angles and wind conditions during image collection, resulting in sun glint and image blur that sometimes obscured manual counting. Identifying old nests (e.g. Fig.1, top and bottom rows) was difficult, potentially further increasing manual counting errors. Our cross-validation approach was motivated by the need to account for uncertainty, and generally accounted well for this error (Fig. 4, Supplementary Fig. 1 & 2), but had a cost in terms of increased training data requirements.

Another challenge is the potential antipathy towards use of drones, when sometimes the literature present them in terms of taking over the role of surveyors. This is a fallacy because equally large amounts of human effort are needed in collection and processing of drone imagery, deriving the training and test data, and developing detection routines. Just as Fraser *et al.* (1999), almost two decades ago, demonstrated improved aerial counting from a kite-mounted camera, drones are now becoming part of the toolkit. Further, researchers and managers can be excited about access to fast and accurate counting, without adequately considering the potential uncertainty, labour and skills required for effective use of drones for monitoring large and complex wildlife aggregations, and that drones still cannot produce all the required biodiversity metrics for monitoring (Callaghan *et al.* 2018).

4.3 Recommendations

There are major improvements in data collection, interpretation and understanding which can come through using drone imagery, including cost savings and potentially improved accuracy. Applications will continue to grow, assisted by development of semi-automated methods such as ours. Drones should be viewed as a tool to complement ecological and environmental monitoring practitioners, rather than a replacement option. We suggest development of semi-automated approaches should focus on adaptability to deliver key monitoring indicators (Baxter & Hamilton 2018), and that detection methods themselves should aim for three main properties: 1) use predictor data that is easily derived from common drone-based (or airborne) imagery; 2) minimal parametrisation among environments, ensuring any parametrisation should be accessible to non-expert users; and 3) implementation on widely available platforms, not requiring significant local computing resources but able to manage large volumes of image data.

5 Data accessibility

Our nest mapping routines were implemented in the Google Earth Engine (https://earthengine.google.com/). All of the statistical analyses, including nest counting, were performed in the R programming environment (Team 2018). The Earth Engine and R code are available on Github (https://github.com/mitchest/bird-colony-count-drones) and archived on Zenodo (*eventual Zenodo DOI link*). Raw drone data cannot be released publicly for most of the colonies, because they are on private land, but the code provided includes the summarised data required for our analyses. We have developed a web-app through the Earth Engine (https://mitchest.users.earthengine.app/view/ibis-drone-count), using the public part of the *Eulimbah* colony, so users can explore drone data, predictor variables and nest map classification interactively.

Acknowledgments

Financial and logistical support grants (Australian Research Council LP150100972), the Commonwealth Environmental Water Office, the New South Wales Office of Environment and Heritage, Bush Heritage Australia and local land owners. We operated under two animal ethics approvals from the University of New South Wales Animal Care and Ethics Committee (16/3B and 16/131B). **Author statement:** All authors contributed to study design, ML, CC, JM and KB carried out field work, ML, JW and NM led the data processing and statistical analysis, and all authors wrote the manuscript.

References

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

Anderson, K. & Gaston, K.J. (2013) Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment,* **11,** 138-146.

Baxter, P.W. & Hamilton, G. (2018) Learning to fly: integrating spatial ecology with unmanned aerial vehicle surveys. *Ecosphere,* **9**.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M.L. & Bareth, G. (2015) Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation,* **39,** 79-87.

Boyd, W.S. (2000) A comparison of photo counts versus visual estimates for determining the size of snow goose flocks. *Journal of Field Ornithology,* **71,** 686-690.

Brandis, K., Koeltzow, N., Ryall, S. & Ramp, D. (2014) Assessing the use of camera traps to measure reproductive success in straw-necked ibis breeding colonies. *Australian Field Ornithology,* **31,** 99.

Brandis, K.J., Kingsford, R.T., Ren, S. & Ramp, D. (2011) Crisis Water Management and Ibis Breeding at Narran Lakes in Arid Australia. *Environmental Management,* **48,** 489-498.

Breiman, L. (2001) Random forests. *Machine Learning,* **45,** 5-32.

