CEAMEC shiny App v1.0 User Manual

Qian TANG

tangqiannus@gmail.com

2021-12-02

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 in *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*).

Access the App

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

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

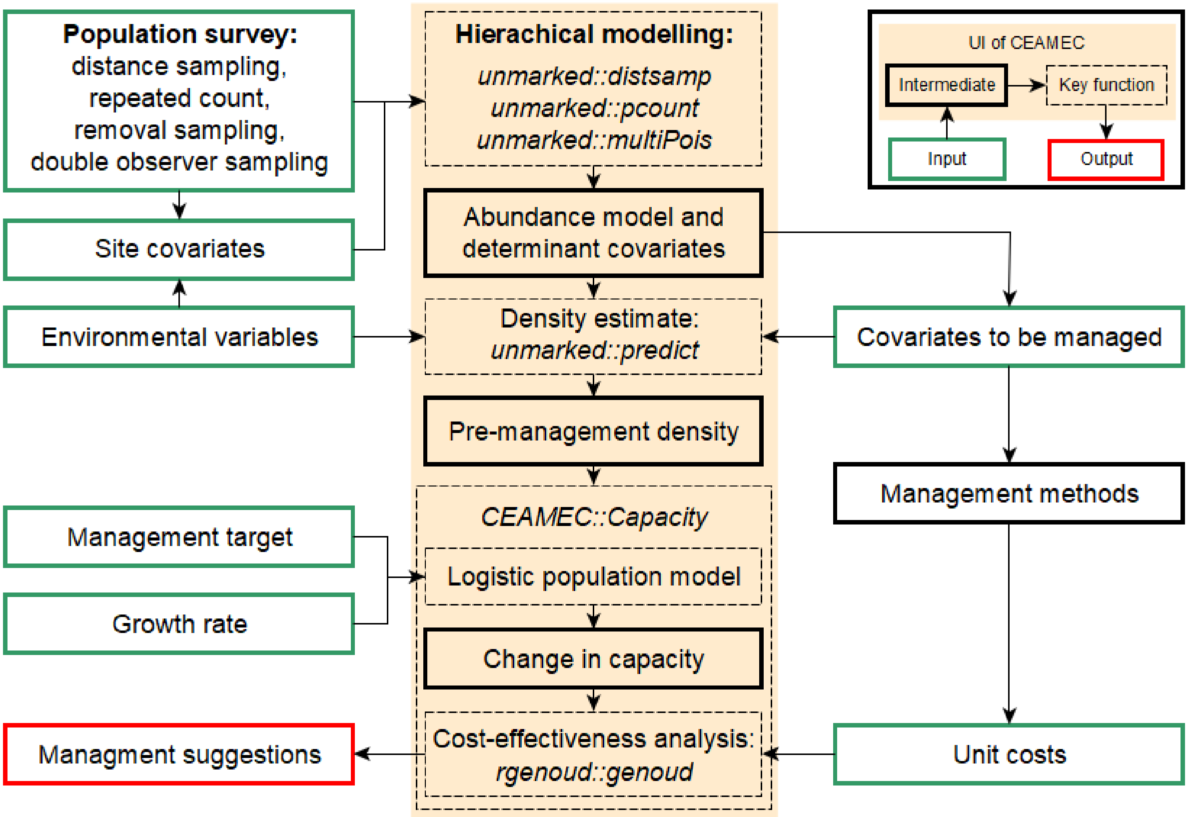
Otherwise, users can run *CEAMEC* at a local device after installing from GitHub with the code below:

```
list.of.packages <-
c("shiny","rgdal","leaflet","shinycssloaders","shinythemes","tibble","unmarked","DT","data.table","xlsx")
req.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(req.packages)) install.packages(req.packages, dependencies = TRUE)
shiny::runGitHub('CEAMEC', 'qt37t247')
```

The UI

UI of *CEAMEC* 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 corresponding different types of population surveys and different hierarchical modelling methods. Under the "field data input" tab, there are four sections (titles of sections highlighted with brown bold font). The first section 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 different 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.

