CEAMEC shiny App v1.0 User Manual

Qian TANG

tangqiannus@gmail.com

2021-03-27

1. Overview

CEAMEC (**C**ost-**E**ffective **A**nimal **M**anagement via **E**nvironmental **C**apacity), a *Shiny* app in the HTML user interface (UI) programmed with *R* language, is a tool to provide managers with cost estimation of resource-based management strategies in the population control of over-abundant nuisance species in the anthropogenic environments. Integrated with hierarchical modelling functions in *R* package *unmarked* (Fiske & Chandler, 2011) to identify the association between population density and the environmental resources, *CEAMEC* computes the change between pre-management (observed and may be subject to extant management) and post-management (user-defined management target) environmental carrying capacity and optimizes the quantity of different resources to be manipulated at the lowest cost. In this version, *CEAMEC* works for population survey data of distance sampling, repeated counts, removal sampling and double observer sampling (corresponding to *unmarked*'s hierarchical modelling functions of *distsamp*, *pcount* and *multinomPois*).

2. Access *CEAMEC*

Users can run *CEAMEC* online in the Shiny Cloud:

https://qt37t247.shinyapps.io/CEAMEC-master/

Otherwise, users can run *CEAMEC* through *R* on a local device after installing from GitHub:

https://github.com/qt37t247/CEAMEC

Example datasets and *R* scripts for the demonstration of non-interactive *CEAMEC* run (without using UI) are available from GitHub.

3. The UI

UI of *CEAMEC* (see Figure 1) comprises two tabs: the "field data input" tab for survey data input and hierarchical modelling; the "CEAMEC" tab for density visualization, cost-effectiveness analysis and results output. Under the "field data input" tab, there are three sub-tabs (namely "Distance sampling", "Repeated count" and "Removal sampling or double observer sampling") corresponding different types of population surveys and different hierarchical modelling methods. Under each sub-tab of the "field data input" tab, there are four sections (titles of sections highlighted with brown bold font). The first section (“**Distance sampling survey information**” or “**Data file composition**”) varies in the three sub-tabs because of the input data structure is different among different survey methods. The later three sections are similar across sub-tabs: “**Modelling with covariates**” section is for the generation of models by inputting the combinations of covariates; “**Models with covariates**” section is for the selection of best abundance model and covariates to be managed; “**Extent and dimension of study area**” section is for specifying size of management units and uploading environmental variables for the density estimation. Detailed explanations of all items in the UI are listed in the tooltips section at the end of the manual.

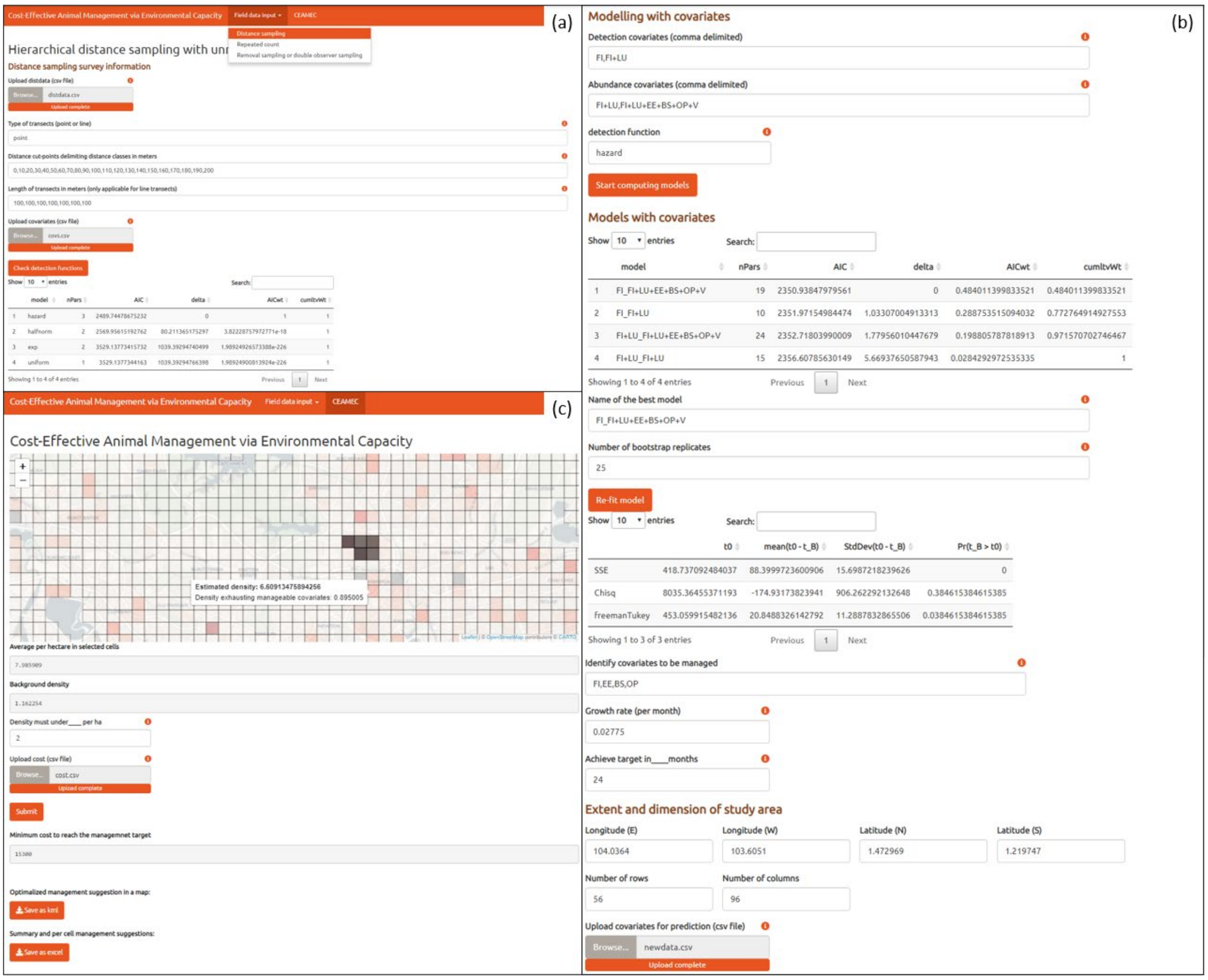


Figure 1. Screenshot of fully excuted *CEAMEC* UI with the demonstration dataset of pigeons in Singapore. (a) and (b) present interface under the “field data input” tab (Distance sampling, sub-tab); (c) presents interface under the “CEAMEC” tab.

