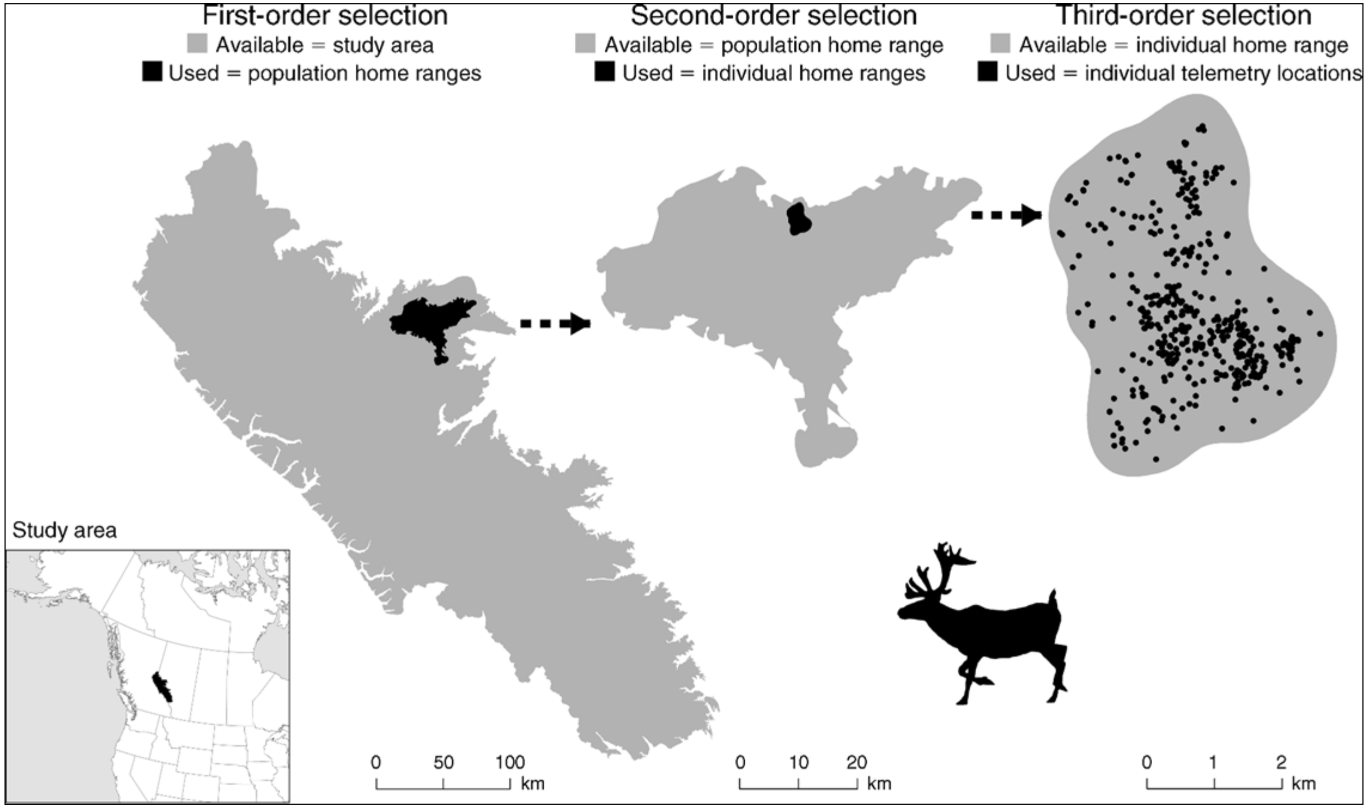
Resource Selection Analysis



Luca Börger (@lucaborger)
modified after J. Potts (U. Sheffield, UK)
UNESP Rio Claro, 23 March 2018

Swansea University
Prifysgol Abertawe

College of Science
Coleg Gwyddoniaeth

Resource selection analysis

The utilisation distribution of an animal population is proportional to the quality of available resources (Manly et al., 2002).