Workflow



Data Preparation

Users need to prepare raw **survey file** in the comma-seperated values (csv) format before running *CEAMEC*. In general, the survey 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. The names of site covaraites should be consistent throughout the run of *CEAMEC*. Detailed structure of the survey files varies based on the types of survey, which can be found with the links:

Distance sampling:

Example data file provided in *CEAMEC* Github page: distdata.csv and cov.csv

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

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

Repeated count:

Example data file provided in *CEAMEC* Github page: mld_pcount.csv

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

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

Removal sampling:

Example data file provided in *CEAMEC* Github page: oven_removal.csv

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

Double observer sampling:

Example data file provided in *CEAMEC* Github page: fake_double.csv

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

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. Models with different combinations of site covariates are tested in the *CEAMEC* using the functions of *unmarked* to search for the best model. Moreover, it is recommended to use multi-session surveys across seasons for species display significant seasonal behavior.

Survey data input

After the preparation of survey data, users may open *CEAMEC* and upload the survey file at the first section of the “Field data input” tab.

For distance sampling, users need to upload a separate file for the site covariates and specify whether the survey uses point transects or line transects in the text box. 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). 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 will appear 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.

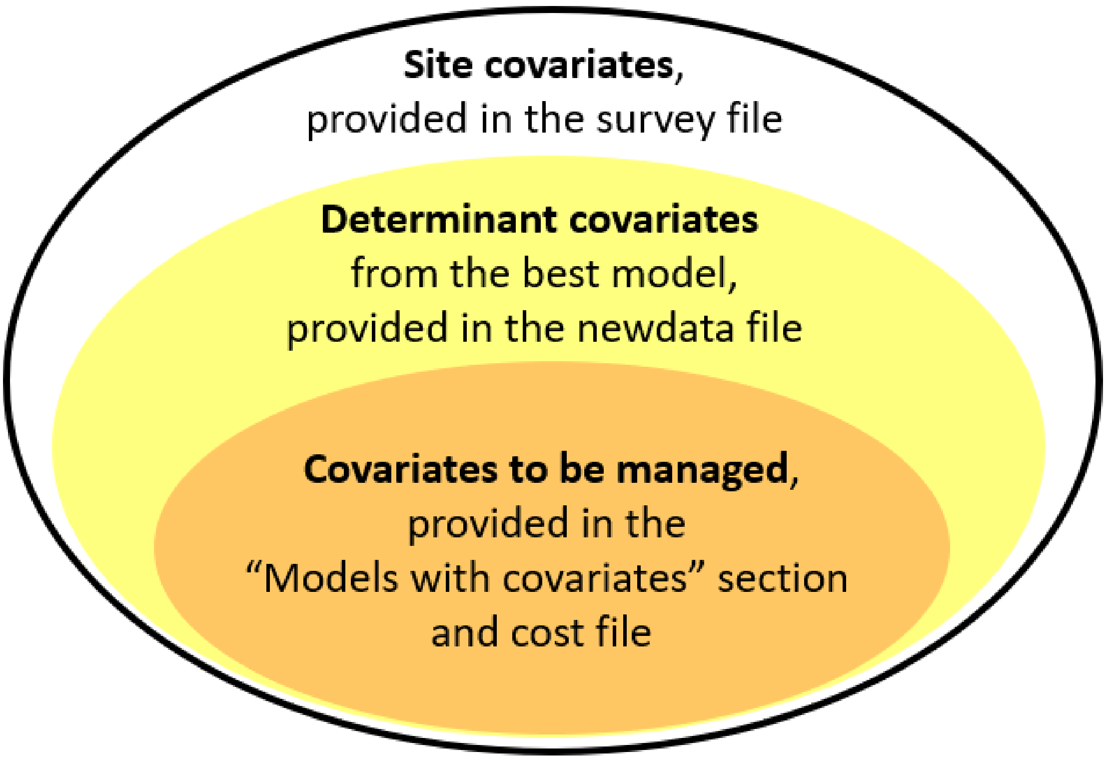
For repeated count, removal sampling and double observer sampling, there are text box to specify the column names of survey data, observation covariates and site covariates in the survey file. Moreover, users need to define the area of the survey. In addition, for repeated count, users could check which function best fit the abundance distribution by click on the “Check abundance distribution” button. A table will appear 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 Akaike information criterion (AIC) value, to create models with the covariates.