4. Workflow

In general, *CEAMEC* intakes population survey data along with the variables collected along with the survey, the covariates, and computes the correlation between abundance and covariates using hierarchical modelling implemented in *R* package *unmarked* (Fiske & Chandler, 2011). The model, which best describe the correlation between abundance and the environment, is then used to compute the change of carrying capacity, which is calculated with user-defined management target. Based on the change of carrying capacity, *CEAMEC* applies genetic algorithm to compute cost-effectiveness for the combinations of resource reduction (resources correspond to covariates) with the unit monetory costs provided by users for the management of different resources.

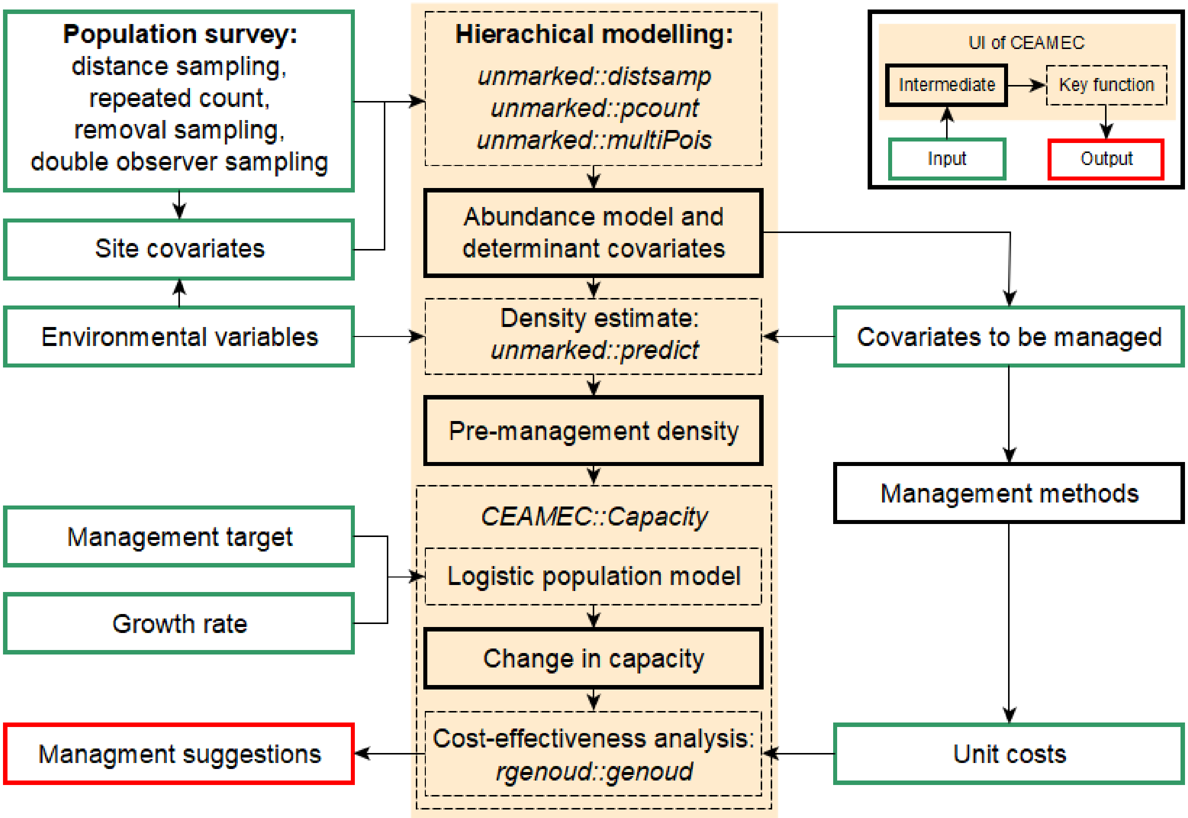


Figure 2. Workflow of *CEAMEC*.

4.1. Data Preparation

Users need to prepare **survey data file** in the comma-seperated values (csv) format. In general, the survey data file contains survey data, observation covariates (variables encountered at the same site among multiple visits, normally time, weather, seasons) and site covariates (for distance sampling data, site covariates need to be uploaded in a separate file). Site covariates are sampling-site-specific environmental variables. In the *CEAMEC*, we adopted key functions from *R* package *unmarked* for the hierarchical modelling, please kindly refer to the *unmarked* publication (Fiske & Chandler, 2011) for the mathematical process and technical details. The names of site covairates should be consistent throughout the run of *CEAMEC*. Because the names of site covariates are also used in combinations for the names of models, we recommend to use abbreviations for the covariates' name in the input file to make the subsequent flow concise. Description of input data file structure can be found in the corresponding tooltips (distance sampling, repeated count, removal sampling and double observer sampling) in the tooltip section in this manual.

To better quantify and itemize the environmental resources in the management approaches, we'd suggest using numbers that are greater than 1 for all the numeric environmental variables by altering the units. For example, instead of using "0.43" kilometer, users may need to use "430" meters. Or, instead of using "0.652", users may need to use "65.2" percent for the environmental variables in proper fractions.