→ central dogma of resource selection analysis (RSA)

→ Define ‘resources’, ‘quality’, ‘available’ … not an easy task!

→ Transform into quantifiables:

$$u(\mathbf{x}) = \frac{A(\mathbf{x})W(\mathbf{x})}{\int_{\Omega} A(\mathbf{x}')W(\mathbf{x}')d\mathbf{x}'},$$

Resource selection analysis

$$u(\mathbf{x}) = \frac{A(\mathbf{x})W(\mathbf{x})}{\int_{\Omega} A(\mathbf{x}')W(\mathbf{x}')d\mathbf{x}'},$$

x = a vector denoting position,

u(x) = the utilisation distribution of the animal (i.e. the probability density of an arbitrary animal being at position x),

A(x) = availability of the point x,

W(x) = quality of the point x (the W stands for `weight'), and is the study area.

The denominator in Equation (1) is a constant, simply there to ensure that u(x) is a probability distribution defined on Ω (i.e. it integrates to 1)

Resource selection analysis

$$u(\mathbf{x}) = \frac{A(\mathbf{x})W(\mathbf{x})}{\int_\Omega A(\mathbf{x}')W(\mathbf{x}')d\mathbf{x}'},$$

Now define:

$$K_\mathrm{r} = \int_\Omega A(\mathbf{x}')W(\mathbf{x}')d\mathbf{x}',$$

→ Thus we obtain:

$$u(\mathbf{x}) = K_\mathrm{r}^{-1} A(\mathbf{x})W(\mathbf{x})$$

Which is a mathematical description of the central dogma of RSA:

“the utilisation distribution is proportional to the quality of available resources"

Resource selection analysis

Resource: anything that we hypothesise might co-vary with space use, either negatively or positively

→ covariates, function of space

→ denoted by Zi(x) for i = 1; 2; … ; n

→ vector of resources:

$$\mathbf{Z}(\mathbf{x}) = \begin{pmatrix} Z_1(\mathbf{x}) \\ Z_2(\mathbf{x}) \\ \dots \\ Z_n(\mathbf{x}) \end{pmatrix}$$

Must be carefully chosen, biological/ecological criteria!

Resource selection function

From the vector of resources we construct the weighting function W(x):

$$W(\mathbf{x}) = \exp(\beta_0 + \mathbf{Z}(\mathbf{x}) \cdot \beta),$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{pmatrix}$$

where beta is a vector denoting the weighting of each covariate, i.e. how much the covariates affects space use. The equation can also be written as:

Resource selection function

$$W(\mathbf{x}) = \boxed{\exp(\beta_0 + Z_1\beta_1 + Z_2\beta_2 + \cdots + Z_n\beta_n)}$$

Resource selection function

Resource
selection function

$$W(\mathbf{x}) = \boxed{\exp(\beta_0 + Z_1\beta_1 + Z_2\beta_2 + \cdots + Z_n\beta_n)}$$

value of beta → `quality' of the `resource' Zi

a negative value of i means that Zi is negatively correlated to space use

a positive value means Zi is positively correlated to space-use.

Mathematical perspective: the task of RSA is to infer the values of the betas.

Biologically: assessing the `quality' of `resources'.

