ECOGRAPHY

Software Note

The CountEm software: simple, efficient and unbiased population size estimation

Javier González-Villa and Marcos Cruz

J. González-Villa and M. Cruz (https://orcid.org/0000-0002-4767-530X) ✉ (marcos.cruz@unican.es), Dept of Mathematics, Statistics and Computer Science, Univ. of Cantabria, Av. Los Castros 48, ES-39005 Santander, Spain.

Ecography
43: 251–255, 2020
doi: 10.1111/ecog.04862

Subject Editor: Michael Borregaard
Editor-in-Chief: Miguel Araújo
Accepted 13 October 2019

Population size estimation is essential in ecology and conservation studies. Aerial photography can facilitate this laborious task with high resolution images. However, in images with thousands of individuals exhaustive manual counting is tedious, slow and difficult to verify. Computer vision software may work under some particular conditions but they are generally biased and known to fail in several situations. The CountEm software is a simple alternative based on geometric sampling. It provides a fast and unbiased size estimation for all sorts of populations. The only requirement is that the discrete objects (e.g. animals) in the target population are unambiguously distinguishable for counting in a still image. Typical relative standard errors in the 5–10% range are obtained after counting ~200 properly sampled animals in about 5 min irrespective of population size. The CountEm ver. 1.4.1 is presented here, which includes a guided mode with a simple software interface.

Keywords: CountEm, geometric sampling, population monitoring, population size estimation, quadrats, wildlife management

Introduction

Population size estimation is fundamental to wildlife management, ecology or environmental science, but it is a challenging task in many cases. Traditional field methods can be expensive, biased and labor intensive, particularly in large and hard to access areas (Hollings et al. 2018). Alternative methods based on imagery have been gaining ground over the last decades (Hodgson et al. 2018, Lyons et al. 2019). Exhaustive manual counting on images (Kadlec and Drury 1968, Leonard and Fish 1974, Löffler and Margules 1980, Chabot et al. 2015) is the simplest among these methods. Some software packages have been developed to facilitate manual photo counts (Schneider et al. 2012, Gerum et al. 2017) but they are still time-consuming, observer dependent and difficult to verify for populations above a few thousands (Hollings et al. 2018). There have been major advances over the past three decades in population size estimation with automated computer vision software, such as Descamps et al. (2011) and references in Chabot and Francis (2016), Hollings et al. (2018). These methods can work in some particular cases with regular, or non-overlapping patterns on homogeneous backgrounds. However, they may perform poorly in more complex situations or in

NORDIC SOCIETY OIKOS

www.ecography.org

multispecies detection, and are generally biased to unknown degree (Chabot and Francis 2016, Hollings et al. 2018).

The CountEm ver. 1.4.1 software, which is freely available at <http://countem.unican.es> under a freeware license, is presented here. The CountEm method (Cruz et al. 2015, Cruz and González-Villa 2018, 2019) offers an unbiased population size estimation based on an alternative approach, namely geometric sampling (Howard and Reed 2005, Cruz-Orive 2017). Similar software has previously been developed for quantitative microscopy, see for instance Computer Assisted Stereological Toolbox (CAST) Grid stereology software (Olympus) or newCAST (Visiopharm), but CountEm is the first software that uses geometric sampling for macroscopic images. It can be applied to any kind of discrete objects or 'particles' of interest (e.g. birds, humans, trees, etc.), whereas automated computer vision software usually needs specific remodeling for different patterns.

The main idea is to estimate the population size, N, by properly sampling and counting ~200 particles irrespective of population size, yielding relative standard errors in the 5–10% range in ~5 min. Therefore, it is fast, accurate and reliable compared to the mentioned alternatives. The only practical requirement for applying CountEm is that all the particles in the population should be unambiguously identifiable for manual counting in the considered image. Systematic sampling is performed with a uniform random grid of quadrats and can be applied to populations of any size and spatial distribution. Thus, the approach is design based, see Cruz-Orive (2017), and it warrants unbiasedness (namely absence of systematic errors).