CEAMEC intakes a variety of population survey data and computes hierarchical models with the corresponding functions of *unmarked*. As hierarchical models aim to explore the correlation between the abundance of species and the environmental context, it is always ideal to acquire descent number of environmental variables, especially the ones users think to contribute to the high density of the targeted species. Moreover, it is recommended to use multi-session surveys across seasons for species display significant seasonal behavior.

4.2. Survey data input

For distance sampling, users need to upload a separate file for the site covariates, in addition to the survey data file, and specify whether the survey uses point transects or line transects in the input textbox "**Type of transect (point or line)**". Moreover, users need to key in the cut-off distances for the detection modelling and the length of each line transect (applicable only if line transects are used) in the input textbox "**Distance cut-points delimiting distance classes in meters**" and "**Length of transects in meters (only applicable for line transects)**" respectively. After input the files and parameters, users need to click on the "**Check detection funcions**" button to see which detection function best fit the observation. A table appears after half a minute, presenting the comparisons of null models with four detection functions. In the later step, users may need to use the best-fit detection function, with the lowest Akaike information criterion (AIC) value, to create models with the covariates.

For repeated count, removal sampling and double observer sampling, there are input textboxes to specify the column names for survey data, observation covariates and site covariates in the survey data file. Users need to define the area of the survey in the input textbox "**Area of each survey site in hectare**". In addition, for repeated count, users could check which function best fit the abundance distribution by click on the "**Check abundance distribution**" button. A table appears after half a minute, presenting the comparisons of null models with three abundance distribution. In the later step, user may need to use the best-fit abundance distribution, with the lowest AIC value, to create models with the covariates.

4.3. Hierarchical modelling

After inputting the survey data, in the "**Modelling with covariates**" section, users need to specify which combinations of covariates are used in modelling detection (input textbox "**Detection covariates (comma delimited)**") and abundance (input textbox "**Abundance covariates (comma delimited)**") respectively. Always input "1" in both detection covariates and abundance covariates for the testing of null model. For distance sampling, users need to specify the detection function in the input textbox "**detection function**", which is normally the one with lowest AIC value from the four detection functions computed in the previous section. For repeated count, users need to specify the latent abundance distribution in the input textbox "**Latent abundance distribution (Poisson (P), negative binomial (NB) or zero-inflated Poisson random variable (ZIP))**", which is normally the one with lowest AIC value from the three abundance distributions computed in the previous section. Click the botton "**Start computing models**" to start while users are unable to access the UI until computing is finished.

4.4. Model selection and manamgenet method design

Once the hierarchical modelling with all the listed combinations of covariates is finished, a table presents at the top of the "**Models with covariates**" section. According to the table, users may enter the model name of the best model of their choice, normally the one with lowest AIC value, in the input textbox "**Name of the best model**". *CEAMEC* also provides an option for users to validate the model of choice by inspecting the distribution of simulated refitted dataset. Users may input a integer number (>10 is recommended but larger number leads to longer waiting time) in the input textbox "**Number of bootstrap replicates**" and hit "**Re-fit model**" botton. After bootstrapping, a table presents three sets of statistics: sum of squared errors ("SSE"), Pearson's Chi-squared ("Chisq"), and Freeman-Tukey Chi-squared ("freemanTukey") comparing observed data with simulated refitted data. Ideally, we consider the model is adequately fit when the observed data is not significantly different from the simulated data ("Pr(t_B>t0)" is greater than 0.05). The abundance model in the model of choice is used for the subsequent analyses of density estimation and cost-effectiveness computation. The covariates in the best model are the determinant covariates (Figure 3). Users may subset the determinant covariates for the management method designs, which are the covariates to be managed

(Figure 3), considering that resources in some of the determinant covaraites are not easily identified or easily managed.

CEAMEC only allow numeric covariates (either continuous or integers) to be managed, if there are categorical determinant covariates that may contribute essentially to the species' density, please consider converting them into numeric. For example, instead of using vegetation categories as a site covariate, users may consider using the area or proportion of a specific vegetation category. The determinant covariates not to be managed are still included in the density estimation and subsequent analyses as contributors of background density.

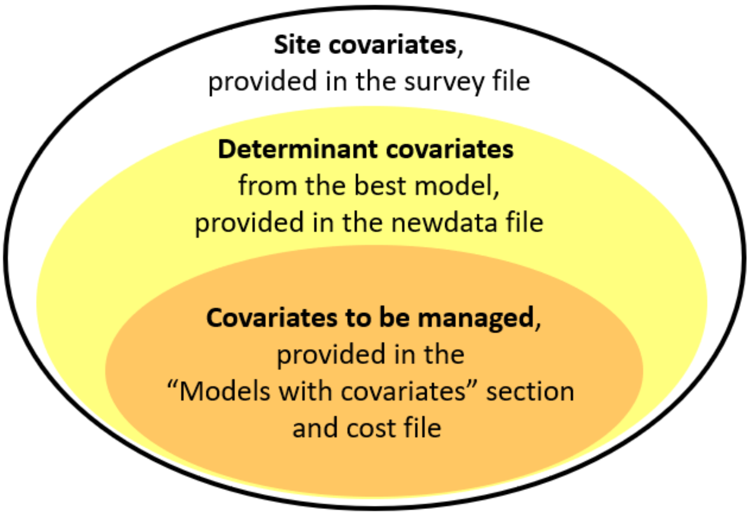


Figure 3. Graphic explanation of relations among covariates used in *CEAMEC*.