Buckland, S.T., Burt, M.L., Rexstad, E.A., Mellor, M., Williams, A.E. & Woodward, R. (2012) Aerial surveys of seabirds: the advent of digital methods. *Journal of Applied Ecology,* **49,** 960-967.

Callaghan, C.T., Brandis, K.J., Lyons, M.B., Ryall, S. & Kingsford, R.T. (2018) A comment on the limitations of uavs in wildlife research–The example of colonial nesting waterbirds. *Journal of Avian Biology*.

Caughley, G. (1977) Sampling in Aerial Survey. *Journal of Wildlife Management,* **41,** 605-615.

Chabot, D. & Bird, D.M. (2012) Evaluation of an Off-the-shelf Unmanned Aircraft System for Surveying Flocks of Geese. *Waterbirds,* **35,** 170-174.

Chabot, D. & Bird, D.M. (2015) Wildlife research and management methods in the 21st century: Where do unmanned aircraft fit in? *Journal of Unmanned Vehicle Systems,* **3,** 137-155.

Chabot, D., Craik, S.R. & Bird, D.M. (2015) Population Census of a Large Common Tern Colony with a Small Unmanned Aircraft. *Plos One,* **10**.

Chabot, D., Dillon, C. & Francis, C.M. (2018) An approach for using off-the-shelf object-based image analysis software to detect and count birds in large volumes of aerial imagery. *Avian Conservation and Ecology,* **13**.

Chabot, D. & Francis, C.M. (2016) Computer-automated bird detection and counts in high-resolution aerial images: a review. *Journal of Field Ornithology,* **87,** 343-359.

Descamps, S., Bechet, A., Descombes, X., Arnaud, A. & Zerubia, J. (2011) An automatic counter for aerial images of aggregations of large birds. *Bird Study,* **58,** 302-308.

Drever, M.C., Chabot, D., O'Hara, P.D., Thomas, J.D., Breault, A. & Millikin, R.L. (2015) Evaluation of an unmanned rotorcraft to monitor wintering waterbirds and coastal habitats in British Columbia, Canada. *Journal of Unmanned Vehicle Systems,* **3,** 256-267.

Evans, L.J., Jones, T.H., Pang, K.Y., Saimin, S. & Goossens, B. (2016) Spatial Ecology of Estuarine Crocodile (Crocodylus porosus) Nesting in a Fragmented Landscape. *Sensors,* **16**.

Fraser, W.R., Carlson, J.C., Duley, P.A., Holm, E.J. & Patterson, D.L. (1999) Using kite-based aerial photography for conducting Adelie penguin censuses in Antarctica. *Waterbirds,* **22,** 435-440.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment,* **202,** 18-27.

Groom, G., Petersen, I.K., Anderson, M.D. & Fox, A.D. (2011) Using object-based analysis of image data to count birds: mapping of Lesser Flamingos at Kamfers Dam, Northern Cape, South Africa. *International Journal of Remote Sensing,* **32,** 4611-4639.

Haralick, R.M. (1979) Statistical and Structural Approaches to Texture. *Proceedings of the Ieee,* **67,** 786-804.

Hodgson, J.C., Mott, R., Baylis, S.M., Pham, T.T., Wotherspoon, S., Kilpatrick, A.D., Segaran, R.R., Reid, I., Terauds, A. & Koh, L.P. (2018) Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution,* **9,** 1160-1167.

Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T. & Robinson, A. (2018) How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. *Methods in Ecology and Evolution,* **9,** 881-892.

Kingsford, R.T. (1999) Aerial survey of waterbirds on wetlands as a measure of river and floodplain health. *Freshwater Biology,* **41,** 425-438.

Kingsford, R.T. & Porter, J.L. (2009) Monitoring waterbird populations with aerial surveys - what have we learnt? *Wildlife Research,* **36,** 29-40.

Liu, C.C., Chen, Y.H. & Wen, H.L. (2015) Supporting the annual international black-faced spoonbill census with a low-cost unmanned aerial vehicle. *Ecological Informatics,* **30,** 170-178.

