Towards monitoring large and complex wildlife aggregations with drones

Mitchell Lyons*, Kate Brandis, Nick Murray, John Wilshire, Justin McCann, Richard Kingsford & Corey 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 (depending on title)**:** remote sensing, modelling, monitoring, aerial vehicle, breeding, colonial, waterbird, automated detection, machine learning

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 increased use for monitoring and managing wildlife populations. Despite demonstration of their relative advantages over traditional ecological methods, we have not seen a shift to their use more generally by ecologists and managers, particularly at large spatial scale and for environments with high biotic or abiotic complexity
- In this paper we present a generalisable semi-automated approach that first uses a machine learning approach to map targets of interest in drone imagery, and then a predictive modelling approach that estimates target counts. We use a case study of four large breeding waterbird colonies, ranging in size from ~20,000 to ~250,000 birds, where the ecological goal was to map and count bird nests. Google Earth Engine was used for the nest mapping, so users do not require significant local computing resources, and R for the nest count modelling.
- The approach we developed was able to be applied to all four colonies without any modification to the mapping or modelling routine, and was able to deal with large amounts of variation in nest size, shape, colour and density, as well as variation in background type and heterogeneity (vegetation, water, sand, soil etc.). Our approach estimated nest numbers at about the same accuracy as manual counting from the drone-imagery
- Our approach represents a significant improvement in cost-benefit for monitoring large and complex aggregations of wildlife, and a new potential solution for monitoring very large and complex aggregations where ground counts are virtually impossible. Importantly our approach also has a relatively low technical threshold for user's application to their own data. All code used for mapping and modelling nests is provided and we also provide an online web-app for users to explore the drone imagery and mapping predictor layers.

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 instrumentation (Anderson & Gaston 2013). The growing use of drones to capture image data has been driven by their relative ease of use, allowing people from a range of fields to capture extremely high spatial resolution data at a temporal resolution tailored to their needs. This has led to much excitement about the use of drones for wildlife monitoring (e.g.(Chabot & Bird 2015)), and novel applications for monitoring both populations and features of a range 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 their ability to collect high quality data in close proximity to large aggregations of wildlife, drones offer an attractive opportunity to improve methods for monitoring population and status. The relative advantages of aerial counting for these activities to complement ground surveys is long established, including reduced detection error, increased precision, ability to retrospectively analyse data and reduced observer effects. For example, Fraser *et al.* (1999) used aerial images of penguin colonies from a kite and Boyd (2000) used aerial (aeroplane) photography of geese flocks to show that aerial counting was more accurate and precise than ground counting. Recently, the same advantages over ground counts have been demonstrated with aerial photos captured with drones (Hodgson *et al.* 2018). However, application of drones in this context has generally been limited to either following sporadic individuals or, apart from a few exceptions (Chabot & Bird 2012; Chabot, Craik & Bird 2015; Afán, Máñez & Díaz-Delgado 2018), monitoring fairly small aggregations (i.e. < 5-10,000 individuals).

At large spatial scales (km's) and for large aggregations (e.g. >5,000-10,000 individuals), aerial surveys have long been seen as the only feasible methodology for providing information on parameters like counts of individuals, breeding-pairs and nests (Caughley 1977; Kingsford & Porter 2009). High altitude imagery from aeroplanes allows large areas, if not whole aggregations, to be captured in single images. (Boyd 2000) demonstrated counting of multiple large flocks of geese from aerial imagery, noting that ~30 photos captured flocks of ~10,000 geese. Using drones, large aggregations require many thousands of photos to provide coverage.

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).

Notwithstanding the promise of these applications, several major shortcomings have been reported, including in their ability to i) scale to cases or applications with large spatial extents, ii) deal with large numbers of highly mobile individuals or iii) account for complex environmental conditions (Baxter & Hamilton 2018; Hollings *et al.* 2018). A common example of when these three limitations are encountered is for large breeding bird colonies, which we use as case studies in this paper. Despite the interest in automated methods for counting aggregations of birds, there has not been a shift towards their use by ecologists and managers for monitoring bird colonies (Chabot & Francis 2016), and manual interpretation methods remains popular (Buckland *et al.* 2012; Drever *et al.* 2015).

Hollings *et al.* (2018) highlighted that two key reasons for this are that most methods have only been demonstrated 1) at small spatial scales relative to real-world applications, even if the number of individuals is very large, and 2) in homogenous environments that do not represent the complexity likely to be found in real world applications. Additionally, automated methods are often sought for larger colonies or groups, when the methods become cost-effective (Chabot & Francis 2016), and these larger groups usually present the complex conditions and image characteristics that inhibit generalisation of automated methods. These include: structural and spectral differences of birds and nests at various breeding stages (e.g. empty nests, adult/juvenile/chick/egg occupied nest, variable nest material, variable nest shape and arrangement); birds that do not stay stationary during nesting; differences in the structural and spectral properties of the background imagery (mud, sand, water, live/dead vegetation etc.); or variable population density across space.

A disconnect between the methods literature and ecological applications has also inhibited uptake (Chabot & Francis 2016). The direct barrier in terms of technical skills to implement and modify automated methods is only part of the issue. Automated methods almost exclusively focus on counting individuals, yet often, the primary biological parameter of interest is not individuals but the number or status of breeding-pairs or nests (Chabot, Craik & Bird 2015; Sarda-Palomera *et al.* 2017; Callaghan *et al.* 2018). Considerations of detectability and ecological context are a critical consideration during planning (Baxter & Hamilton 2018). Historically, individual counts have sometimes been used as a surrogate for these parameters, but with increased spatial resolution of drone-based imagery, these salient features like nest counts (Chabot, Craik & Bird 2015; Lyons *et al.* 2018a) and nesting success (Sarda-Palomera *et al.* 2017) can be directly quantified.