To design appropriate management methods, we have presented a case study of pigeons in Singapore in the following section as an example on how to design management methods corresponding with specific determinant covariates and calculate unit costs for each management method. In general, to calculate the cost of managing a covariate, *CEAMEC* adopt a linear equation with respect to period of management (t) and changes of covariates (Δx):

$$V = a\Delta x\, t + b\Delta x + ct + d$$

where a is cost per unit of the covariate per unit time, b is cost per unit of the covariate, c is cost per unit time, and d is fixed cost. Unit costs (a, b, c and d) of all covaraites to be managed need to be stored in a csv file (the **cost file**) and to be uuploaded for the cost-effectiveness compulation in the later step.

At the end of "**Models with covariates**" section, users also need to input the growth rate, which is often estimated from reproductive biology studies of the species. Also, users need to specify the period of management here.

4.5. Density estimation and management unit selection

With the best model chosen, users may estimate the population density of the area of study by providing environmental variables over the area of interest in a **newdata file**. In addition to uploading the newdata file, users also need to provide the extent and dimension of the raster in the "**Extent and dimension of study area**" section. The resolution of the environmental variables used in the newdata file for pre-management density estimation defines the basic spatial unit, the management unit, in the *CEAMEC* analyses. *CEAMEC* considers each management unit, where an estimated density value is given, as a closed system in which dynamics of the population is subjected to the logistic population growth model with negligible exchange of members to adjascent spatial units. Subsequently, the cost-effectiveness and the management suggestions are calculated independently for each management unit.

After uploading the newdata file, users may switch to the "CEAMEC" tab to see a map of density estimation at the top (may wait ~10s to generate). Estimated density are considered as pre-management density. The pre-management density is the species density before the planning management project, which may have been already subjected to extant management effort. The map is also used by users to select the management units for the subsequent optimization process.

4.6. Cost-effectiveness computation

Once management units are selected, users can see the average density and background density of the selected cells, refer to which, users may input the post-management density in the input textbox "**Density must under_____ per ha**". Finally, users need to upload the cost file, which is prepared during the management design. Hit the "Submit" button and wait. The process takes one minutes per management unit selected. Once the process finished, the total cost of management for all selected management units is displayed. Users can download the map, in the kml format, and visualize the optimal management suggestions for the selected management units with Google Earth or other GIS tools. In addition, users can download an excel sheet of multiple tabs, in which the first tab records the summary of the most cost-effective manamgent suggestions for the selected management units whereas subsequent tabs (one management unit per tab) record the comparisons between the optimal management suggestions with other manamgent scenarios.

5. Example data demonstration

We have provided an example dataset for the demonstration of using *CEAMEC*. The dataset is modified from a case study of feral pigeon population modelling in Singapore based on distance sampling survey carried out in 2016 (Tang et al., 2018) with additional covariates. We underlined textinputs and fileinputs and linked them to relevant content in tooltips and appendix. The timing provided were measured on a laptop of six-core, Intel i7 (2.6GHz) 12-processor with 16GB RAM.

Field data input tab (Distance sampling sub-tab)

Distance sampling survey information

We uploaded the distance sampling survey data file, distdata.csv (file description see appendix 8.1.), in the fileinput "**Upload distdata (csv file)**". We specified "point" for the input textbox "**Type of transects (point or line)**" as we used point transects and 0 to 200 meters with 10-meter intervals ("0,10,20,30,40,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190,200") for the input textbox "**Distance cut-points delimiting distance classes in meters**". We uploaded the coveriates with the file covs.csv (file description see appendix 8.1.) in the fileinput "**Upload covariates (csv file)**". We clicked on button "**Check detection functions**" to run four null models with different detection functions (the process took ~18 seconds). We confirmed that the detection of pigeons in this distance sampling case study best fits a "hazard" model (based on the lowest AIC value across all four detection functions, see Figure 1a).

Modelling with covariates

We entered "FI,FI+LU" in the input textbox of "**Detection covariates (comma delimited)**" and "FI+LU,FI+LU+EE+BS+OP+V" in the input textbox "**Abundance covariates (comma delimited)**" (Figure 1b). If you input *x* combinations of covariates in detection covariates textbox and *y* combination of covariates in abundance textbox, this step will produce $x \times y$ models. In our example, *CEAMEC* generated below four models for best model comparation (please refer to the description of covs.csv in the appendix 8.1. for the of expansion of covariates' names):

1. FI_FI+LU: FI corresponds detection while FI and LU correspond abundance;
2. FI+LU_FI+LU: FI and LU correspond detection while FI and LU correspond abundance;
3. FI_FI+LU+EE+BS+OP+V: FI corresponds detection while FI, LU, EE, BS, OP and V correspond abundance;
4. FI+LU_FI+LU+EE+BS+OP+V: FI and LU correspond detection while FI, LU, EE, BS, OP and V correspond abundance.

We entered "hazard" in the input textbox "**detection function**" as it exhibited the best fit with the field observations as shown in the result of "**Check detection functions**".

Models with covariates

After model computation (the process took ~16 minutes), models were sorted by ascending AIC values in the table generated (Figure 1b). We selected "FI_FI+LU+V+EE+BS+OP" for the input textbox "**Name of the best model**" as it exhibited the lowest AIC value. The model consists of a detection model and an abundance model, whose covariates are connected by "_". The name of the best model also suggest that the detectability of pigeons is likely correlated to the number of feeding incidents (FI), whereas the abundance of pigeons is likely correlated to six covariates (FI, LU, V, EE, BS and OP). The determinant covariates identified are "FI" (number of feeding incidents), "LU" (land use types), "V" (vegetation types), "EE" (number of eating establishments), "BS" (number of bus stops) and "OP" (length of overpasses).

We input "25" in the input textbox "**Number of bootstrap replicates**" to simulate data with the chosen model for 25 times and compare with the observed data to check adequacy of model fit. Then we clicked on button "**Re-fit model**" to proceed the simulation (the process took ~16 minutes). As result, a table lists statistics of comparisons between simulated data and observed data is generated after bootstrapping (see Figure 1b). Users may decide whether to proceed with the model (a model with Pr(t_B>t0) is greater than 0.05 is consider adequately fit).