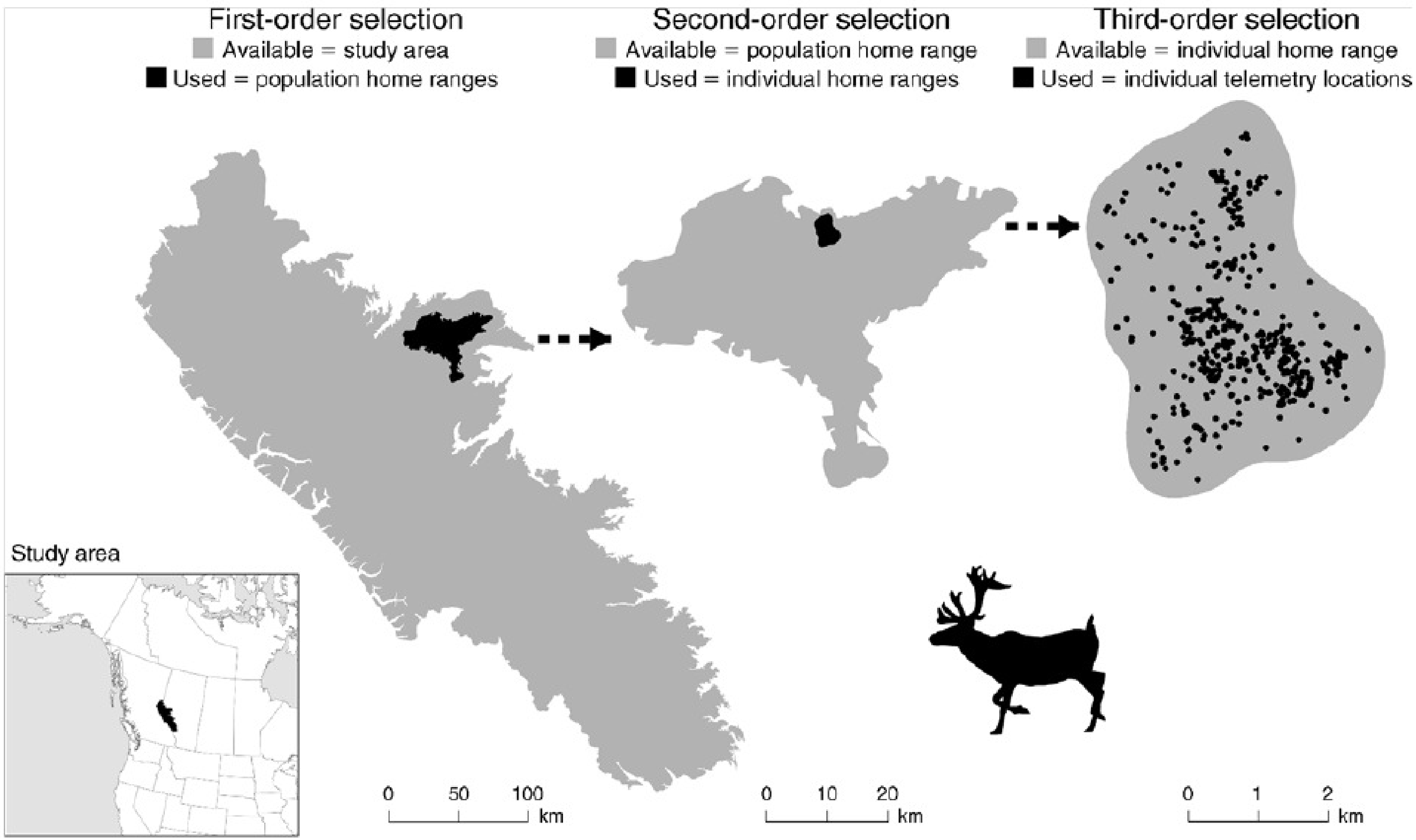
Available resources

Models typically compare locations where animals are found to. . .

- a set of ‘available’, ‘control’, ‘background’, or ‘pseudo-absence’ locations
- many ways to select points (depending on scale of inference)

‘Preference’ = used/availability depends critically on what the researcher deems is available!

Johnson, D. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 61:65-71.



Fourth order: local selection (e.g., within a feeding site)

DeCesare, et al. 2012. Transcending scale dependence in identifying habitat with resource selection functions. Ecological Applications 22(4):1068- 1083.

Orders of selection in words. . .

First order: physical or geographical range of a species.