Lyons, M., Brandis, K., Callaghan, C., McCann, J., Mills, C., Ryall, S. & Kingsford, R. (2018a) Bird interactions with drones, from individuals to large colonies. *Australian Field Ornithology,* **35,** 51-56.

Lyons, M., Brandis, K., Wilshire, J., Murray, N., McCann, J., Kingsford, R. & Callaghan, C. (2019) A protocol for using drones to assist monitoring of large breeding bird colonies. *Australian Zoologist*, **submitted**.

Lyons, M.B., Keith, D.A., Phinn, S.R., Mason, T.J. & Elith, J. (2018b) A comparison of resampling methods for remote sensing classification and accuracy assessment. *Remote Sensing of Environment*, **208,** 145-153.

Pimm, S.L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R. & Loarie, S. (2015) Emerging Technologies to Conserve Biodiversity. *Trends in Ecology & Evolution*, **30,** 685-696.

Polakowski, W.E., Cournoyer, D.A., Rogers, S.K., DeSimio, M.P., Ruck, D.W., Hoffmeister, J.W. & Raines, R.A. (1997) Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency. *Ieee Transactions on Medical Imaging*, **16,** 811-819.

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

Seymour, A.C., Dale, J., Hammill, M., Halpin, P.N. & Johnston, D.W. (2017) Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery. *Scientific Reports*, **7**.

Team, R.C. (2018) R: A language and environment for statistical computing.

Vermeulen, C., Lejeune, P., Lisein, J., Sawadogo, P. & Bouche, P. (2013) Unmanned Aerial Survey of Elephants. *Plos One*, **8**.

Tables

Table 1. Location and information on drone-surveyed waterbird colonies. All colonies were in New South Wales, Australia. Nests were manually counted from the drone-based imagery. Nest count error was calculated from in situ ground counts cross-referenced with manual nest counts from drone imagery.

Location (*Colony name*)	**Date**	**Approx. colony size**	**Manual nest count**	**Manual nest count error**	**Estimated number of birds***
Lachlan River (*Merrimajeel*)	Oct 2016	60-65 Ha	96,989	±6.1%	200-250,000
Macquarie Marshes (*Zoo Paddock*)	Nov 2016	60-65 Ha	20,411	±8.8%	40-50,000
Murrumbidgee River (*Eulimbah*)	Nov 2016	15-20 Ha	13,343	±8.4%	30-40,000
Lachlan River (*Block Bank*)	Sep 2017	7-10 Ha	7717	±12.1%	15-20,000

*From (Lyons *et al.* 2018a) – the estimated number of birds incorporates site-specific information.

Table 2. Manual and semi-automated counting results for drone-surveyed waterbird colonies. Colonies were divided into a grid of quadrats and nests were manually counted with accuracy from in situ counts. *k*-fold nest estimates were derived from our semi-automated approach, using 40 different random subsets of quadrats.

Colony name	**50 x 50 m quadrats in grid**	**Manual nest count (± manual error)**	**Mean and range of *k*-fold nest estimates**	**Full count effort (hours)**	***k*-fold count effort (hours & speed-up)**
Merrimajeel	233	96,989 (91,073–102,905)	99,645 (90,383–106,727)	40	5 (8x)
Zoo Paddock	244	20,411 (18,615–22,207)	21,432 (16,627–27,361)	42	5 (8x)
Eulimbah	71	13,343 (12,222–14,464)	13,479 (12,212–14,879)	12	2.5 (5x)
Block Bank	33	7717 (6783–8651)	7777 (7152–8425)	5.5	2 (3x)

Figures

Figure 1. Example drone imagery showing the variation in nest types and environments across four breeding colonial waterbird colonies. Images from top row to bottom row are from the following colonies: *Merrimajeel, Zoo Paddock*, *Eulimbah* and *Block Bank* (details in Table 1).