In this paper, we develop a framework for mapping and counting nests in large colonies of breeding birds from drone acquired imagery. We use the case study of colonial waterbird breeding colonies because they contain several of the key challenges currently inhibiting uptake of drone-based methods, including i) targets that vary in colour in shape, ii) spatially large and complex environments in including dry and inundated land and vegetation, and iii) many thousands of highly mobile animals that cannot be contained to single images. We developed a compartmentalised workflow that focuses on ecological outcomes that can be adapted to different methods and software platforms. We particularly focus on a methodology that is transferrable between target types and that has a relatively low technical ability threshold. We captured imagery over four breeding waterbird colonies in New South Wales, Australia, ranging in size from ~20,000 to >200,000 birds. Our framework was developed on some of the largest ever bird colonies to be surveyed by drone, and includes flight planning, image acquisition and processing, manual and automated methods for mapping and accurately counting nests. While this paper uses bird colonies as the case study, our methodology requires only minimal technical expertise, and should generalise to monitoring challenges for other large and complex biological phenomena.

2 Data & methods

The primary motivation in this study was mapping and counting nests for breeding colonial waterbirds. The key requirements for our methodology was that it must work on both small (~10,000 – 20,000 birds) and large (200,000+ birds) colonies and require no changes to the mapping and modelling implementations, and that there should be a low technical threshold for modifying the approach. We developed a modular approach that involved 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 nests numbers.

2.1 Study location and bird colony details

Straw-necked Ibis (*Threskiornis spinicollis*) are a nomadic bird species of Australia, and form very large breeding colonies when ecological conditions are favourable. We surveyed four colonies: *Merrimajeel, Zoo Paddock, Eulimbah and Block Bank* (location and size details in Table 1). Coordinating with on-ground ecologists, we attempted to survey the colonies at around their maximum size, to provide corresponding estimates of population. Straw-necked Ibis typically make their nests in flooded wetlands and flood plains, making use of inundated vegetation as nesting material raised above ground/water level. Their nests can be single isolated nests or in 'clumps' of up to 100-200 nests. The nests are generally round or oblong in shape, but are often irregular when in large clumps. The nests are typically made of trampled vegetation, so can be dark green to brown, and as they continue to be used they become white with guano, until eventually they are abandoned and lose any obvious structure or colour. Additionally, at any time point the nests may be empty, occupied by adults, eggs or juveniles, or some combination thereof. A colony usually has a mixture of nest and juvenile ages. The vast majority of birds in the colony were Straw-necked Ibis, which are a dark glossy blue-black colour on their back and wings, and have 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.*). Figure 1 shows some example imagery of birds and nests at the colonies.

2.2 Drone data and processing

Drone image data were collected during Spring and Summer of 2016 and 2017 (exact dates in Table 1). The data was collected using a DJI Phantom 3 Professional quad-copter, with the stock single sensor red/green/blue (RGB) camera. Colonies are typically located within large flooded extents, so multi-rotor drones are the only option, due to a lack of landing area for fixed-wing drones. An amphibious vehicle or canoe was used to enter the colonies, on which a platform can be set up to launch a drone. Flights were conducted using parallel flight lines at an altitude of ~100 m above take off and at a speed of 5-10 m/s (Lyons *et al.* 2019). We aimed to acquire imagery with ~70% forward and lateral overlap to ensure adequate photo overlap for post-processing. Depending on weather and environmental conditions, we were able to survey between 10-40 hectares per flight, so multiple flights were required to survey each of the colonies. Animal ethics and interaction 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). Pix4D uses a photogrammetry technique called 'structure from motion' which uses points identifiable in overlapping images to generate 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 results in some error in absolute geographic location, but this did not concern us since were only interested in identifying the relative position of features in the image mosaics.

2.3 Semi-automated approach

The objective for the bird colonies in our study was mapping and counting of nests. This was to support the broader ecological goal of delineating the spatial organisation of nests and nesting success in response to environmental drivers. While estimates of the population of individuals is of general interest, this number is more variable over short time periods compared to nest numbers. Additionally, ibis are quite mobile during nesting (e.g. at the Merrimajeel colony, there was usually >10,000 birds in the air), so counting individuals from drone-based imagery was not expected to produce accurate results (i.e. double-counting).

As described above, the nests are highly variable in shape and colour, and sometimes have low contrast to the surrounding environment (Fig. 1). We initially tested a point process algorithm ((Descamps *et al.* 2011); could not handle large data sizes), an object-based image analysis routine (*sensu* Chabot, Dillon and Francis (2018); difficult to identify >3,000-5,000 nests with one ruleset) and a machine learning/modelling approach ((Hodgson *et al.* 2018); could not identify >1,000 nests with one parameterisation – see Data accessibility for access to modified Matlab routine). Being unable to find generalised combinations of parameters that worked either within or between the colonies we surveyed was in line with recent findings on the limitations of automated and semi-automated methods (Hollings *et al.* 2018). We thus aimed to develop a modular approach that could adapt to variable target properties as well as be scaled to large spatial extents across 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 training and validation set

A comprehensive training and validation data set is critical for developing counting methods, we thus decided to manually count all the nests in the imagery over the colonies. We developed a systematic approach to compiling the manual nest counts, which involved splitting each colony into a grid of 50 x 50 m quadrats, and digitally annotating 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 allowed us to simulate real-world scenarios when users would choose only a limited number of training quadrats to manually count nests. During the field work, we counted nests (in situ) for GPS-tagged clumps at each colony, and we used these data to test the accuracy of the drone-based manual counting.

2.3.2 Nest mapping

We applied a supervised machine learning approach to map nest area at each colony. We defined nest area as any material or bird that constituted a nest or nest clump. 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 having to alter the approach for each colony.

