Proof of absence modelling – data requirements

This document provides a description of the data that the Proof-of-absence model (Anderson et al., 2013) must use to fulfil its functional requirements.

1. Proof-of-Absence Model overview

For any eradication to be declared with confidence, surveillance is required to confirm (or otherwise) that eradication has been successful. The surveillance phase involves looking for any individuals that may have survived the control phase. If any individuals are found, that is evidence the eradication was unsuccessful. When a surveillance network does not detect any individuals across an area of interest (i.e. zero pest detections), it does not necessarily mean absence of the targeted pest (i.e. complete removal of the targeted pest from the area of interest), but increases our confidence in absence. The Proof-of-Absence model (Anderson et al., 2013) quantifies this confidence into a probability of absence based on the frequency of zero detections derived from a known surveillance network.

In essence, the model provides a quantitative statement that 1) previous pest control has been sufficient to eradicate the targeted pest from the area of interest, and 2) the surveillance network is sufficient to detect a recent immigrant pest into the area of interest. The conclusion depends on:

1) How confident in eradication were we before doing the surveillance (model prior) ?
2) How hard did we look for any pest that may have survived the eradication phase (model sensitivity)?
3) How confident do we want to be in our declaration of eradication (targeted probability of absence)?

The modelling approach superimposes a grid-cell system on the area of interest (extent). Each grid cell is characterized by a relative risk indicating where a pest could survive in the landscape and/or where an immigrant pest could settle after incursion in the area of interest. There are two general types of surveillance data that can be analysed by the model. The first is point detection devices, such as traps, cameras and chewcards, Each point detection device on the landscape has the potential to detect a pest that has a home range centre in grid cells that surround the surveillance device. The detection parameters for the model are g0 (the nightly probability of capture of an individual by a surveillance device placed at the centre of the animal's home range) and σ (the spatial decay parameter for a half-normal home-range kernel, to model the decline in detection probability with distance between the home-range centre and the control device). The second surveillance data type is grid-cell surveillance. An example is the probability that the public will detect a pest species in a given grid cell over the surveillance period. The associated parameters are the mean and standard deviation of the probability of detection over the surveillance period.

Following a surveillance survey in which there is evidence of the targeted pest, it is concluded that the control operation has not been successful and eradication cannot be declared (assuming these are not false detections). Following a surveillance survey in which the targeted pest is not detected (negative surveillance), the probability of absence of the targeted pest for the full area of interest is calculated as a function of the sensitivity of the surveillance system. This sensitivity of the surveillance system represents the probability of detecting one single pest in a grid cell given it was present. It represents how good our detection/surveillance network is.

The Proof-of-Absence model relies on three different types of spatial data: 1) the boundaries of the area of interest; 2) the location of the surveillance devices, sentinel captures or grid surveillance; and 3) an optional relative risk of pest survival after control (Figure 1).