Hierarchical modelling

After inputting the survey data, users need to specify which covariate(s) will be used in modelling detection and abundance respectively in the “**Modelling with covariates**” section. Covaraites can be used in combinations that connected with “+”. Multiple covairates and combinations of covariates (delimited with “,”) can be used to generate multiple models. *CEAMEC* will generate all possible combinations for the given combinations of detection covariates and abundance covariates. For example, if users key in *D* combinations for detection covariates and *A* combinations for abundance covariates, they will expect $D \times A$ models to be computed. Each model will cost 10 seconds to a minute to compute, users should consider limiting numbers of input combinations to avoid extremely long waiting time. 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, 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, 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.

Model selection and manamgenet method design

Once the hierarchical modelling with the selected covariates is finished, a table of all the models will be presented at the top of the "**Models with covariates**" section. Follow the table, users may type the model ID of the best model of their choice, normally the one with lowest AIC value, in the text box. The model includes both detection model and abundance model as indicated in the model ID (detection covariates followed by abundance covariates delimited with "_"). The abundance model will be used for the subsequent analyses of density estimation and cost-effectiveness computation. Once the best model is determined, the covariates in the abundance model of the best model are the determinant covariates that have the strongest correlations to the abundance of the examined populations. The resources in the determinant covariates may support the level of population abundance. Limiting those resource may be helpful to control the number of individuals. Considering that resources in some of the determinant covaraites are not easily identified or easily managed to cause the population density, users need to select a subset of the determinant covariates as covaraites to be managed for the design management methods (see the illustration below).



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 categorical vegetation types for the site covariates, users may consider using the area or proportion of a certain vegetation type. The determinant covariates not selected will still be included in the density estimation and sequent analyses to support background density of the species.

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, and the length of the management period. Also, users need to specify the period of management here.

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**. The newdata file, a data frame of csv format, should include all determinant covariates (each covariate per column, start from the second column). The first column of the newdata file is the "layer", which is the cell ID that is generated during rasterization process. Each row of the newdata file represent each cell in the raster. 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 will be calculated independently for each management unit. We recommend users to determine the cell size by either running hierarchical models in parallel with different cell sizes to find the cell size where covariates have the strongest correlation with the density; or referring the size to the activity range of the species.