Since all nests had been manually identified, we implemented a sampling method to choose a subset of nests to train the random forest classifier. We randomly placed points across the colony that were at least 30 m apart, and then randomly chose a number of those points to train the classifier. To approximate the 50 x 50 m quadrats, a 30 m buffer was placed around each of the chosen points, and all manually counted nests within the buffer were selected as training nests. The classifier also requires non-target features, so 'non-nest' points were randomly spread across the remaining colony; 1000 points for the smaller colonies (Eulimbah and Block Bank) and 10,000 points for the larger colonies (Merrimajeel and Zoo Padock).

We derived a number of 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), as it allowed a 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 considered 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 (see Data accessibility section).

2.3.2 Nest counting

To estimate the number of nests as a function of the mapped nest area for each colony, we used a predictive modelling framework. 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 is 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 were not pursued further.

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 having limited resources for manual counting, we implemented another resampling approach to determine whether a given sample of the manually counted quadrats can provide 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 between 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 are so variable that it was often not possible to accurately manually count nests, even in 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 counts, which ranged from ±6% to ±12% (Table 1). The largest colony had a manual count of 96,989 nests, and with an estimated population of over 200,000 birds at the time (Lyons *et al.* 2018a), regardless of the counting error, it is one of the biggest ever to be monitored via drone.

3.2 Semi-automated approach

The exact same Google Earth Engine code was able to be applied to each colony, showing that the nest area mapping routine was robust to the differing conditions and targets within and among each of the colonies. We found that around 10 of the 30 m training buffer areas 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 opted to leave a quantitative assessment of accuracy for the subsequent nest count estimation. Figure 2 shows some examples of the metrics used as predictors for the random forest classifier, and Figure 3 shows an example of the classified nest area for each colony.

The first resampling routine demonstrated that there was quite a large amount of variation in nest estimates given any random draw of quadrats, but that only a much smaller subset of the quadrats was required to capture most of the variation (Supplementary Fig. 1). There was no noticeable gain in using the GLM estimation method over the straight area ratio method. Inspecting the results of the nest count estimates for individual quadrats showed that there was a large amount of variation among estimates for individual quadrats, further motivating 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. For the two largest colonies, *Merrimajeel* and *Zoo Paddock*, we judged that to be 30 quadrats (~12% of all 50 x 50 m quadrats) to provide accurate estimates. For the two smaller colonies, *Eulimbah* and *Block*

Bank, we judged that to be 10 and 15 quadrats respectively (~20% and ~30% of total quadrats respectively). Figure 4 shows the *k*-fold nest count estimates (compared to manual count and error margin) for a set of 40 random draws for each of the colonies. The estimation was most accurate for the smaller two colonies. 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.

Parameterisation and thresholds for what is adequate will vary with the requirements for monitoring outcomes. The code provided allows users to fully customise the amount of training data for the random forest classifier and its internal parameterisation, the number of quadrats to sample for nest count estimation, the number of random draws to perform, the *k* for *k*-fold and the number of *k*-fold repeats (see Data accessibility section).

4 Discussion

The aim of this study was to develop a generalised approach that could map and count a key population metric – the number of nests – in four large and complex bird colonies using remotely sensed data captured via drones. An underlying driver was for the approach to be simple and robust enough to be applied in multiple environments, and modular such that users could readily change parameters or substitute their own or more appropriate methods as needed. We demonstrated a semi-automated approach founded on applying a machine learning classifier to high-resolution drone imagery to identify nests, and a modelling method to estimate a nest count. The methods were able to be applied exactly the same across all four colonies, and we found that only a relatively small amount of training data was required to estimate nest numbers with similar accuracy to manually counting from the drone imagery.

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 human-input time (Chabot & Bird 2015; Baxter & Hamilton 2018; Hodgson *et al.* 2018; Hollings *et al.* 2018). For very large and complex aggregations, like the bird colonies we surveyed, it is rarely possible to perform comprehensive on-ground counts. The primary motivator for automated methods is saving time. As we reported, out automated method provided significant time-savings. For example, it took ~40 hours to manually count all nests in the *Merrimajeel* colony, but ~5 hours to count enough nests to provide the same estimates using the *k*-fold cross-validation method (~8x faster). The time saving was less attractive for the smaller colonies; for example, the semi-automated approach was ~3.5x faster at *Block Bank*.

Explicitly quantifying cost-benefit is difficult, due to varying user ability and conditions across the whole exercise, including data acquisition in the field, drone image processing, modelling and programming, and even the level of detail and accuracy required for monitoring outcomes. While cost-benefit will improve over time, we concede that the realised benefits will vary between users. For example, if just a one-off monitoring exercise is required, then the most cost-effective option for users without prior experience, in terms of resources and delivery-time, is probably to just manually count features. Lowing the processing and skill overheads for applying automated methods is the critical factor for improving the cost-benefit ratio.

Advantages and challenges for increased uptake of automated methods

Transferability across environments and spatial scales has been identified as a key limitation preventing the uptake of automated methods within the broader scientific and management communities (Chabot & Francis 2016; Hollings *et al.* 2018). Using our semi-automated approach, the exact same routine/code was applied to each colony, which will facilitate easier transfer for other users on their own data. Large variation in target and background features, potentially combined with processing artefacts, results in high spatial variation in target features across the imagery. Most current detection approaches that are more algorithmic in how they treat colour and shape of target features naturally rely on consistency in their colour and shape, as well as consistency of the target background. The random forest classifier enabled efficient handling of redundant predictor data (Breiman 2001), which provides capacity to account for a range of spatial and spectral variation in target features, as well as potential image blur and illumination artefacts (Fig. 3, top row).