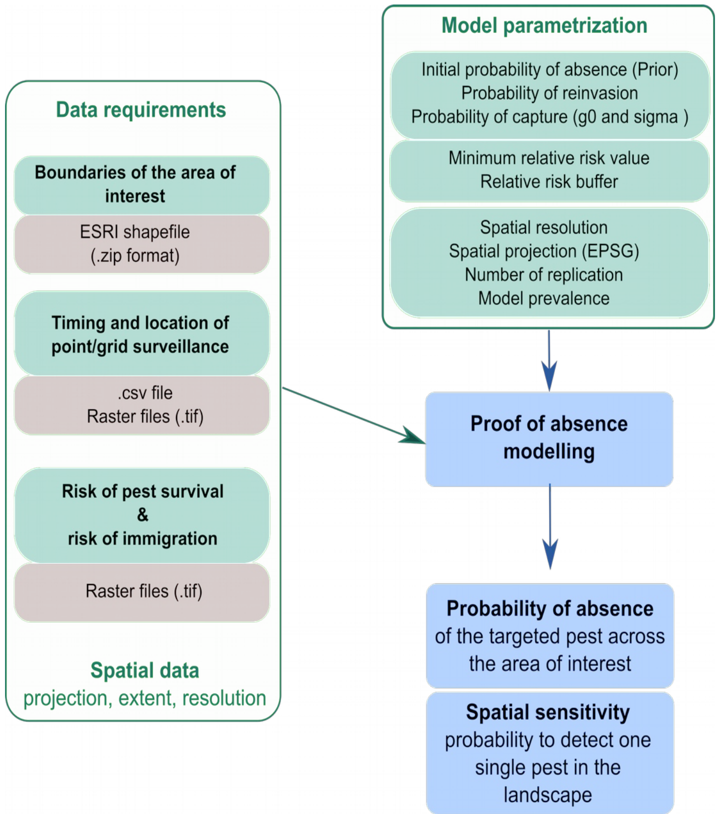


Figure 1: Proof-of-Absence model workflow

2. Web application - description of the model parameters

Page 1 – General description

Variable	Unit or format	Default value	Abbreviation	Description
Name	Character	-	-	Name of the project
Description	Character	-	-	Description of the project
Shapefile of the geographic boundaries of the treatment area	ESRI shapefile distributed as .zip file (compression format)	-	zoneShapeFName	ESRI shapefile of the geographic boundaries of the area of interest. **See section 3.1 for more details about how to format the shapefile.** *The term "shapefile" is quite common, but is misleading since the format consists of a collection of files with a common filename prefix stored in the same directory. The primary way to make shapefile data available for the Proof-of-Absence modelling is to add it to a .zip file (compression format). The .zip files*

				must contain at least the files components with the extension .shp, .shx, .dbf and .proj (or .prj).

Page 2 – Surveillance type (point or grid)

Variable	Unit or format	Default value	Abbreviation	Description
Grid surveillance (point surveillance is optional)	-	Not selected	-	Select this option when the surveillance systems takes the form of a **grid surveillance** (i.e. survey by members of the public). This option allows to use both grid and point surveillance (i.e. detection devices) simultaneously, or just grid surveillance.
Point surveillance	-	Selected	-	Select this option when the surveillance system takes the form of a **point surveillance** only.

Page 3 – Surveillance data

Variable	Unit or format	Default value	Abbreviation	Description
Timing and location of point surveillance (.csv format)	.csv	-	surveyFName	**For point surveillance only** A .csv text file indicating - year of surveillance (year) - name of the surveillance device used (species) - geographic coordinates of the surveillance device in the landscape (Easting/Northing) - length of device deployment (Tnights) **See section 3.2 for more details about how to format the .csv file**
Probability of detection	Number between 0 and 1	-	g0	**For point surveillance only** Nightly detection probability of an individual pest with a device at the centre of its home range. Provide mean and standard deviation.
Scale parameter of the home range size	meters	-	σ	**For point surveillance only** Spatial decay parameter for a half-normal home-range kernel, to model the

				decline in detection probability with distance between the home-range centre and the control device. Provide mean and standard deviation.
Timing of grid surveillance and mean probability of detecting a pest in each surveyed grid cell	.csv	Selected	gridName	**For grid surveillance only** A .csv text file indicating: - names of binary (0/1) raster map indicating the grid cells that are surveyed in the treatment area - year of surveillance - mean probability of detecting the targeted pest in each surveyed grid cell - standard deviation around the probability of detecting the targeted pest. **See section 3.3 for more details about how to format the .csv file**
Surveillance grids	.tif/ .img format distributed as .zip file (compression format)		gridSurveyRasters	**For grid surveillance only** A binary (0/1) raster map indicating the grid cells that are surveyed for each year of surveillance. For example, if the survey surveillance occurs in 2019 and 2023 , two raster maps need to be generated. Grid cells with one indicate areas that have been surveyed. All raster(s) need to be compressed into a .zip file. **See section 3.3 for more details about how to format the .zip file**