CountEm is written in Python 3.6.4. The user interface is implemented using the PyQt 4.11.4 bindings of the Qt framework (The Qt Company 2016). The Pillow (PIL fork) 5.1.0 and Numpy 1.11.3 libraries are used to manipulate images and matrices respectively.

Methods

The main idea of the CountEm method (Cruz et al. 2015) is to estimate population size, N, by performing geometric sampling on the population, Y, with a systematic sampling grid of quadrats of side length t, and separation between quadrat centers T (Fig. 1a). To ensure unbiasedness, the grid has to be superimposed uniformly at random on the target image. In this context a population is defined as a set of discrete objects which are visible on an image. The discrete objects (e.g. animals) are generally called particles. The grid is an infinite union of congruent quadrats. The number of particles captured by the grid, namely the sample size Q, is counted manually. The population size is estimated with the following unbiased estimator:

$$\hat{N} = \frac{T^2}{t^2} \times Q$$

The grid parameters $\{t, T\}$ have to be selected in order to obtain a sample size, $Q \sim 200$ and a number of non-empty quadrats $n > 30$, which usually yields relative standard errors in the 5–10% range (Cruz and González-Villa 2018). An alternative, more intuitive grid parametrization was proposed in Cruz and González-Villa (2018). The parameters considered are the sampling fraction $f = t^2/T^2$ and the initial number of quadrats $n_0 = B_x B_y / T^2$ where B_x, B_y represent image width and height, respectively. Selecting a suitable grid should be easier using the parameters $\{f, n_0\}$, because a low sample size (or a low number of non-empty quadrats) can be corrected by increasing f (or n_0).

A guided mode has been implemented in CountEm ver. 1.4.1. This mode guides the users through concrete, iterative steps starting from a rough visual estimation of population size $\hat{N}$. The software returns $\hat{N}$ and $\mathrm{sd}_{\mathrm{Cav}}(\hat{N})$, namely the predicted standard error of $\hat{N}$, calculated using the new modification of the Cavalieri estimator, as explained in Gomez et al. (2019). The predicted coefficient of error of $\hat{N}$, $\mathrm{ce}_{\mathrm{Cav}}(\hat{N}) = \mathrm{sd}_{\mathrm{Cav}}(\hat{N}) / \hat{N}$, is also given.

The precision of the method was tested with Monte Carlo replications on 51 human crowd images (Cruz and González-Villa 2018). The relative standard errors were typically in the 5–10% range.

Software interface

The first step in CountEm ver. 1.4.1 is mode selection. The standard mode allows the experienced user to choose the grid parameters at will, whereas in the guided mode the recommended buttons in each step are marked in yellow, while some of the options of the standard mode are hidden in order to visually simplify the window.

The 'inputs' window (Fig. 2a) allows to load the image and choose the sampling parameters defining the grid. In the guided mode, the software will automatically choose a suitably sized sampling grid, based on a visual estimation of the total number of visible particles in the image, $\tilde{N}$. The visual estimation does not need to be accurate since the guided protocol allows to adjust the parameters on several iterations if necessary. A rough estimate obtained in a few seconds is sufficient. The software sets the sampling fraction to $f = 200 / \tilde{N}$, and the number of initial quadrats to $n_0 = 100$. Manual choice of grid parameters is available in the standard mode.

The 'select area' button allows the user to create a polygonal region of interest (RoI) by interactively clicking on the screen display. Multiple RoI's can be defined iteratively. The 'process' button, generates a uniform random superimposition of the grid of quadrats on the image. The quadrats lying outside the selected RoI's are not shown, reducing the counting time. The actual area of the selected region is not used in the calculations, however. Hence the borders of the RoI do not need to be drawn accurately.