Recent research beginning to tackle this issue of consistency across target and background features has also used a remote sensing mapping mentality (e.g. (Afán, Máñez & Díaz-Delgado 2018; Chabot, Dillon & Francis 2018)). The ability to apply a single consistent detection routine to multiple cases will be a key factor for broader uptake across scientific and management applications (Hollings *et al.* 2018). It is important to note though that approaches like Descamps *et al.* (2011) and Hodgson *et al.* (2018) require less training data than the approach we developed here. Thus they represent more significant time savings if they can be further developed to work across larger spatial extents.

Similarly to that demonstrated in (Chabot, Dillon & Francis 2018), we think a mapping driven mentality represents increased ability to deal with large volumes of image data. Many existing methods deal with image tiles in the order of 1-10 Mb. The bird colonies we surveyed are in the order of 500 Mb to 5 Gb, so existing methods may require significant modification to perform efficiently, if at all. Use of the Google Earth Engine platform enables handling of large data, and will allow future expansion into web-based tools where users only supply imagery and training data, reducing local expertise and computing resource requirements.

Our approach was successful for identifying nest material as well as individual birds when they were captured in imagery away from their nests (see Fig. 3, 3rd row). This demonstrates that a mapping driven approach would also be suitable for identifying individuals if that was a main aim. Indeed (Chabot, Dillon & Francis 2018) successfully used an object-based mapping approach for identifying and counting individual Snow Geese. For small and simple tasks (e.g. counting just a few thousand birds or nests) the *k*-fold estimation process we used could be replaced with simple thresholding or classification of the predictor metrics. For example, thresholding and vectorizing the predictor layers in Figure 2 (e.g. bottom row) would produce very accurate counts, but the thresholds become too variable to apply consistently as spatial scale increases.

For the colonies that we surveyed, mapping the nest area required about a third of the training data needed to estimate the actual nest counts. This is important because for some monitoring applications, mapping the area or location of the phenomenon may be sufficient, which increases the cost-benefit saving. The main challenge for our approach was converting the mapped area to nest estimates. Although generally successful, we were unable to rectify over estimation for the *Zoo Paddock* colony. This colony was a very large, but only sparsely populated. Further exploration of modelling methods that account for density might be beneficial for situations such as this.

A potential limitation of our approach is that uncertainty in the manual counting stage can propagate through to the mapping and counting stages. One of the challenges when transferring automated methods to larger spatial scales or more complex environments is variation in image quality (Hollings *et al.* 2018). Indeed, our surveys were in flooded wetlands that had limited access points and take-off/landing was challenging. The remoteness and ethics requirements also dictated limited time in the colony itself. Together, this resulted in imagery being acquired across a wide range incident sun angles and during sub-optimal wind conditions. Resulting artefacts, like sun glint and image blur, occasionally inhibited manual counting. Older nests (e.g. Fig.1, top and bottom rows) are more difficult to identify, which attributed to manual counting errors up to ~12%. The cross-validation method we implemented does a good job at accounting for this error (Fig. 4, Supplementary Fig. 1 & 2), but also leads to the need for larger amounts of training data.

Managing expectations for drone-based monitoring

An unintended side effect of the intense focus on accurately counting individuals is backlash against the "drones vs. humans" attitude that a some of the literature presents. This idea is in fact a fallacy because drones don't typically count anything at all – large amounts of human effort goes into collecting and processing drone imagery, deriving the training and test data, and developing the detection routines. Almost two decades ago Fraser *et al.* (1999) showed that aerial counting from a kite-mounted camera was more accurate than ground surveys, but it's unlikely one would propose that kites count birds better than humans.

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) they use predictor data that is easily derived from common drone-based (or airborne) imagery, 2) they require minimal parametrisation among environments, and any parametrisation should be accessible to non-expert users, and 3) they are implementable on a platform that is widely available, does not require significant local computing resources, and can handle large volumes of image data.

Researchers and managers become excited about the proposition of fast and accurate counting, but invariably are disappointed when faced with the cold reality of using drones to monitor large and complex biological phenomena. Progress in the field is indeed rapid, but it is important that we continue to make progress without over-selling the capabilities of research- and consumer-grade drones. There is still much work to be done in terms of extracting complete end-to-end scientific and management applicable information from drone-based monitoring. Drones should be viewed as a tool to complement ecological and environmental monitoring practitioners, rather than a means to replace them.

5 Data accessibility

The 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*). Due to ecological sensitivity and being on private land, the raw drone data cannot be released publicly for most of the colonies, but the code provided includes the summarised data required to perform all the analyses in this paper. Part of the *Eulimbah* colony can be shown, and we have developed a web-app through the Earth Engine (https://mitchest.users.earthengine.app/view/ibis-drone-count) so that users can explore the drone data, the predictor variables and the 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.

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. & 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,* **in press**.

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 the surveyed bird colonies. All bird colonies were located within New South Wales, Australia. Nests were manually counted from the drone-based imagery. Ground-based nest count error is based on in situ counts cross-referenced with manual nest counts from drone imagery. *From (Lyons *et al.* 2018a) – the estimated number of birds incorporates site-specific information.

Location	**Date**	**50 x 50 m quadrats in grid**	**Manual nest count**	**Manual nest count error**	**Estimated number of birds***
Lachlan River (Merrimajeel)	Oct 2016	233	96,989	±6.1%	200-250,000
Macquarie Marshes (Zoo Paddock)	Nov 2016	244	20,411	±8.8%	40-50,000
Murrumbidgee River (Eulimbah)	Nov 2016	71	13,343	±8.4%	30-40,000
Lachlan River (Block Bank)	Sep 2017	33	7717	±12.1%	15-20,000