Page 4 – Relative risk map

Variable	**Unit or format**	**Default value**	**Abbreviation**	**Description**
Relative risk map of pest survival (optional)	.tif/.img format distributed as .zip file (compression format)	-	relativeRiskFName	As it is expected that the risk of pest survival in the area of interest after control is not equally distributed across the landscape, one can define a raster map with the value in each raster cell indicating where a pest is most likely to survive to control. The raster should be provided in a .zip format. **See section 3.3 for more details about how to format the relative risk map**

Minimum relative risk value	-	1	minRR	**If a relative risk map is provided** The minimum relative risk value which represents suitable habitat. Can be used to treat risk map areas below this value as non habitat.
Relative risk buffer	meters	100	RRTrapDistance	**If a relative risk map is provided** An option to set a searched area around all detection devices. Used to treat areas as suitable habitat regardless of the relative risk map.

Page 5 – Prior and probability of (re)introduction

Variable	Unit or format	Default value	Abbreviation	Description
Prior (minimum, mode, maximum)	-	-	prioMin prioMode priorMax	Estimated probability (or practitioner confidence) that eradication was successful before any surveillance was carried out. Provide a minimum, most likely and maximum value.
Probability of (re)introduction (minimum, mode, maximum)	-	-	introMin introMode introMax	Reinvasion / reintroduction probability between periods of surveillance. Provide a minimum, most likely and maximum value.

Page 6 – General settings

Variable	Unit or format	Default value	Abbreviation	Description
Spatial resolution	meters	100	resolution	Height and width of raster grid cells when search area is converted to raster cells
Number of iterations	-	-	numIterations	Number of simulations/replications. Starting parameters are drawn randomly from appropriate distributions for each iteration.
EPSG code - spatial reference	-	-	epsg	EPSG code of the used coordinate reference system. *EPSG Geodetic Parameter Dataset is a structured dataset of Coordinate Reference Systems and Coordinate Transformations, accessible through this online registry (www.epsg-registry.org)*

Design prevalence	-	1	startPu	The model will quantify the probability of absence at a specified prevalence *Pu*. This prevalence can be interpreted as the probability of detecting at least one pest in the landscape given *Pu* grid cells are occupied in the targeted area. In the context of eradication, this value is set to one, i.e. we modelled the probability of detecting at least one pest in the landscape given one occupied grid cell.
Prevalence increase rate	-	0	putIncreaseRate	Increase in designed prevalence. Population growth in a surviving population means that surveillance may be able to detect a greater number of surviving animals over time. Zero values (default) set *Pu* as constant over time.
Multiple zone ?	-	Not selected	MultiZone	Used when applying a multi-zone approach.

3. Data description

3.1 Characteristics of spatial data

The Proof-of-Absence model relies on three different types of spatial data: 1) the boundaries of the area of interest; 2) the location of the surveillance devices, sentinel captures or grid surveillance; and 3) an optional relative risk of pest survival after control (Figure 1).

Spatial data have six important characteristics:

- ***the geometry*** distinguish between two primary types of spatial data: vector (points, line, polygon) vs. raster (pixel based)
- ***the data format*** refers to the format for encoding data for storage
- ***the coordinate reference system*** of a spatial object tells the model where the raster is located in geographic space. Each coordinate reference system entity is assigned a unique EPSG code.
- ***the spatial extent*** is the geographic area that the data covers (min/max longitude/latitude).
- ***the spatial resolution of raster data*** represents the area on the ground that each pixel covers
- ***the attributes table of vector data*** are descriptions, measurements, and/or classifications of geographic features in a vector map. They are stored in a table like a spreadsheet. Each column in the table is called a field. Each row in the table is a record. Each field can store a specific type of data, such as a number, date, or piece of text.