There are two alternative modes in the 'Counting' window of the Software interface, namely 'Overview' and 'Image view' (Fig. 2b). The 'Overview' shows the full image with the grid

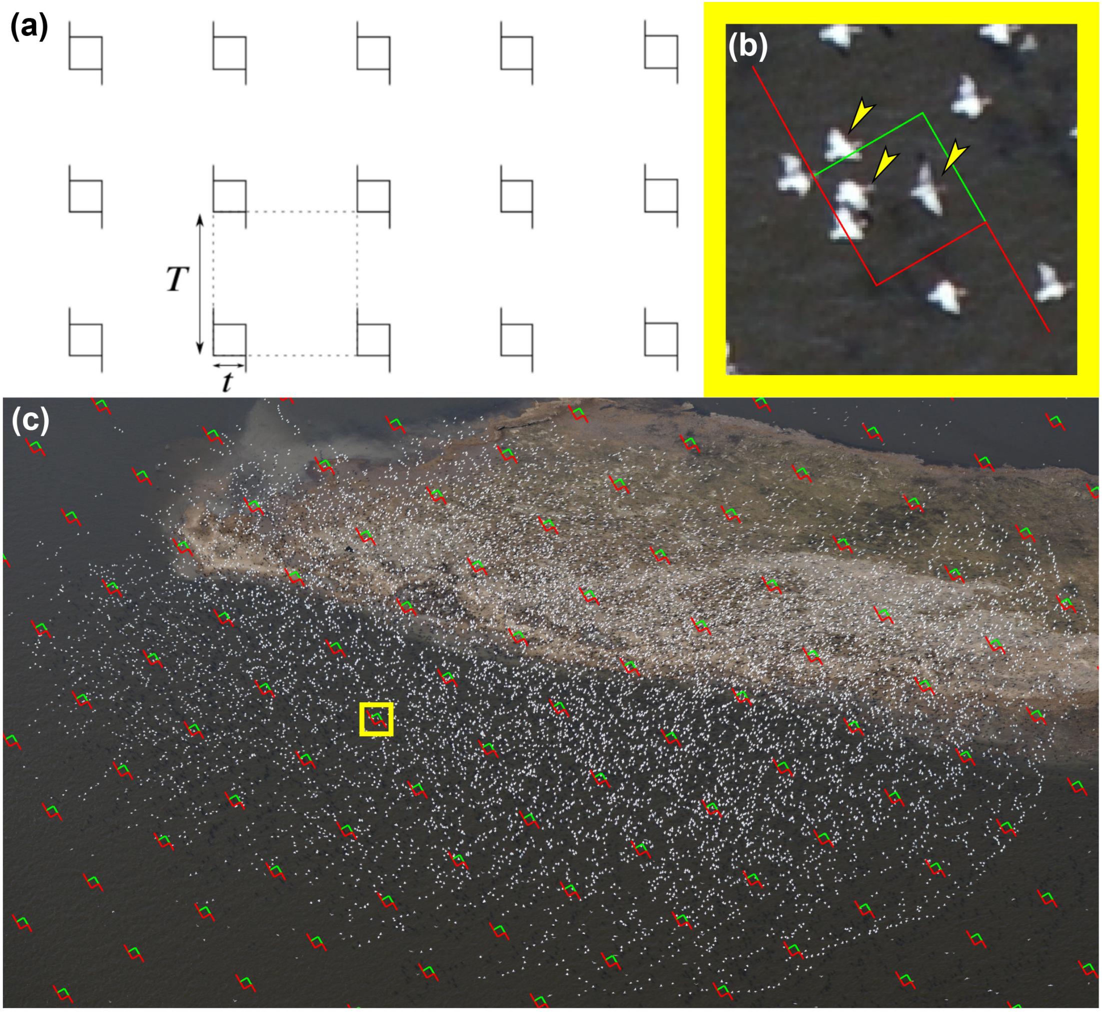

Figure 1. (a) A portion of the grid of quadrats used for systematic sampling. The sampling fraction is t^2/T^2. (b) Magnified version of the quadrat marked in (c). Only the three marked birds should be counted, applying the forbidden line counting rule. (c) Image of the GSGO survey (Francis St-Pierre, Canadian Wildlife Service) with $N = 13\,744$ manually counted snow geese. A grid of quadrats of the type shown in (a), with $T = 440$ and $t = 44$ pixels ($n_0 = 100$, $f = 0.01$), has been superimposed with a tilt of 60°.