We input names of four covariates to be managed ("FI,EE,BS,OP") in the input textbox "**Identify covariates to be managed**". We chose these four from the six determinant covariates for two reasons: first, they are directly associated with resources that support the population density of pigeons in Singapore: "FI", "EE" and "BS" are associated with food sources whereas "OP" is associated with sheltered roosts; second, they make it straightforward to demonstrate the relationship between the quantity of covariates and the quantity of resources during management method design (please read the details of manamgenet method design in response to the quantity of covariates and estimation of costs in the appendix 8.2.).

We set "0.02775" in the input textbox "**Growth rate (per month)**" for pigeons as suggested by previous studies (Johnston and Janiga, 1995). This growth rate is based on the assumption that around one third of pigeons in the entire population breeds every year; that each pair produces an average of five fledged offspring per year; and that around half of the population dies every year. We set "24" in the input textbox "**Achieve target in __ months**" for period of management.

Extent and dimension of study area

We set the study area to encompass the entire Singapore with geographic limits at 104.0364°E to 103.6051°E (east-west) and 1.472969°N to 1.219747°N (north-south). We rasterized the study area into 56 (rows) × 96 (columns) raster cells (500m × 500m for each raster cell). Finally, we uploaded newdata.csv (file description see appendix 8.1.) in the fileinput "**Upload covariates for prediction (csv file)**".

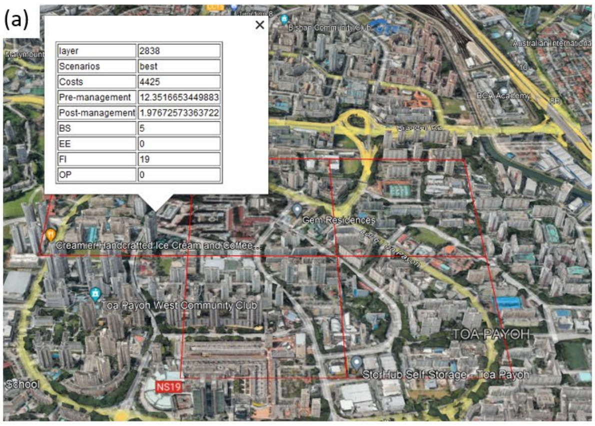
CEAMEC tab

Switching to the "**CEAMEC**" tab and wait for a few seconds, a map presented at the top displaying estimated densities across all management units (Figure 1c). Redder colour hues imply a higher pigeon density in each management unit and vice versa. Hovering over a management unit with the cursor triggers a pop-up that displays the pre-management density and the background density of the management unit. As we only declared four of the six determinant covariates as being subject to management and manipulation, the remaining two determinant covariates ("LU" and "V") contributed to the background density, which is the minimum pigeon density the management unit can reach when all the four covariates available for manipulation are being made exhaustive use of. In this example application, we selected five management units with a relatively high pigeon density (average of ~8 pigeons per hectare, see output textbox "**Average per hectare in selected cell**" in Figure 1c) and background density of ~1.16 pigeons per hectare (see output textbox "**Background density**" in Figure 1c). We input "2" in the input textbox "**Density must under __ per ha**" to set out reducing the density below two pigeons per hectare for all five selected management units. We uploaded the cost.csv (file description see appendix 8.1.) and hit the "**submit**" button to initiate the cost-effectiveness computation (the process took ~8 minutes).

In the end, *CEAMEC* produced an optimal management plan that entails a cost of 15,300 Singapore dollar (see output textbox "**Minimum cost to reach the management target**" in Figure 1c) to reduce pigeons to fewer than two per hectare within 24 months for the five management units.

We clicked "**Save as kml**" and downloaded the kml file to view the detailed management suggestions for each management unit. The kml file can be opened in Google Earth and visualized as polygons (a management unit per polygon) over the map (Figure 4a). In the pop-up table above the clicked management unit, *CEAMEC* lists the management unit ID ("layer"), management costs, pre- and post-management densities of pigeons and the quantity of covariates to be managed. The example result can be interpreted as follows: in the management unit (cell number: 2838), 4425 Singapore dollar must be spent for averting 19 feeding incidents and installing warning signs at five bus stops in order to reduce pigeons from more than 12 per hectare to less than two per hectare in two years. We clicked "**Save as excel**" and downloaded the Excel file for the summary of optimal management suggestions across all selected management units in the first tab (Figure 4b) and per management unit comparison between the best management suggestion and other, financially less optimal combinations of management methods in the subsequent tabs (Figure 4c).

(a)



(b)

Scenarios	Costs	Pre-management	Post-management	BS	EE	FI	OP	layer
best	4425	12.3516653449883	1.97672573363722	5	0	19	0	2838
best	2250	4.54928443173079	1.92087916790396	6	0	8	0	2839
best	2850	6.04928498733963	1.95738326833398	6	0	11	0	2840
best	4475	13.494078248473	1.8734680019136	7	0	19	0	2935
best	1300	3.48522998297907	1.97559315740411	8	0	3	0	2936

(c)