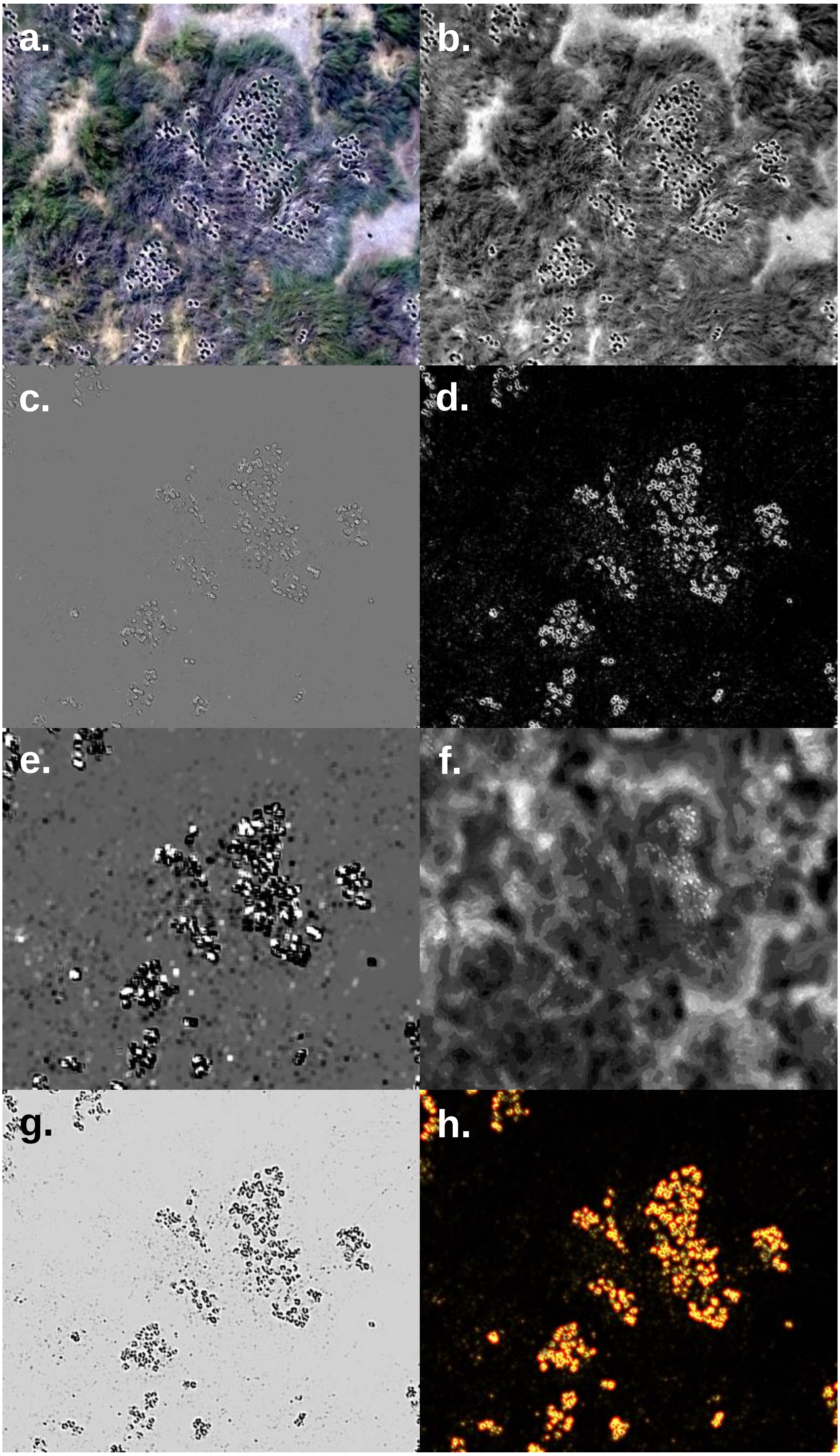

a.
b.
c.
d.
e.
f.
g.
h.