3.2 Shapefile of the geographic boundaries of the treatment area.

Geometry: Spatial polygon (vector data)

Data format: ESRI Shapefile distributed as .zip file (compression format).

The term "shapefile" is quite common, but is misleading since the format consists of a collection of files with a common filename prefix stored in the same directory. The primary way to make shapefile data available for the Proof-of-Absence modelling is to add it to a .zip file (compression format). The .zip files must contain at least the files components with the extension .shp**,** ***.shx, .dbf and .proj (or .prj)***.

The **spatial extent of the study** must be delineated around the area of interest (Figure 2).
The **projection** is the choice of the user but must be consistent across all data in the analysis.

The targeted area can be tackled as a unique management zone (Figure 2a) or can be divided into several management zone (Figure 2b).

The **attribute table** of the shapefile of the area of interest must have at least the following four fields:

- **zoneName (string)** – name of the management zone
- **zoneID (int)** – unique ID for each management zone
- **Pu_zone (float or int)** – designed prevalence of each management zone. Given negative surveillance results, the model will assess the likelihood that the area is free of the targeted pest at a specified prevalence *Pu_zone*. This prevalence can be interpreted as the probability of detecting at least one pest in the landscape given *Pu* grid cells are occupied in the targeted area. In the context of eradication, this value is by default set to ONE (probability of detecting one single pest in the targeted are given ONE occupied grid cell).
- **RR_zone (float or int)** – Zone-level relative risk of pest survival or introduction into the management zone. The zone-level relative risk could be related to reinvasion risk, control history, or other factors. It is often assumed that the reinvasion risk might decrease with increasing distance from the boundaries of the treatment area.