	A	B	C	D	E	F	G	H
1	Scenarios	Costs	Pre-management	Post-management	BS	EE	FI	OP
2	BS+0+0+0	150	12.3516653449883	9.60971573076191	6	0	0	0
3	0+EE+0+0	75600	12.3516653449883	10.7271801138585	0	105	0	0
4	BS+EE+0+0	75750	12.3516653449883	8.09609855760107	6	105	0	0
5	0+0+FI+0	4700	12.3516653449883	2.29854351619833	0	0	21	0
6	BS+0+FI+0	4850	12.3516653449883	1.50162158393739	6	0	21	0
7	0+EE+FI+0	80300	12.3516653449883	1.79340097541699	0	105	21	0
8	BS+EE+FI+0	80450	12.3516653449883	1.16225397065568	6	105	21	0
9	BS+0+0+OP	150	12.3516653449883	9.60971573076191	6	0	0	0
10	0+EE+0+OP	75600	12.3516653449883	10.7271801138585	0	105	0	0
11	BS+EE+0+OP	75750	12.3516653449883	8.09609855760107	6	105	0	0
12	0+0+FI+OP	4700	12.3516653449883	2.29854351619833	0	0	21	0
13	BS+0+FI+OP	4850	12.3516653449883	1.50162158393739	6	0	21	0
14	0+EE+FI+OP	80300	12.3516653449883	1.79340097541699	0	105	21	0
15	BS+EE+FI+OP	80450	12.3516653449883	1.16225397065568	6	105	21	0
16	best	4425	12.3516653449883	1.97672573363722	5	0	19	0

Figure 4. Results from the demonstration run of *CEAMEC*. (a) screenshot of the kml file opened in Google Earth, (b) summary table of optimal methods, and (c) per management unit table of methods comparisons.

6. Tooltips

6.1. Field data input tab

6.1.1. Distance sampling sub-tab

CEAMEC adopts hierarchical distance sampling functions from the *R* package *unmarked*, in-depth explanations of distance sampling survey input file structure and parameters can be found in the *unmarked* documents in the links below:

https://rdrr.io/cran/unmarked/f/inst/doc/distsamp.pdf

https://rdrr.io/cran/unmarked/man/unmarkedFrameDS.html

Distance sampling survey information section

Upload distdata (csv file): A csv format data frame where each row is a detected individual. Must have two columns. One for distances to the detected individuals and the other for transect names.

Example file in GitHub: distdata.csv

Type of transects (point or line): Either 'point' or 'line' for point- and line-transects.

Distance cut-points delimiting distance classes in meters: Numeric vector of distance cut-points delimiting the distance classes. Distance classes delimit detections for modelling the correlation between distance with probability of detection.

Length of transects in meters (only applicable for line transects): A vector of length R containing the trasect lengths. This is ignored if using point transects.

Upload covariates (csv file): A csv format data frame of environmental variables (covariates) that vary at the site level. Number of rows must match number of transects. Number of columns should equal to number of covariates with one column per covariate.
Example file in GitHub: covs.csv

Check detection functions: Run null models (without covariates) across four detection functions and identify which function best fit the correlation between distance with probability of detection. The detection process is modeled as multinomial: *y_ij ~ Multinomial(N_i, pi_i1, pi_i2, ..., pi_iJ)*, where *pi_ij* is the multinomial cell probability for transect *i* in distance class *j*. These are computed based upon a detection function *g(x | sigma)*, such as the half-normal (“halfnorm”), negative exponential (“exp”), uniform (“uniform”) or hazard rate (“hazard”). Output a table comparing the four null models.

Modelling with covariates section
Detection covariates (comma delimited): List all covariates or combinations of covariates to model detection (comma delimited). A combination of covariates is writen as names of covariates connected with '+'. Consider non-interactive run if many combinations are tested, otherwise very time-consumming.
Abundance covariates (comma delimited): List all covariates or combinations of covariates to model detection (comma delimited). A combination of covariates is writen as names of covariates connected with '+'. Consider non-interactive run if many combinations are tested, otherwise very time-consumming.

Detection function: One of the four detection functions: “halfnorm”, “hazard”, “exp”, or “uniform”. You could choose based on the result table if you ran check detection functions (normally the detection function with lowest AIC).

Start computing models: Compute models with all combinations of detection covariates and abundance covariates. If you input x combinations of covariates in detection covariates textbox and *y* combinations of covariates in abundance textbox, this step will produce $x \times y$ models. Each model includes two formulas for the covariates' correlation with detection probability and abundance respectively. At the end of modelling, a table will be generated to compare the statistics of all the models.

Models with covariates section

Name of the best model: One of the models computed with covariates. You could choose based on the table produced by "**Start computing models**" (normally the model with lowest AIC). As the model includes both detection model and abundance model, model name comprises detection model name followed by abundance model name connected with "_".

Number of bootstrap replicates: Number of bootstrap replicates (must be integer) to check adequacy of model fit. Can be time consuming (>1 hour) if a large number (>100) is chosen.

Re-fit model: This step simulates datasets based upon the chosen best model, refits the model, and evaluates a user-specified fit-statistic for each simulation. Comparing this sampling distribution to the observed statistic provides a means of evaluating goodness-of-fit or assessing uncertainty in a quantity of interest.

Identify covariates to be managed: Names of covariates (comma delimited).

Growth rate (per month): Growth rate can be estimated from reproductive experiments or field observations. If using anual growth rate please divide by 12.

Achieve target in____months: Length in months for the period of management.

Extent and dimension of study area section

All coordinates should be in longitude-latitude coordinate system. Any coordinates in projected coordinate system are needed to be converted.

Upload covariates for prediction (csv file): A csv format data frame containing environmental variable of rasterized area of study. Each row represents each cell in the rasterized study area with cell ID indicated at the first column. All determinant covariates should be included with one covariate per column starting from the second column.

Example file in GitHub: newdata.csv

6.1.2 Repeated count sub-tab

CEAMEC adopts hierarchical modelling functions for repeated count survey from the *R* package *unmarked*, in-depth explanations of repeated count survey input file structure and parameters can be found in the *unmarked* documents in the links below:

https://rdrr.io/cran/unmarked/man/unmarkedFramePCount.html

https://studylib.net/doc/6696451/fitting-royle-s-n-mixture-model-with-package-unmarked-in-...

Upload repeated count data (csv file): A csv format data frame of the repeated count data with observation and site covariates appended. A transect per row. Columns contains counts (one session per column), observation covariates (one session per column) and site covariates (one covariate per column). Differnt seesions can be identified with '.#' (e.g. '.1','.2','.3') in the column names.