as in Fig. 1c and can be used to assess whether the grid is convenient. If sample size Q or number of non-empty quadrats, n are low (or high), the grid parameters should be changed. This adjustment can be done manually after clicking the 'Back' button, or using the 'Estimated number of non-empty quadrats' and 'Estimated sample size' fields, and the 'Check parameters' button. The latter, is the only available option in the guided mode. The 'Check parameters' button evaluates if the grid parameters are suitable. If necessary, new parameter values, f', n' are determined to generate a new grid and repeat counting. A magnified quadrat is shown in the 'Image view'. Users have to manually count the number of particles in the zoomed quadrat applying the forbidden line rule (Gundersen 1977) to cope with edge effects: a particle is counted in a quadrat only if it touches the quadrat but does not hit the extended, forbidden line of the quadrat (in red in Fig. 1b). The result should be written in the 'Count' cell. Sizes of different types of populations (e.g. different bird species) can be estimated simultaneously by adding the desired number of fields. The 'Next' and 'Previous' buttons can be used to switch to the preceding and the following quadrat respectively. After counting the particles in the last quadrat, the progress bar (top left) turns green and 'Continue' can be clicked.

The resulting values of $\hat{N}, Q, n, \mathrm{sd}_{\mathrm{Cav}}(\hat{N}), \mathrm{ce}_{\mathrm{Cav}}(\hat{N})$ for each object are displayed in the 'Results' window (Fig. 2c). The numerical input and output values, and the overview and/or individual quadrat images can be exported to csv and tiff files respectively.

Example

The greater snow goose (GSGO) image shown in Fig. 1c is analyzed here. It was taken in eastern Canada using fixed-wing aircraft, for the GSGO Spring Survey, to monitor the greater snow goose population in southern Québec and Ontario.