Second order: home range (of individual or social group)

Third order: habitat within a home range

Fourth order: local selection (e.g., within a feeding site)

Definition of availability

Literature: a mess …

Problem can be (almost totally) resolved by using integrated step selection analysis (iSSA)

We first though use a naïve approach for didactical reasons

In the first equation the denominator requires to integrate over a `study area' . We can thus define availability by saying $A(x) = 1$ if $x$ is in $\Omega$ and $A(x) = 0$ otherwise. In other words, the study area is completely `available' to the animals, but anywhere outside the study area is not.
This definition is, of course, entirely dependent upon what the user defines as the `study area‘ and may have little-to-no biological relevance

Inference procedure for RSA
→ inferring the value of the beta(i)

1st define the likelihood function:

$$L_E(\beta_0, \beta_1, \ldots, \beta_n | \mathbf{X}) = \prod_{j=1}^{N} u(\mathbf{x}_j | \beta_0, \beta_1, \ldots, \beta_n)$$

RSA then seeks to find the beta values that maximise the likelihood function. Many numerical estimation methods can be used. The form of our RSF equation:

$$W(\mathbf{x}) = \exp(\beta_0 + Z_1\beta_1 + Z_2\beta_2 + \cdots + Z_n\beta_n)$$

indicates that we can use a very convenient method that's simple both to visualise and to implement:

Conditional Logistic Regression (CLR).

Conditional logistic regression for RSA

$$L_E(\beta_0, \beta_1, \ldots, \beta_n | \mathbf{X}) = \prod_{j=1}^{N} u(\mathbf{x}_j | \beta_0, \beta_1, \ldots, \beta_n)$$

We sample randomly from the distribution A(x) to give a set of \`available‘ positions (the Controls) to compare with each measured position xj (the Case). The values that maximise our likelihood function above are approximately the same as those that maximise the likelihood function for conditional logistic regression:

$$L_A(\beta_0, \beta_1, \ldots, \beta_n | \mathbf{X}) = \prod_{j=1}^{N} \frac{\exp(\beta_0 + Z_1(\mathbf{x}_j)\beta_1 + Z_2(\mathbf{x}_j)\beta_2 + \cdots + Z_n(\mathbf{x}_j)\beta_n)}{\sum_{k=1}^{M_j+1} \exp(\beta_0 + Z_1(\mathbf{s}_{kj})\beta_1 + Z_2(\mathbf{s}_{kj})\beta_2 + \cdots + Z_n(\mathbf{s}_{kj})\beta_n)}$$

where s(Mj+1)j = xj. This means it can be maximised using the built-in clogit() function in R (i.e. you are safe! ☺).

Conditional logistic regression for RSA

Now let: $S_j = \mathbf{s}_{1j}, \ldots, \mathbf{s}_{(M_j+1)j}$

Then the theory of conditional logistic regression gives the following formula for the probability that a randomly chosen element s element of Sj is the Case (Manly et al. 2002):

$$\mathsf{Prob}(\mathsf{Case} = \mathbf{s}|\beta_0, \beta_1, \ldots, \beta_n) = \frac{\exp(\beta_0 + Z_1(\mathbf{s})\beta_1 + Z_2(\mathbf{s})\beta_2 + \cdots + Z_n(\mathbf{s})\beta_n)}{1 + \exp(\beta_0 + Z_1(\mathbf{s})\beta_1 + Z_2(\mathbf{s})\beta_2 + \cdots + Z_n(\mathbf{s})\beta_n)}$$

This equation has a nice visual interpretation. The function on the right-hand side describes a logistic curve in the variable ζ where:

$$\zeta = \beta_0 + Z_1(\mathbf{s})\beta_1 + Z_2(\mathbf{s})\beta_2 + \cdots + Z_n(\mathbf{s})\beta_n$$

As the betas are varied, so the position of each point s on the -axis is varied. The Case is placed at position 1 on the vertical axis (since this is what was actually observed) and the Controls at position 0 (since these were not observed). Then maximising the likelihood function (see previous slide) corresponds to minimising the distance between these M + 1 points and the logistic curve. Let us see such a figure.