Figures

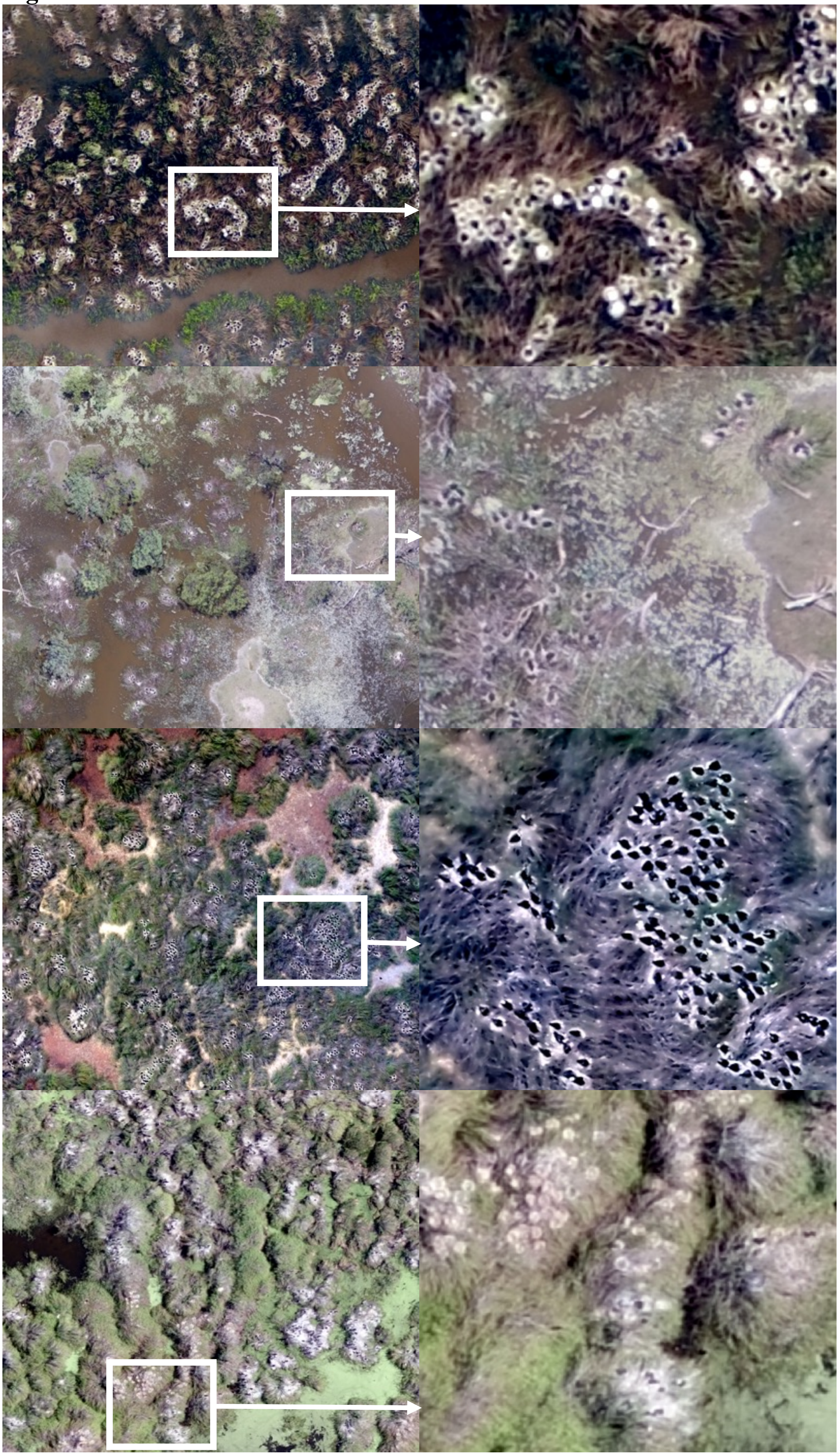


Figure 1. Example drone imagery showing the variation in nest types and environments across four breeding colonial waterbird colonies. Images from top row to bottom row are from the following colonies: *Merrimajeel, Zoo Paddock*, *Eulimbah* and *Block Bank*. Table 1 gives location and size details for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