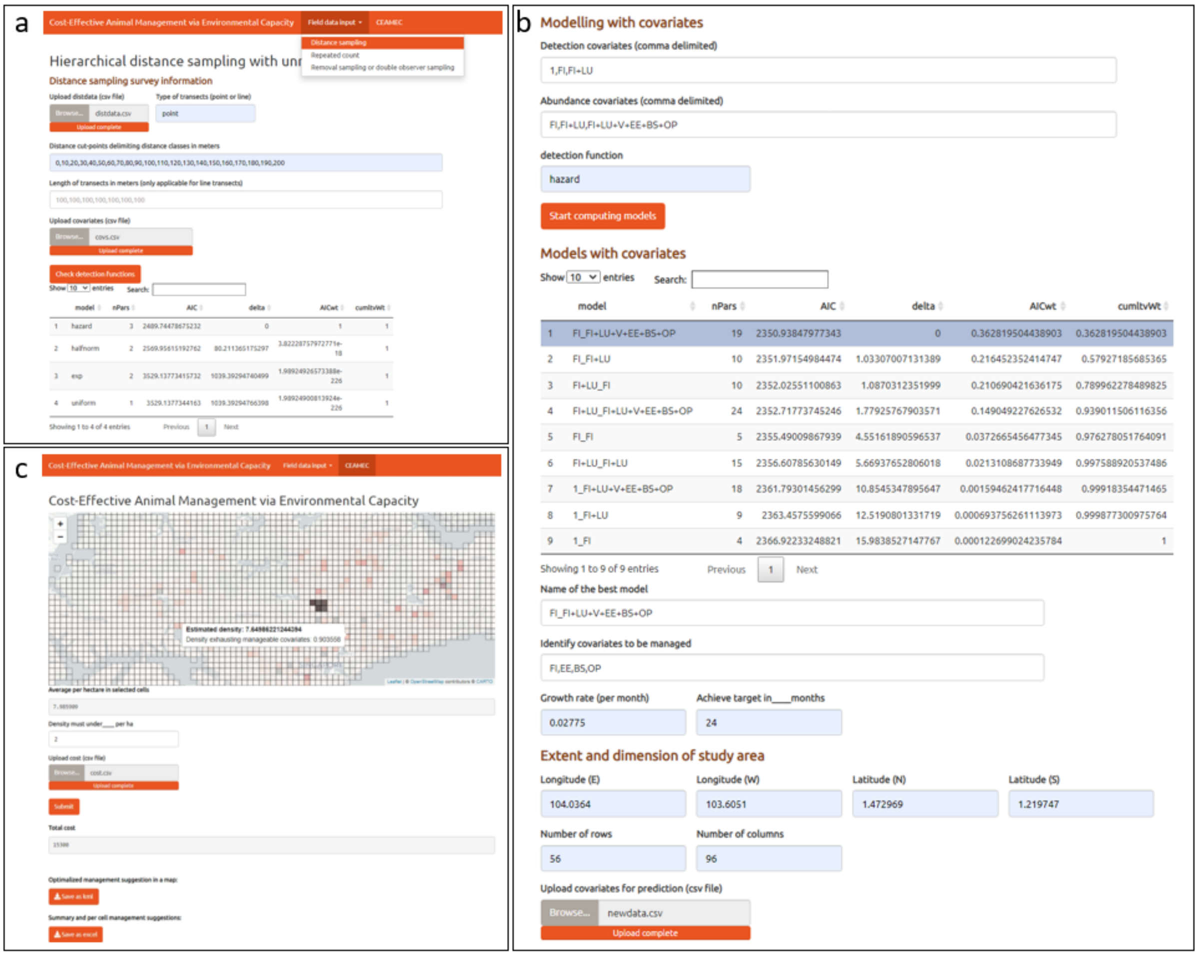
After uploading the newdata file, users may switch to the “CEAMEC” tab, at top of which, a map of density estimation is presented. Estimated density are considered as pre-management density. The pre-management density is the species density before the planning management project, which may have been already subjected to extant management effort. The map is covered by raster cells and each cell in the map represent one management unit. Redder colour hues imply a higher population density in each management unit and vice versa. Hovering over a management unit with the cursor triggers a pop-up that displays the density and the background density, the minimum density can reach when all covariates to be managed are being made exhaustive use of. Single click on a management unit to select it for the subsequent cost-effective calculation. Another click on the selected management unit to deselect it.

Cost-effectiveness computation

Once management units are selected, users can see the average density for the selected cells, refer to which, users may input the post-management density after the given period of management. Please also pay attention to the background density of every management unit selected. If the post-management density is set lower than the background density, the management costs will be calculated to infinite. Finally, users need to upload the cost file, which is prepared during the management design. Hit the “Submit” button and wait. The process takes two minutes per management unit selected. During the process, users are unable to access the UI. Once the process finished, the total cost of management for all selected management units will be 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.

Example data demostration

We have provided an example dataset for the demostration of using *CEAMEC*. The dataset is modified from a feral pigeon population modelling study in Singapore based on distance sampling survey carried out in 2016 (Tang et al., 2018) with additional covariates. Please refer to the UI below for the fully executed demonstration.