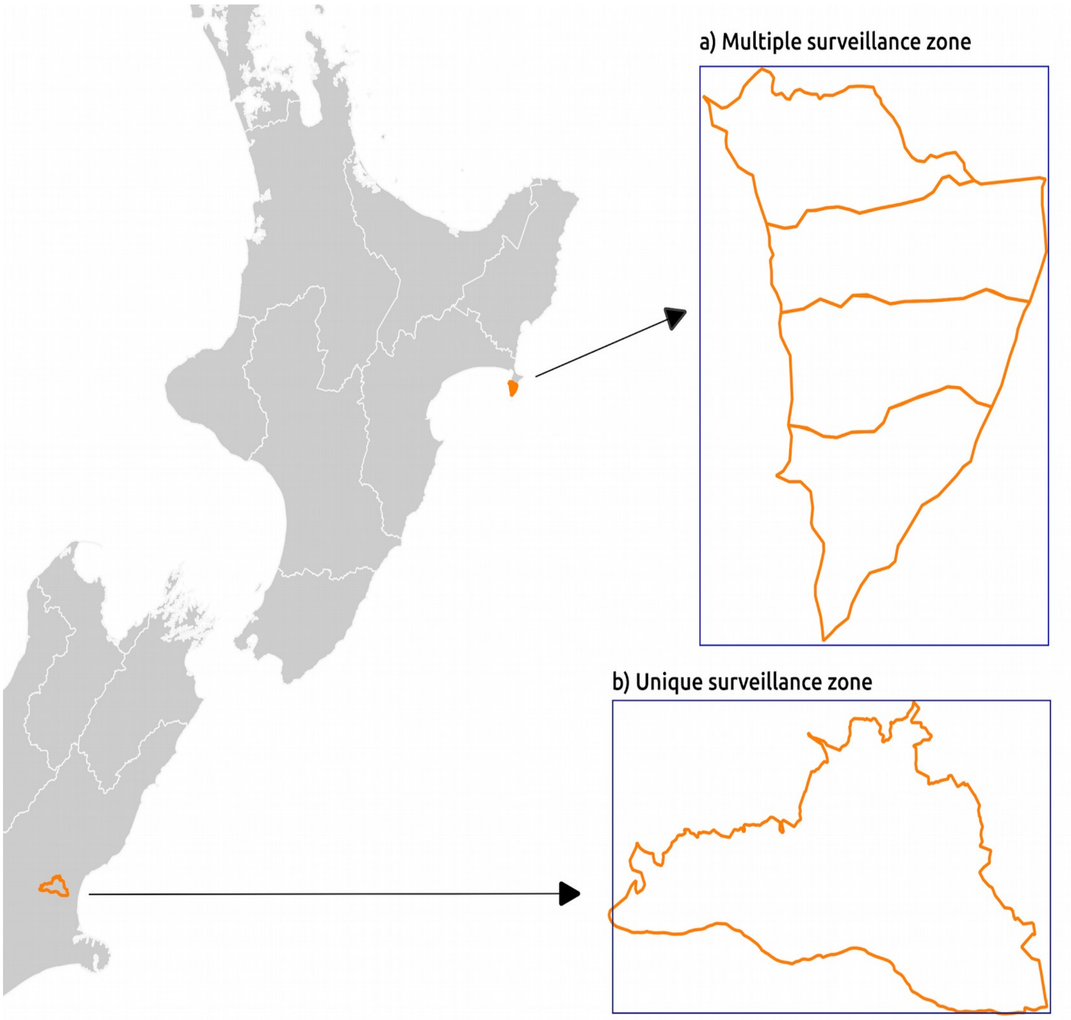


Figure 2: Example of shapefile of the boundaries of the area of interest (red lines) for a) multiple surveillance zone (each block is tackled individually as a rolling front, starting in a 1,500 ha zone at the bottom of the peninsula and working northwards) versus b) a unique surveillance zone versus. The blue rectangular indicate the extent of the study. The projection system was set using the New Zealand Transverse Mercator characterized by the EPSG code 2193.

a) Attribute table for multiple management zone

ZoneName	ZoneID	SHAPE_Length	SHAPE_Area	Pu_zone	RR_zone
Zone 1A	1	18442.1699666	1317.02391809	1	1
Zone 1B	2	17447.7969206	1419.43429931	1	2
Zone 1C	3	19212.8396093	*1543.11346030*	1	5
Zone 1D	4	12518253.01769	1251.82530177	1	10

b) Attribute table of a unique management zone

ZoneName	ZoneID	SHAPE_Length	SHAPE_Area	Pu_zone	RR_zone
Zone1	1	102440.895301	242349363.0419	1	1

Table 1: Example of attribute tables for a) a unique management zone and b) multiple management zone. Fields in gray are mandatory fields. Many other fields can be added to the table without affecting the analysis

3.3 A point and/or grid of surveillance devices/sentinels

Surveillance systems can take the form of point surveillance (i.e. detection devices) and/or grid surveillance (i.e. member of the public). Both type of surveillance can be used concurrently in the model.

Point surveillance:

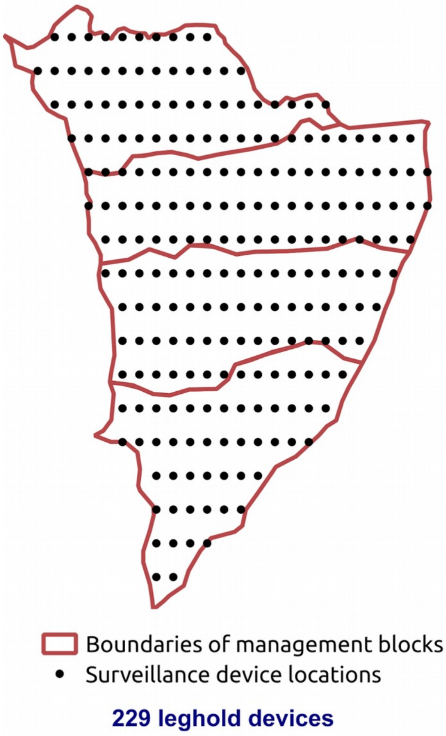
Geometry: Spatial points (vector data)