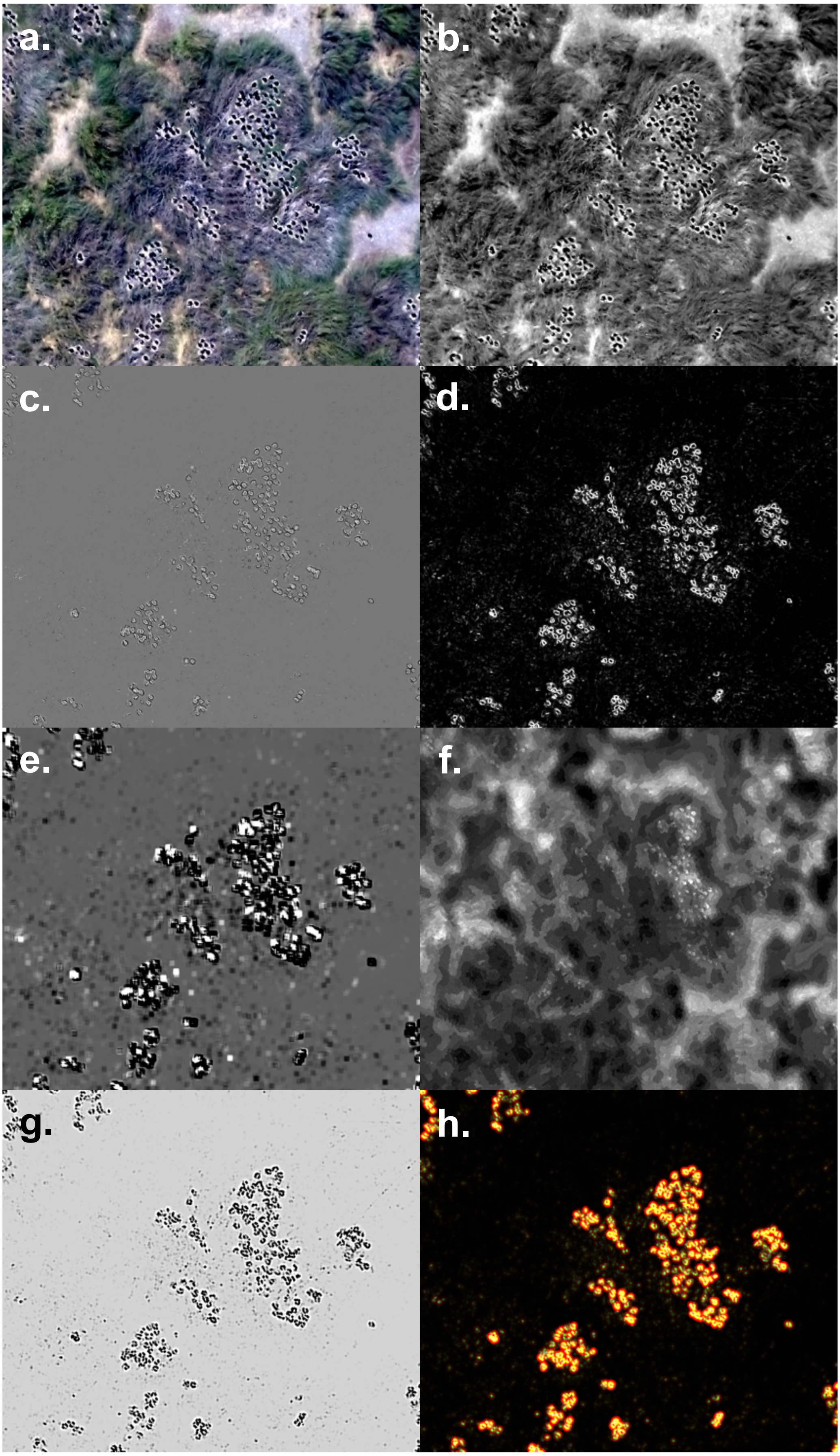

a.
b.
c.
d.
e.
f.
g.
h.

Figure 2. An example of images 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. See 'Data accessibility' section for access to an online web-app to explore these layers interactively.

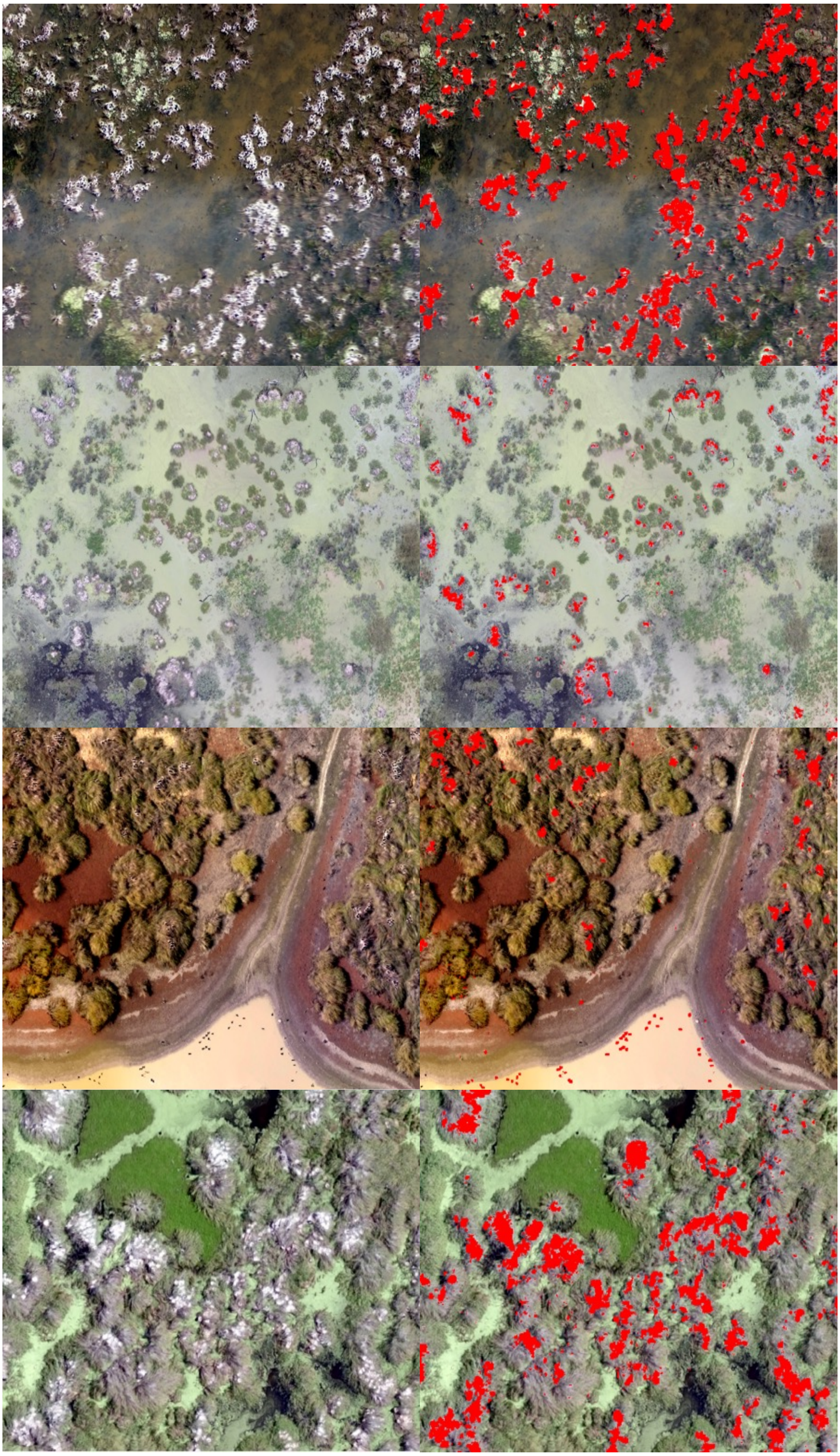


Figure 3. Example nest area classifications for colonial waterbird colonies surveyed via drone and classified using a random forest classifier in the Google Earth Engine. Images from top row to bottom row are from the following colonies: *Merrimajeel, Zoo Paddock*, *Eulimbah* and *Block Bank*. Table 1 gives location and size details for each of these colonies. See 'Data accessibility' section for access to an online web-app to explore the *Eulimbah* layer interactively.