Example file in GitHub: mld_pcount.csv

Data file composition section

Column names for counts (comma delimited): All the names of columns contain counts of all sessions.

Column names for site covariates (comma delimited): All the names of columns contain site covariates.

Column names for observation covariates (comma delimited): All the names of columns contain observation covariates of all sessions.

Area of each survey site in hectare: Normally to be the size of the transect. But if using traps, you may need to estimate the area that the trap may cover.

Check abundance distribution: Run null models (without covariates) across three abundance distributions and identify which best describe the abundance distribution. The latent abundance distribution, *f(N | theta)* can be set as a Poisson, negative binomial, or zero-inflated Poisson random variable, depending on the setting of the mixture argument, mixture = "P", mixture = "NB", mixture = "ZIP" respectively. For the first two distributions, the mean of *N_i* is *lambda_i*. If *N_i ~ NB*, then an additional parameter, *alpha*, describes dispersion (lower *alpha* implies higher variance). For the ZIP distribution, the mean is *lambda*(1-psi)*, where *psi* is the zero-inflation parameter. Output a table comparing the three null models.

Modelling with covariates section

Detection covariates (comma delimited): List all covariates or combinations of covariates to model detection (comma delimited). A combination of covariates is writen as names of covariates connected with '+'. Consider non-interactive run if many combinations are tested, otherwise very time-consumming.

Abundance covariates (comma delimited): List all covariates or combinations of covariates to model abundance (comma delimited). A combination of covariates is writen as names of covariates connected with '+'. Consider non-interactive run if many combinations are tested, otherwise very time-consumming.

Latent abundance distribution (Poisson (P), negative binomial (NB) or zero-inflated Poisson random variable (ZIP)): One of the three mixture: 'P' (Poisson), 'NB' (negative binomial) or 'ZIP' (zero-inflated Poisson). You could choose based on the table produced by "**Check abundance distribution**" (normally the mixture with lowest AIC).

Start computing models: Compute models with all combinations of detection covariates and abundance covariates. If you input *x* combinations of detection covariates and *y* abundance covariates, this step will produce $x \times y$ models. Each model includes two formulas for the covariates' correlation with detection probability and abundance respectively. At the end of modelling, a table will be generated to compare the statistics of all the models.

Models with covariates section

Name of the best model: One of the models computed with covariates. You could choose based on the table produced by "**Start computing models**" (normally the model with lowest AIC). As the model includes both detection model and abundance model, model name comprises detection model name followed by abundance model name connected with "_".

Number of bootstrap replicates: Number of bootstrap replicates to check adequacy of model fit. Can be time consuming (>1 hour) if a large number (>100) is chosen.

Re-fit model: This step simulates datasets based upon the chosen best model, refits the model, and evaluates a user-specified fit-statistic for each simulation. Comparing this sampling distribution to the observed statistic provides a means of evaluating goodness-of-fit or assessing uncertainty in a quantity of interest.

Identify covariates to be managed: Names of covariates (comma delimited).

Growth rate (per month): Growth rate can be estimated from reproductive experiments or field observations. If using anual growth rate please divide by 12.

Achieve target in____months: Length in months for the period of management.

Extent and dimension of study area section
All coordinates should be in longitude-latitude coordinate system. Any coordinates in projected coordinate system are needed to be converted.

Upload covariates for prediction (csv file): A csv format data frame containing environmental variable of rasterized area of study. Each row represents each cell in the rasterized study area with cell ID indicated at the first column. All determinant covariates should be included with one covariate per column starting from the second column.
Example file in GitHub: not available

6.1.3 Removal sampling or double observer sampling sub-tab

CEAMEC adopts hierarchical modelling functions for removal sampling or double observer sampling survey from the *R* package *unmarked*, in-depth explanations of survey input file structure and parameters can be found in the *unmarked* documents in the links below:
Removal sampling:
https://rdrr.io/cran/unmarked/man/ovendata.html

Double observer sampling:
https://rdrr.io/cran/unmarked/man/unmarkedFrameMPois.html

Upload survey data (csv file): A csv format data frame of the count data with observation and site covariates appended. A transect per row. Columns contains counts (one session per column), observation covariates (one session per column) and site covariates (one covariate per column). Differnt seesions can be identified with '.#' (e.g. '.1','.2','.3') in the column names.

Example removal sampling data file in GitHub: oven_removal.csv

Example double observer sampling data file in GitHub: fake_double.csv

Data file composition section

Column names for counts (comma delimited): All the names of columns contain counts of all sessions.

Column names for site covariates (comma delimited): All the names of columns contain site covariates.

Column names for observation covariates (comma delimited): All the names of columns contain observation covariates of all sessions.

Survey type (removal, double or depDouble): Anyone of 'removal' for removal sampling, 'double' for standard double observer sampling, or 'depDouble' for dependent double observer sampling.

Area of each survey site in hectare: Normally to be the size of the transect. But if using traps, you may need to estimate the area that the trap may cover.

Modelling with covariates section

Detection covariates (comma delimited): List all covariates or combinations of covariates to model detection (comma delimited). A combination of covariates is writen as names of covariates connected with '+'. Consider non-interactive run if many combinations are tested, otherwise very time-consumming.

Abundance covariates (comma delimited): List all covariates or combinations of covariates to model abundance (comma delimited). A combination of covariates is writen as names of covariates connected with '+'. Consider non-interactive run if many combinations are tested, otherwise very time-consumming.

Start computing models: Compute models with all combinations of detection covariates and abundance covariates. If you input *x* combinations of detection covariates and *y* abundance covariates, this step will produce $x \times y$ models. Each model includes two formulas for the covariates' correlation with detection probability and abundance respectively. At the end of modelling, a table will be generated to compare the statistics of all the models.