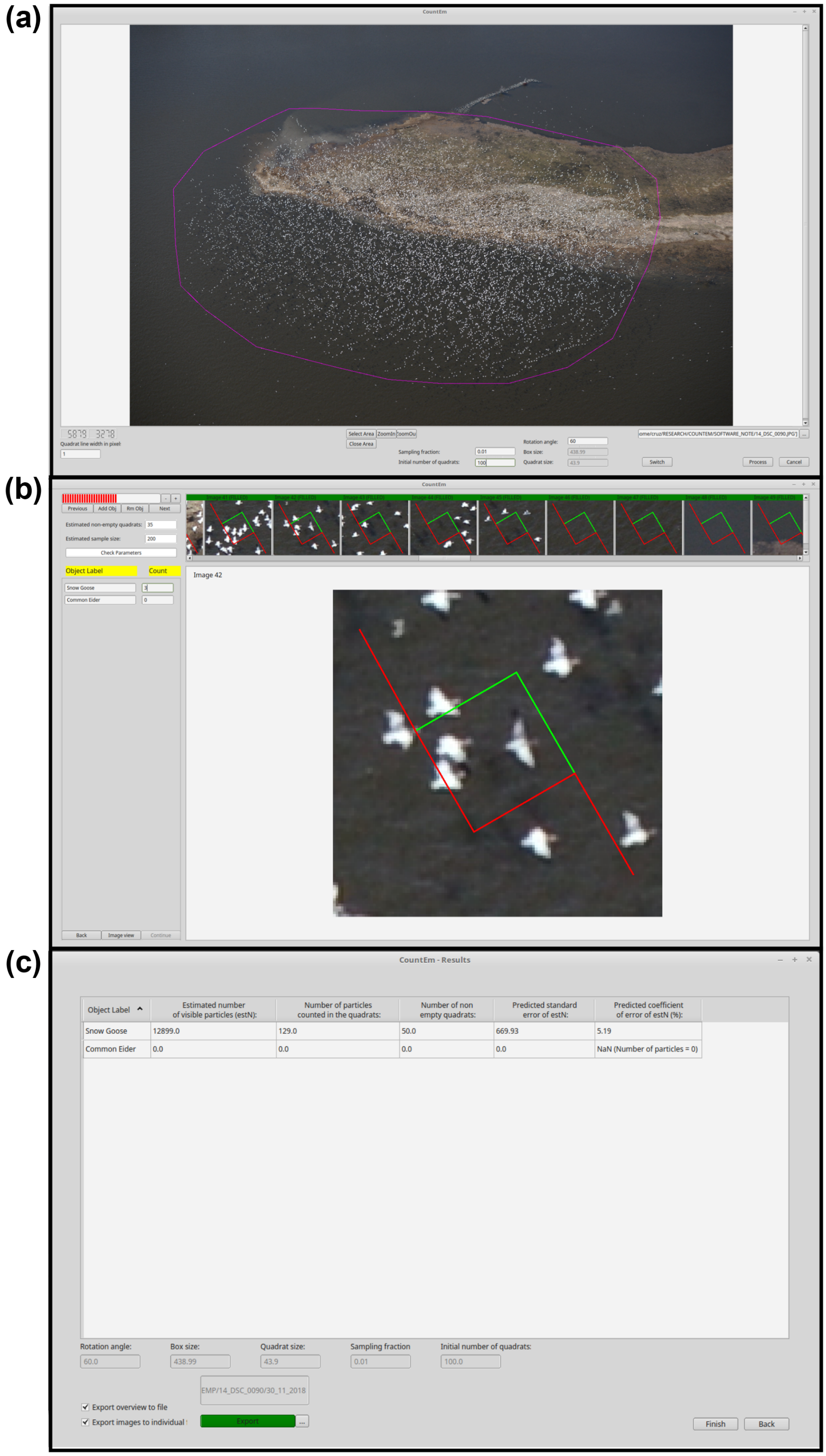

Figure 2. Screenshots of the software interface windows. (a) 'Inputs' window. (b) 'Counting' window. (c) 'Results' window.

The guided mode was chosen. After loading the image, the visual estimation of population size $\tilde{N} = 20\,000$ was introduced. Then we selected the RoI which is the big main flock of snow geese. The few birds outside the flock are gulls and can be ignored since we only aim at estimating the total number of geese. The 'Process' button led us to the next window. A uniform random superimposition of the grid of quadrats is generated and shown in the 'Overview' window as in Fig. 1c. The visual estimations for 'Estimated number of non-empty quadrats' and 'Estimated sample size' and were given, namely $\tilde{n} = 35$ and $\tilde{Q} = 105$ respectively. $\tilde{n}$ can be guessed quickly in the 'Overview' image and $\tilde{Q}$ can be obtained multiplying by the visual estimation of the average number of birds per quadrat (three in our example). Since $\tilde{n} > 30$ and $\tilde{Q} > 100$, the grid was considered to be convenient after clicking the 'Check parameters' button. The 'Image view' Fig. 2b is used to manually count the number of birds in each quadrat. The forbidden line rule was used as shown in Fig. 1b. Only snow geese were counted in our example. However, several species could be counted simultaneously, adding new counting cells with the 'Add obj' button. The 'Continue' button was clicked after filling out all the quadrat counting cells. The results are shown in the final window (Fig. 2c).

The estimated number of snow geese was $\hat{N} \approx 12\,900$ with predicted relative standard error of $\mathrm{ce}_{\mathrm{Cav}}(\hat{N}) \approx 5\%$. The sample size was $Q = 129$ and number of nonzero quadrats $n = 50$ which were within the recommended range. A relative deviation $100 \times (\hat{N} - N) / N = -6.1\%$ is obtained with a single estimation $\hat{N} = 12\,900$ and number of manual annotations $N = 13\,744$. The CountEm counting time was ~5 min.