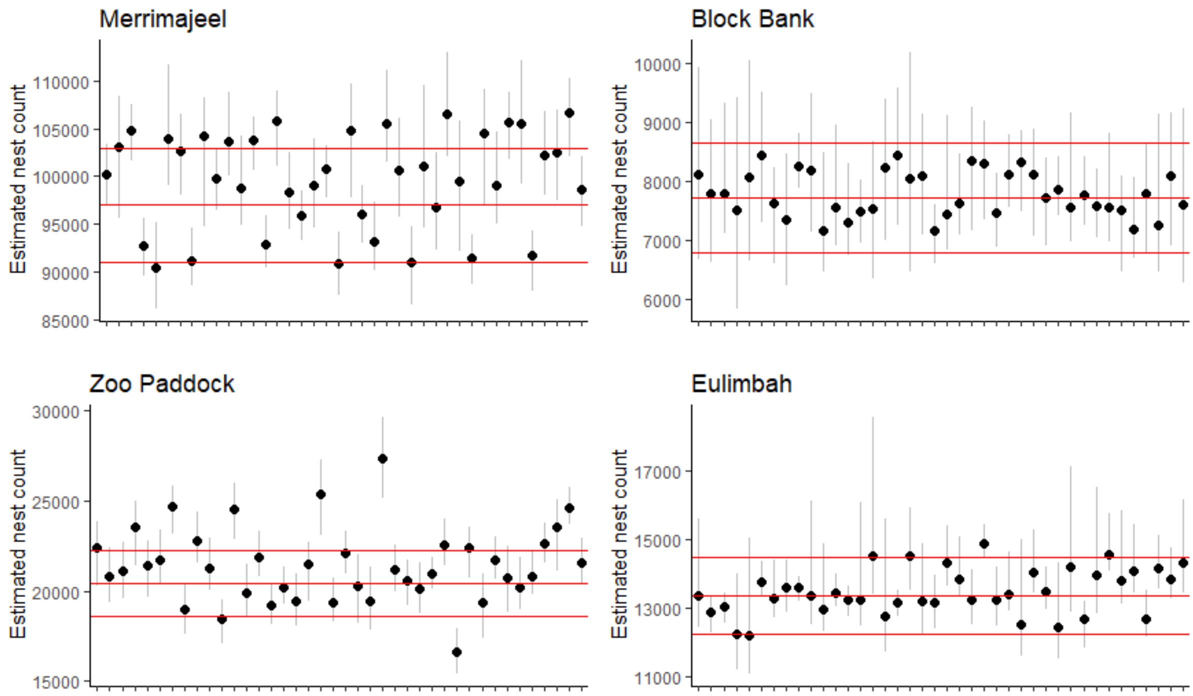
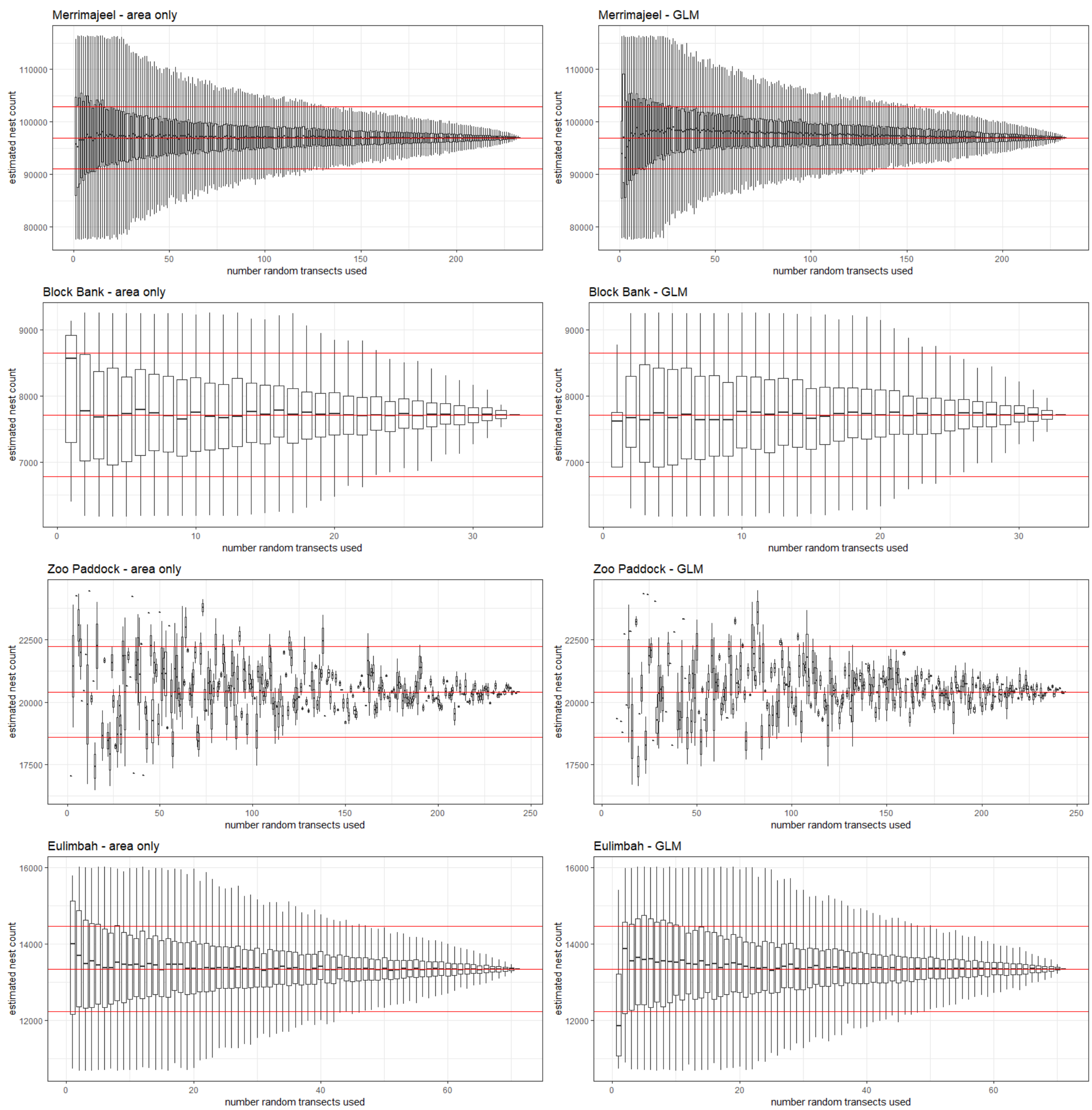
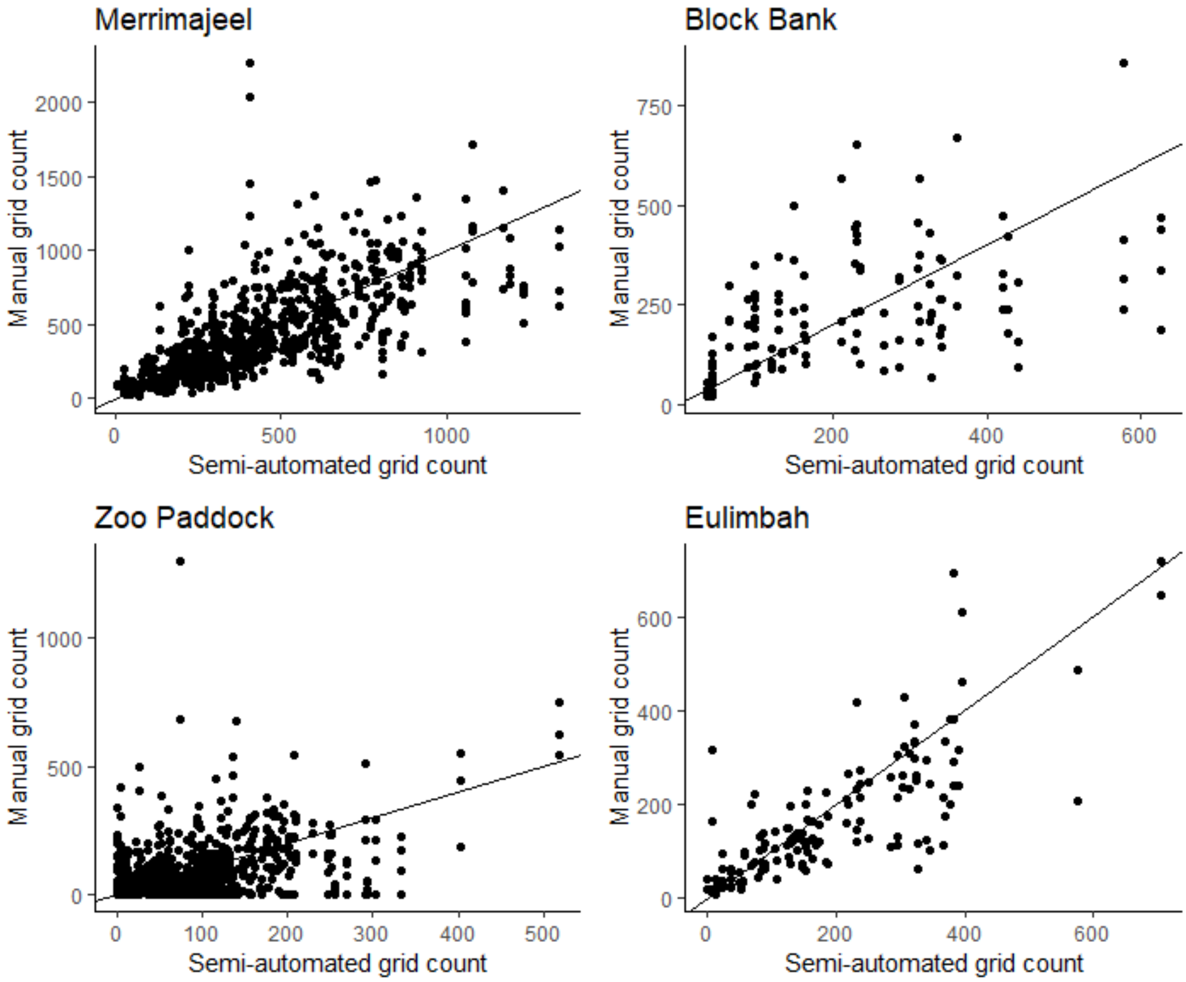


Figure 4. Resampling estimates of nest counts for breeding waterbird colonies surveyed via drone, trained using a classification of nest area and manually counted nests. Each black dot represents the mean of the sampling distribution (10x repeated *k*-fold *k*=10 cross-validation) for a different subset 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. For breeding waterbird colonies surveyed via drone, an example of individual quadrat area estimates from machine learning classifier plotted against the manual count form the corresponding quadrat.