Models with covariates section

Name of the best model: One of the models computed with covariates. You could choose based on the table produced by "**Start computing models**" (normally the model with lowest AIC). As the model includes both detection model and abundance model, model name comprises detection model name followed by abundance model name connected with "_".

Number of bootstrap replicates: Number of bootstrap replicates to check adequacy of model fit. Can be time consuming (>1 hour) if a large number (>100) is chosen.

Re-fit model: This step simulates datasets based upon the chosen best model, refits the model, and evaluates a user-specified fit-statistic for each simulation. Comparing this sampling distribution to the observed statistic provides a means of evaluating goodness-of-fit or assessing uncertainty in a quantity of interest.

Identify covariates to be managed: Names of covariates (comma delimited).

Growth rate (per month): Growth rate can be estimated from reproductive experiments or field observations. If using anual growth rate please divide by 12.

Achieve target in____months: Length in months for the period of management.

Extent and dimension of study area section

All coordinates should be in longitude-latitude coordinate system. Any coordinates in projected coordinate system are needed to be converted.

Upload covariates for prediction (csv file): A csv format data frame containing environmental variable of rasterized area of study. Each row represents each cell in the rasterized study area with cell ID indicated at the first column. All determinant covariates should be included with one covariate per column starting from the second column.

Example file in GitHub: not available

6.2 CEAMEC tab

An interactive map, covered by raster cells, visualizes density estimated with the best model. Hovering over a cell with the cursor triggers a pop-up that displays the estimated density and the background density, the minimum density can reach when all covariates to be managed are being made exhaustive use of. Single left mouse click on a management unit to select/unselect it (multiple management units can be selected for the subsequent cost-effective calculation).

Average per hectare in selected cells: average density of animals of the selected management units.

Background density: minimum density can reach when all covariates to be managed are being made exhaustive use of. Normally contributed by environmental conditions not included in the hierarchcal model or determinant covariates not manageable. If multiple management units are selected, the value is set to the management unit of highest background density.

Density must under____ per ha: Don't set this value lower than the background density suggested in the output textbox "**Background density**".

Upload cost (csv file): A csv format data frame with one covariate to be managed per row and one of the four unit costs (a, b, c, d) per column.
Example file in GitHub: cost.csv

7. References

Fiske, I., & Chandler, R. (2011). *Unmarked*: An *R* package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, *43*(10), 1–23.

Johnston, R. F., & Janiga, M. (1995). *Feral pigeons* (Vol. 4). Oxford University Press on Demand.

Tang, Q., Low, G. W., Lim, J. Y., Gwee, C. Y., & Rheindt, F. E. (2018). Human activities and landscape features interact to closely define the distribution and dispersal of an urban commensal. *Evolutionary Applications*, *11*(9), 1598–1608.

8. Appendix

8.1. Input files for the demonstration run on the case study of pigeons in Singapore

CEAMEC adopts hierarchical distance sampling functions from the *R* package *unmarked*, in-depth explanations of distance sampling survey input file structure and parameters can be found in the *unmarked* documents in the links below:

https://rdrr.io/cran/unmarked/f/inst/doc/distsamp.pdf

https://rdrr.io/cran/unmarked/man/unmarkedFrameDS.html

1. **distdata.csv:** a file stores distance sampling survey data as two-column data frame, where each row is a detected individual. The first column is for distances between the observer and the detected individual; the second column is for transect names.

2. **covs.csv:** a file stores site covariates (one covariate per column) that vary at the site level (one site per row starting from the second row). The first row is for the names of the covariates (we suggest to use abbreviation to shorten the names of models in the subsequent modelling). In the pigeon study we demonstrated, we collected eight covariates: LU is for landuse categories; FI is for number of feeding incidences; EE is for number of eating establishments; V is for vegetation categories; BA is for building age categories; OP is for the length of over passes in meters; BS is for number of bus stops; BR is for the length of bus route in kilometers.

3. **newdata.csv:** a file stores environmental variables (one variable per column starting from the second column) across all management units of Singapore (one management unit per row starting from the second row). The first row is for the names of the covariates, which should be consistent, or subset, with the names in the covs.csv. The first column is for the cell IDs of the management units.

4. **cost.csv:** a file stores unit costs of management methods in response to covariates to be managed (one covariate per row) as five-column data frame. The first row are the names of the unit costs (a, b, c, d in response to different variables, see the detailed explanation of cost calculation in the workflow section). The first column is the the names of covariates to be managed.

8.2. Management method design and cost estimation on the case study of pigeons in Singapore

The following outlines the costs associated with each of these four management methods:

(1) To reduce feeding incidents (FI), we proposed a policy whereby persons engaging in illegal pigeon feeding ('feeders') are identified, approached and educated by management personnel. For each management unit, the resultant cost comprises a fixed cost (*d*) of $500 for the investigation over the entire management unit and a cost per feeding incident (*b*) of $200 for visiting and educating a feeder to avert one feeding incident.

(2) To reduce food sources generated by eating establishments (EE), we proposed a management plan to combine regular inspections with the disposal of exposed food waste. For each management unit, the resultant cost comprises a cost per eating establishment per month (*a*) of $30 covering administrative fees and disposal costs.

(3) To reduce the food sources (through feeding or littering) generated at or near bus stops (BS), we proposed to install "no feeding/littering" signs and warnings that such behaviour will incur fines when caught by surveillance cameras at bus stops. The resultant cost comprises a cost per bus stop (*b*) of $25 for sign installations.

(4) To reduce roosts beneath overpasses (OP), we proposed to install nets to deter pigeon entry into crevices and expansion gaps. For each management unit, the resultant cost comprises a cost per meter of overpass (*b*) of $24 for net installation and a cost per meter of overpass per month (*a*) of $0.12 for net maintenance.

We generated the cost file with rows of selected covariates to be managed and columns of unit costs (a, b, c and d, see equation in the workflow section).