Data format: .csv text file

Each detection devices (e.g. surveillance cameras, trapping devices, chew cards) is characterized by a specific point location in the landscape (i.e. the geographic coordinates of the control device in the landscape) and a length of deployment. Several types of detection devices can be used at the same time. This information is provided in a *.csv file that must have at least the following four fields (Figure 2):

- **Year (int)** – Surveillance session, which could be one year, 1 month, 3 months, or any specified time.
- **Species (string)** – unique string referring to **specific surveillance device**. Multiple devices can be used simultaneously. Choose among (lowercase):
    - chewdetection
    - trackingtunnel
    - camera
    - legholdtrap
    - legholdraised
    - cagetrap
    - killtrap
    - people
    - dog
    - hairsnare
    - acoustic
    - waxtag
    - scat
    - ground
    - water
- **Easting (float) – easting coordinates** of the detection device (the reference system used must match the reference system of the geographic boundaries shapefile)
- **Northing (float) – northing coordinates** of the detection device (the reference system used must match the reference system of the geographic boundaries shapefile)
- Age (int) – Not relevant, set to NA
- Sex (int) – Not relevant, set to NA
- **Tnights (int) – Length of deployment** of the detection devices in number of days per session.

Each control device has the potential to detect a pest in the landscape that has a home range centre in grid cells that surround the detection device.



Delimited text file (.csv)

Year	Species	Easting	Northing	Age	Sex	Tnights
1	legholdtrap	2019808	5643227	NA	NA	56
1	legholdtrap	2020158	5343227	NA	NA	56
1	legholdtrap	2020158	5643927	NA	NA	56
1	legholdtrap	2020508	5643927	NA	NA	56
1	legholdtrap	2021908	5643927	NA	NA	56
1	legholdtrap	2019458	5643027	NA	NA	56
1	legholdtrap	2020158	5646027	NA	NA	56
1	legholdtrap	2021558	5646727	NA	NA	56
1	legholdtrap	2020508	5646727	NA	NA	56
1	legholdtrap	2020508	5644427	NA	NA	56
1	legholdtrap	2020858	5654427	NA	NA	56

229 rows, one per control devices

Figure 3: Example of delimited text file (.csv) for point surveillance data. The year can take the form of a session (1, 3, 5, etc.) or year (2019, 2020, 2022, etc.)

Grid surveillance:

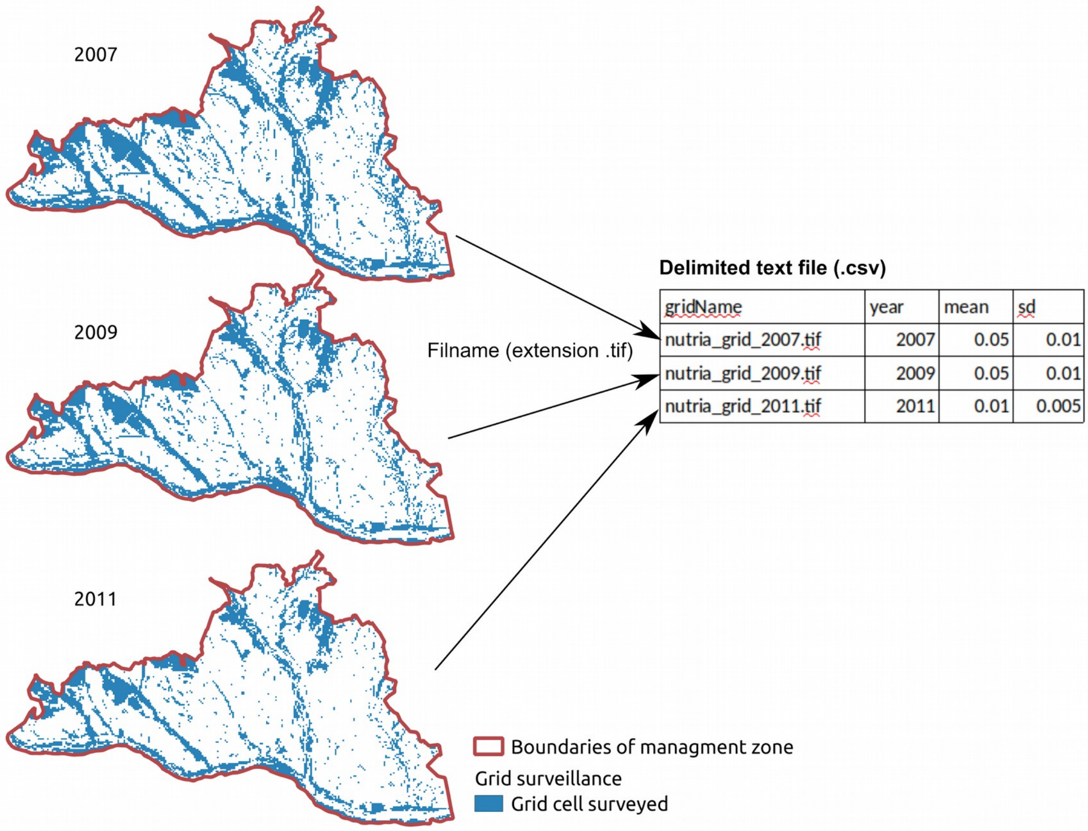
Geometry: raster data

Spatial resolution: 50m

Data format: .tif format distributed as .zip file (compression format) and the associated .csv text file

Grid surveillance relies on two types of data:

1. A binary (0/1) raster map indicating the grid cells that are surveyed for each period of surveillance. For example, if the surveillance occurs in 2019 and 2023 and the spatial surveillance differs in these two years , two raster maps need to be generated. The grid cell with one indicates the area that have been or will be surveyed.
2. A .csv text file indicating the year of surveillance and the mean probability of detecting the targeted pest in each surveyed grid cell and its standard deviation. The .csv file. must have at least the following four fields:
   - **gridName (string)** – name of the binary (0/1) raster map (.tif)
   - **year (int)** – surveillance session, which could be a year (e.g. 2019, 2019, etc) or a session (1,2,3, etc)
   - **mean (float)** – the mean probability of detecting the targeted pest in each surveyed grid cell
   - **sd (float)** – the standard deviation around the probability of detecting the targeted pest.



gridName	year	mean	sd
nutria_grid_2007.tif	2007	0.05	0.01
nutria_grid_2009.tif	2009	0.05	0.01
nutria_grid_2011.tif	2011	0.01	0.005

Figure 4: Example of raster data and associated delimited text file (.csv) for grid surveillance data.

3.4 Risk of pest survival (raster file .tif / .img)

Geometry: raster data
Spatial resolution: 50m
***Data format*:** .tif format distributed as .zip file (compression format)

It is expected that the risk of pest survival in the area of interest is not equally distributed across the area of interest. One can define a relative risk map indicating where a pest could survive to control.

Where a pest could survive ?

Distribution of mammalian pests across the landscape is often interpreted as a direct response to changes in resource availability (vegetation cover and/or prey abundance). The relative risk to habitat suggests that the risk of a mammalian pest surviving in an area is dependent on the land cover[1].
For example, for the Brushtail possum in New Zealand we commonly allocate a carrying capacity (K) to available georeferenced land cover classes in New Zealand. We used the LRIS-LCDB-v41 land cover database, along with the EcoSat indigenous forest layer to provide finer differentiation of forest classes (lris.scinfo.org.nz). The polygon dataset had 31 categories of land cover. The spatial

1While resource availability may not always correlate with population size due to environmental and temporal variability, direct information on the association between species abundance and local landscape features or micro-habitat characteristics, and how these relationships vary over space and time, is generally lacking.

layer was rasterized to a resolution of 50m so that each grid cell was characterized in isolation by the local carrying capacity.