Figure 2. An example of image metrics derived from drone imagery over a waterbird colony (*Eulimbah*), used as predictor variables in the random forest classification: **a** – raw drone imagery; **b** – 'white' brightness image; **c & d** – GLCM 'shade' and 'contrast' of the 'white' metric; **e** – GLCM 'shade' of the blue reflectance; **f** – RGB vegetation index; **g** – difference of gaussians applied to the GLCM 'shade' on the 'white' metric; **h** – an RGB composite of the 'white' metric and the standard deviation within a 2m and 7 m radius for the GLCM 'shade' of the 'white' metric.

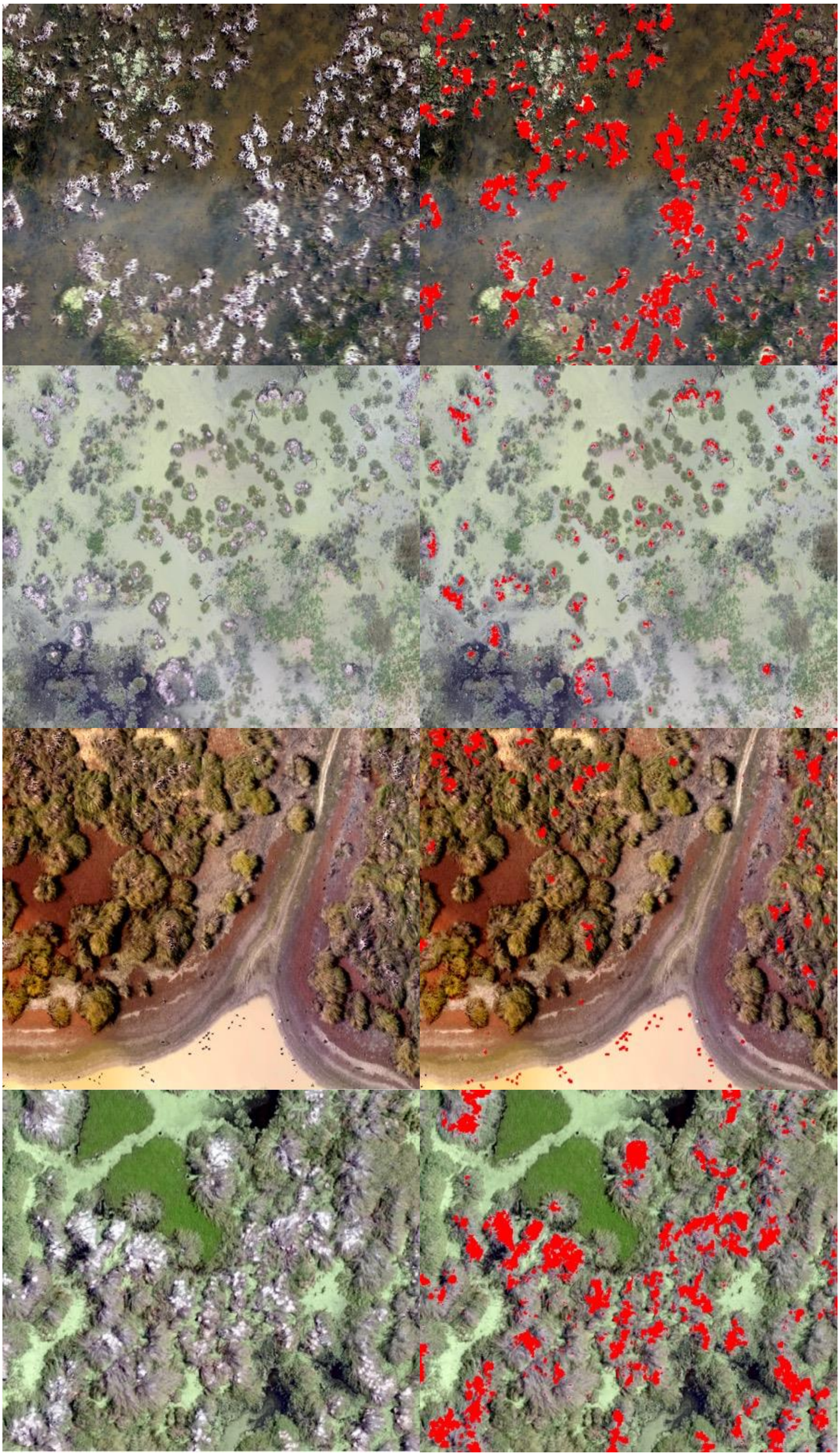


Figure 3. Example nest area classifications for four colonial waterbird colonies, surveyed via drone and classified using a random forest classifier in the Google Earth Engine. Images from top row to bottom row are from the following colonies: *Merrimajeel, Zoo Paddock*, *Eulimbah* and *Block Bank*. Full details in Table 1 & 2.