In Table 1 we show the results of 9 further estimations to give a concrete impression of the method's variability. The mean values are given in the last row. Since the method is unbiased, the mean of the 10 estimations, 13 700, is already close to the true value $N = 13\,744$. The mean counting time was ≈5.3 min and the normalized root mean squared error was 8.1%. This example strengthens the idea that CountEm offers a fast, user-friendly and unbiased population size estimation tool.

Table 1. Results from 10 different flock size estimations on the image shown in Fig. 1c.

Run no.	$\hat{N}$	Q	Counting time (s)	Deviation	(%)
1	12 900	129	320	−844	(6.1)
2	11 700	117	276	−2044	(14.9)
3	13 800	138	329	56	(0.4)
4	14 100	141	324	356	(2.6)
5	12 400	124	294	−1344	(9.8)
6	14 800	148	277	1056	(7.7)
7	14 500	145	297	756	(5.5)
8	14 000	140	289	256	(1.9)
9	15 600	156	262	1856	(13.5)
10	13 200	132	329	−544	(4.0)
Mean	13 700	137	321	−44	(0.3)

Acknowledgments – The authors are grateful to Scott Gilliland and Josée Lefebvre for providing images and useful comments to refine the software, and to Francis St-Pierre (Canadian Wildlife Service) for providing the image used in our example. We also thank Fernando Arce for useful discussions.
Funding – MC acknowledges financial support from the AYA-2015-66357-R (MINECO/FEDER) project.
Conflict of interests – The authors declare that there is no conflict of interest regarding the publication of this article.

References

Chabot, D. and Francis, C. M. 2016. Computer-automated bird detection and counts in high-resolution aerial images: a review. – J. Field Ornithol. 87: 343–359.

Chabot, D. et al. 2015. Population census of a large common tern colony with a small unmanned aircraft. – PLoS One 10(4): e0122588.

Cruz-Orive, L. 2017. Stereology: a historical survey. – Image Anal. Stereol. 36: 153–177.

Cruz, M. and González-Villa, J. 2018. Simplified procedure for efficient and unbiased population size estimation. – PLoS One 13(11): e0208359.

Cruz, M. and González-Villa, J. 2019. Unbiased population size estimation on still gigapixel images. – Sociol. Methods Res. doi: 10.1177/0049124118799373.

Cruz, M. et al. 2015. Efficient and unbiased estimation of population size. – PLoS One 10(11): e0141868.

Descamps, S. et al. 2011. An automatic counter for aerial images of aggregations of large birds. – Bird Study 58: 302–308.

Gerum, R. C. et al. 2017. Clickpoints: an expandable toolbox for scientific image annotation and analysis. – Methods Ecol. Evol. 8: 750–756.

Gomez, A. et al. 2019. Variance prediction for population size estimation. – Image Anal. Stereol. 38: 131–139.

Gundersen, H. J. G. 1977. Notes on the estimation of the numerical density of arbitrary profiles: the edge effect. – J. Microsc. 111: 219–223.

Hodgson, J. C. et al. 2018. Drones count wildlife more accurately and precisely than humans. – Methods Ecol. Evol. 9: 1160–1167.

Hollings, T. et al. 2018. How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. – Methods Ecol. Evol. 9: 881–892.

Howard, C. V. and Reed, M. G. 2005. Unbiased stereology. Three-dimensional measurement in microscopy, 2nd edn. – Bios/Taylor & Francis, Oxford.

Kadlec, J. A. and Drury, W. H. 1968. Aerial estimation of the size of gull breeding colonies. – J. Wildl. Manage. 287–293.

Leonard, R. M. and Fish, E. B. 1974. An aerial photographic technique for censusing lesser sandhill cranes. – Wildl. Soc. Bull. 191–195.

Löffler, E. and Margules, C. 1980. Wombats detected from space. – Remote Sens. Environ. 9: 47–56.

Lyons, M. B. et al. 2019. Monitoring large and complex wildlife aggregations with drones. – Methods Ecol. Evol. 10: 1024–1035.

Schneider, C. A. et al. 2012. Nih image to imagej: 25 years of image analysis. – Nat. Methods 9: 671.