Where an immigrant pest could settle after incursion ?

It is often assumed that the reinvasion risk might decrease with increasing distance from the boundaries of the treatment area.

In the case of Mahia Peninsula, the relative risk of possum incursion suggests that the reinvasion risk would decrease with increasing distance from the northern boundary of the area of interest, with a risk of settlement that is 10 times greater in the north than in the south of the area of interest.

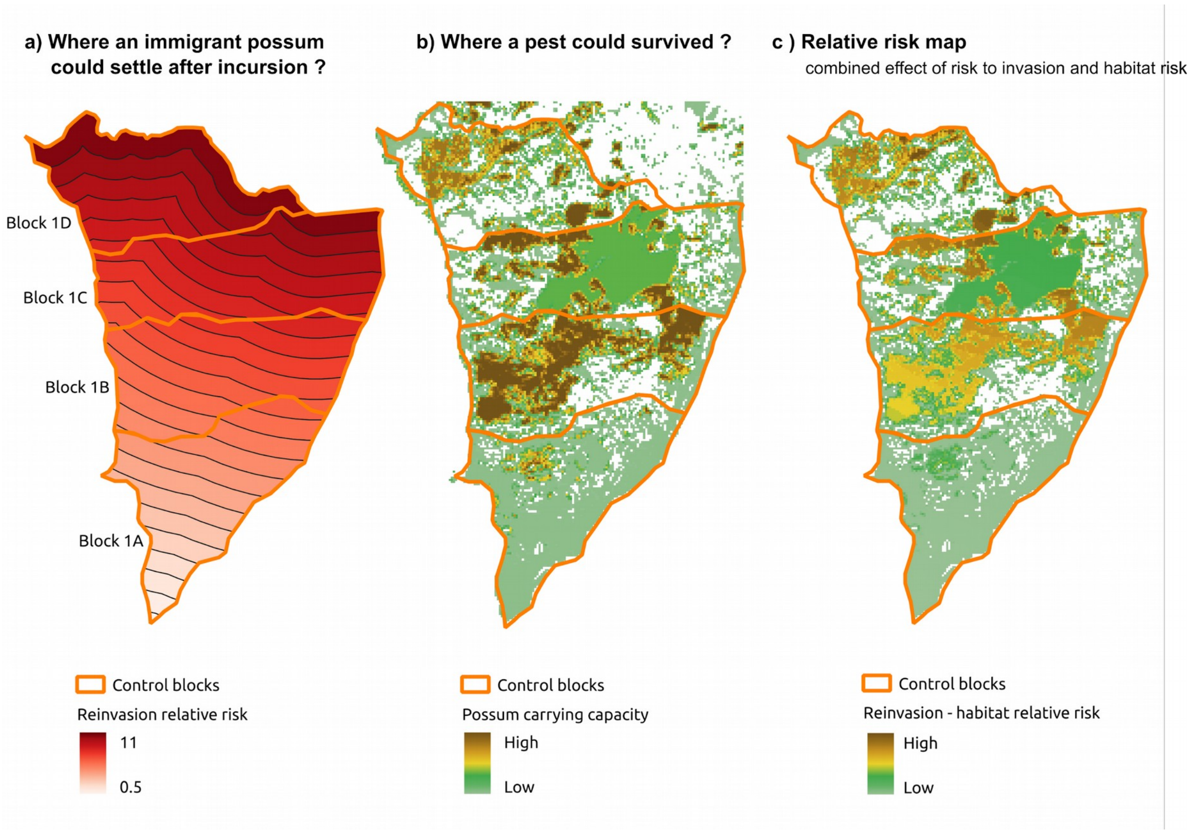


Figure 6: Relative risk map. The relative risk of the centre of activity of a) a possum immigrant and b) a possum survivor on the area of interest as a function of habitat quality. The relative risk map c) shows both the effect of risk to reinvasion and habitat.