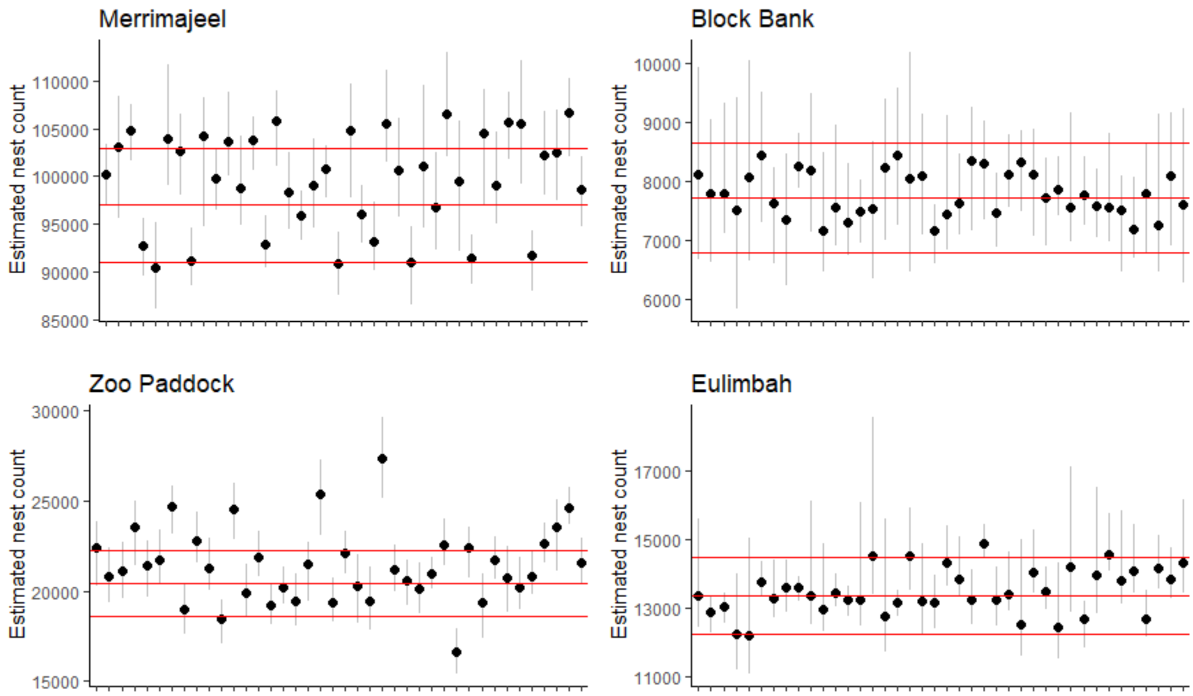
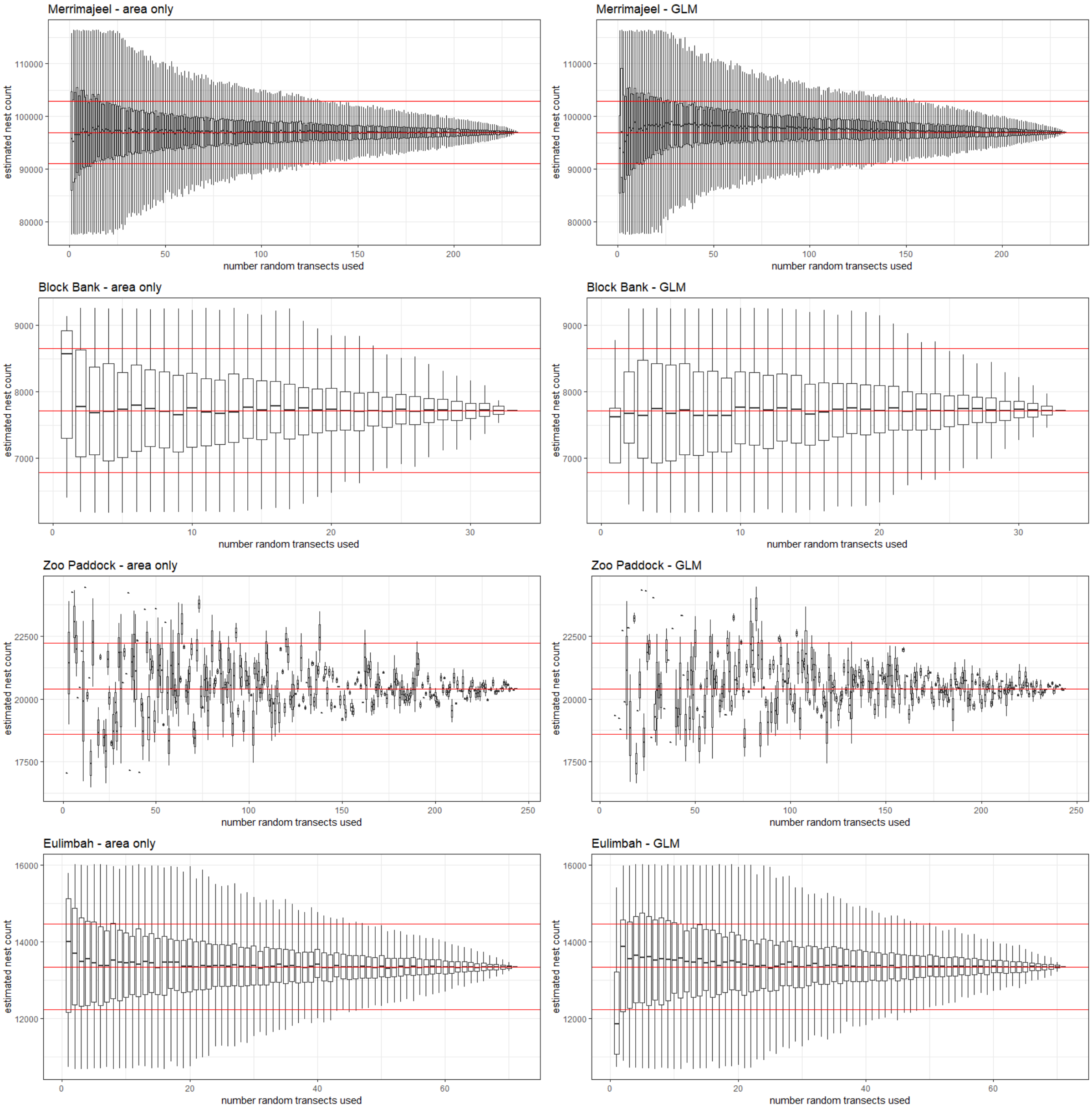
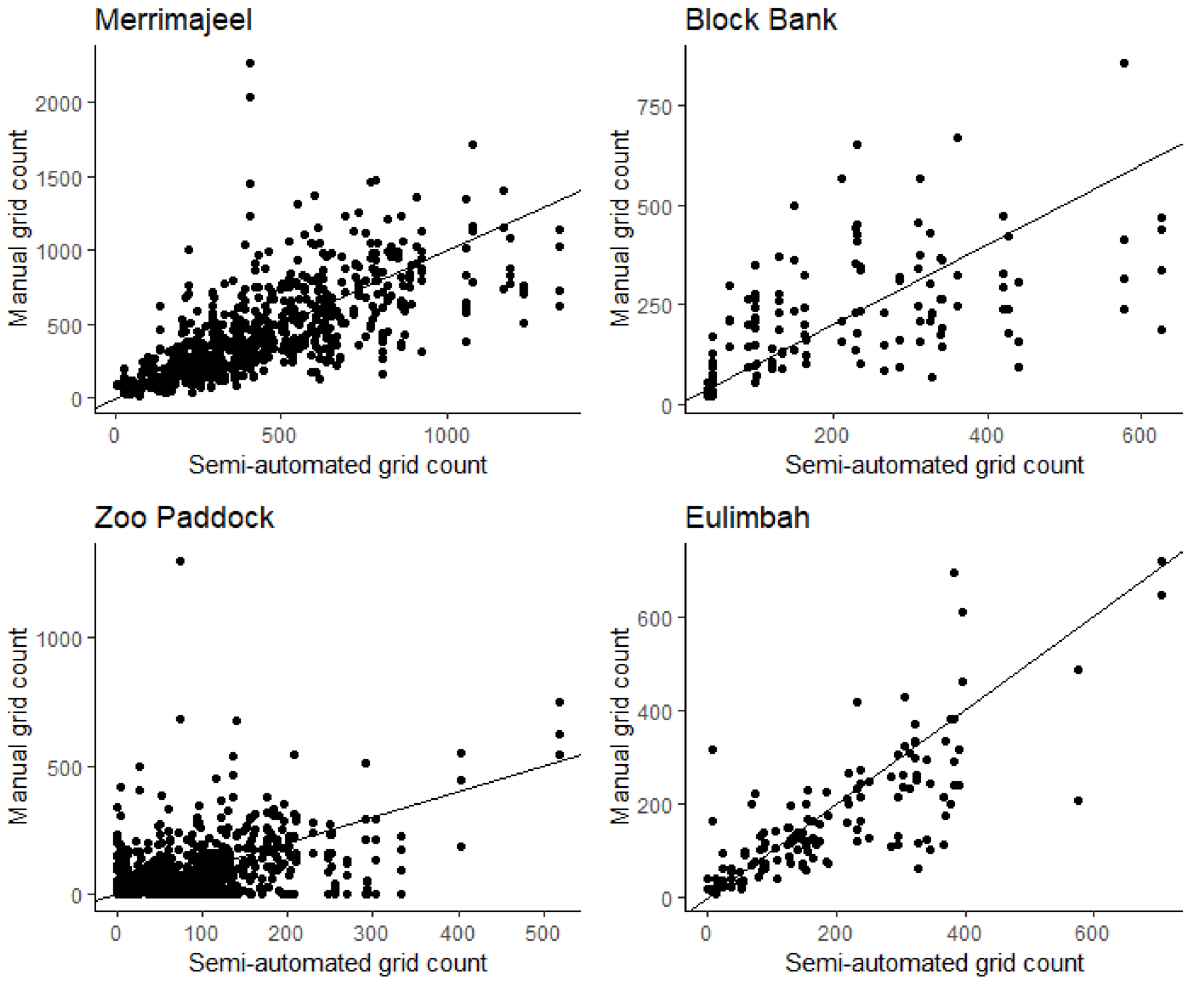


Figure 4. Resampling estimates of nest counts for four breeding waterbird colonies surveyed via drone, trained using a classification of nest area and manually counted nests. Each black dot represents the mean of the sampling distribution (10x repeated *k*-fold *k*=10 cross-validation) for a different subset of the manually counted training nests (corresponding lines denote 95% percentile), and the red horizontal lines denote the manual estimate for the whole colony, and the 95% error margin calculated from on-ground counts.

Supplementary Material



Supplementary Figure 1. Resampling estimates of nest counts for breeding waterbird colonies surveyed via drone, trained using a classification of nest area and manually counted nests (area ratio method on left and GLM method on right). Each box plot represents the sampling distribution (800x Monte Carlo cross-validation) for a different subset of the manually counted training nests, and the red horizontal lines denote the manual estimate for the whole colony, and the 95% error margin calculated from on-ground counts.



Supplementary Figure 2. Individual quadrat nest area estimates from a machine learning classifier plotted against the manual count from the corresponding quadrat, for four breeding waterbird colonies surveyed via drone.