We first created three csv files for the “field data input” tab: a file of distance sampling data (distdata.csv), a file of site covariates (cov.csv) and a file of environmental variables of the entire area (newdata.csv) for the population density estimation. In the “**Distance sampling survey information**” section, we uploaded distdata.csv and cov.csv. After specifying the transect type (point transects) and distance classes, we checked and confirmed that the distance sampling detection best fits a hazard model (based on the lowest AIC value across all four detection models). In the “**Modelling with covariates**” section, we listed a combination of covariates to generate models that appeared worthwhile to explore. Again, “hazard” was selected as the detection function as it exhibited the best fit with the field observations as shown previously.

After model computation, we sorted models by their AIC values using the table at the beginning of the "**Models with covariates**" section. We selected "FI_FI+LU+V+EE+BS+OP" as the best model as it exhibited the lowest AIC value. The model consists of a detection model and abundance model, whose covariates are delimited by "_". The name of the best model also suggest that the detectability of pigeons is likely correlated to the number of feeding incidents, whereas the abundance of pigeons is likely correlated to six covariates. We adopted the abundance model for the subsequent analyses. 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 chose four of the determinant covariates to be managed, as 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. We chose these four covariates to be managed also because they make it straightforward to demonstrate the relationship between the quantity of covariates and the quantity of resources during management method design. The following outlines the costs associated with each of these four management methods:

(1) To reduce feeding incidents, 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, 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, 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, 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. We set a growth rate of 0.02775 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 also set a 24 months period of management.

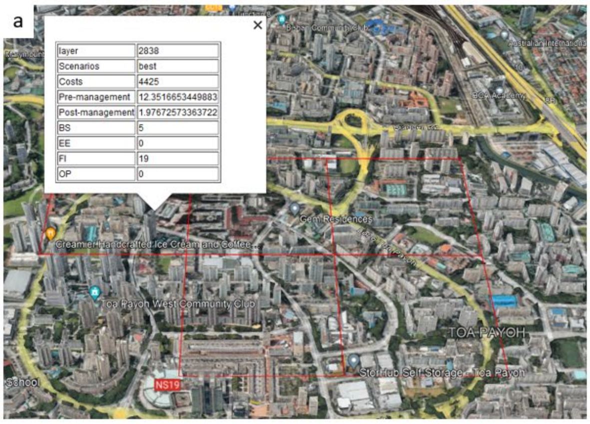
In the "**Extent and dimension of study area**" section, we set the study area to encompass the entire Republic of Singapore with geographic limits at 104.0364°E to 103.6051°E (east-west) and 1.472969°N to 1.219747°N (north-south). We rasterized the study area into 96 × 56 raster cells (500m × 500m for each raster cell), as we had previously determined that pigeon density in Singapore is best explained by environmental features at a 500m × 500m cell size, indicated by the lowest AIC value across models at different spatial resolutions (100m × 100m, 500m × 500m and 1000m × 1000m). Finally, we uploaded the newdata file of environmental variables into *CEAMEC*.

Switching to the "CEAMEC" tab within *CEAMEC*, a map is presented at the top displaying estimated densities across all management units. 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. Clicking the management unit allows a user to select it for the cost-effectiveness computation. In this example application, we selected five management units with a relatively high pigeon density (average of ~8 pigeons per hectare) and set out to reduce the density below two pigeons per hectare. We uploaded the cost file and hit the "submit" button to initiate the cost-effectiveness computation.

The entire process for the five management units took approximately five minutes on a dual core desktop of 16GB RAM. In the end, *CEAMEC* produced an optimal management plan that entails a cost of $15,300 to achieve the management target to reduce pigeons to fewer than two per hectare within 24 months for the five management units. We downloaded the kml file to view the detailed management suggestions for each management unit. We also downloaded the Excel file for a comparison between the best management suggestion and other, financially less optimal combinations of management methods.

The figure below shows the results from the demonstration run of *CEAMEC*. On the left is the kml file opened in *Google Earth*. 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 result can be interpreted as follows: in the management unit (cell number: 2838), $4425 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. On the right is the Excel output. Top is the first tab summarizing optimal management suggestions across all selected management units. Bottom is the per management unit tab of cell 2838 comparing the optimal management suggestion with other, financially less optimal combinations of management methods.

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

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.