Conditional logistic regression for RSA

Here a visual exemplification of the estimation procedure:

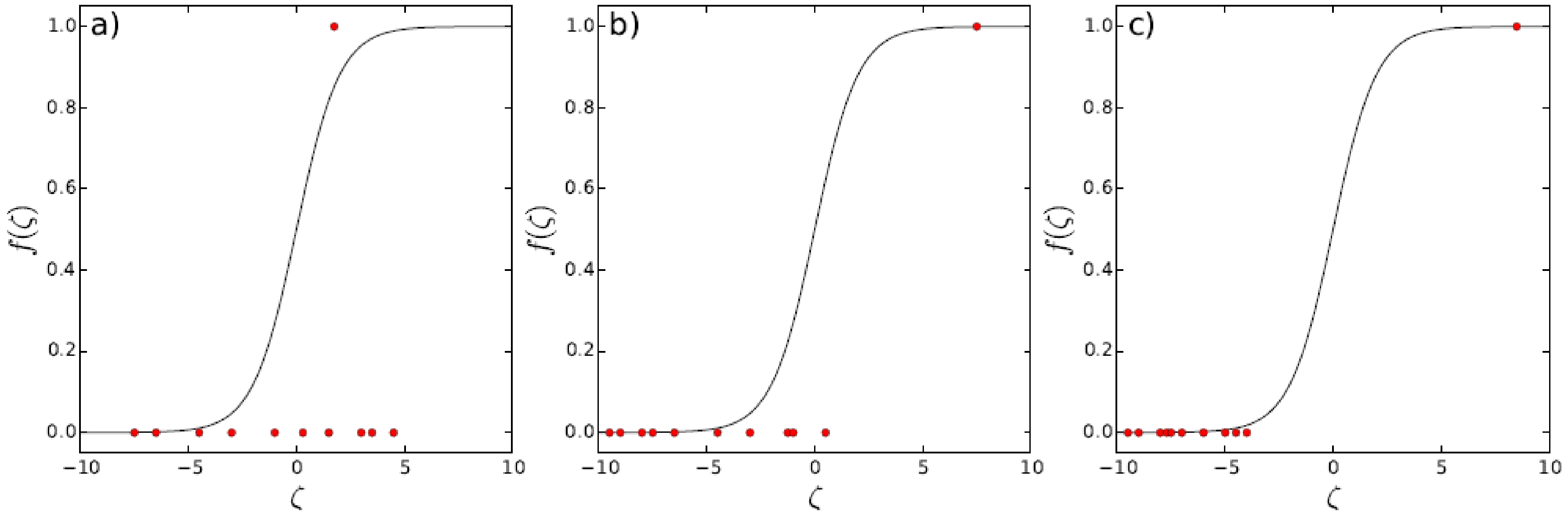


Fig. 1. Mocked-up demonstration of conditional logistic regression. The three panels all show the same logistic curve $f(\zeta) = \exp(\zeta)/[1 + \exp(\zeta)]$ in black. The red dots along the ζ-axis are the ζ-values of Control points, $\mathbf{s}_{1j}, \ldots, \mathbf{s}_{Mj}$ (where $M = 10$), given by Equation (11). The red dot on the line $f(\zeta) = 1$ is the ζ-value of the Case point, $\mathbf{x}_j$. Each panel represents a different hypothetical set of values for $\beta_0, \beta_1, \ldots, \beta_n$, leading to different ζ-values for the Controls and the Case. The closer the red dots are to the black curve, the better the fit. So panel (a) is the worst fit, followed by panel (b), then panel (c). Conditional logistic regression seeks to vary $\beta_0, \beta_1, \ldots, \beta_n$ to find the best fit of the red dots to the black curve.

RSA in R

Using simulated data ('playing God') here an example of the analysis in R. We denote by r(x) the resource quality at a location x. Assume God decided that the utilisation distribution, u(x) is proportional to exp(βr(x)) for some beta. Then play the scientist who has sampled 200 animal locations, under 3 different scenarios of beta values (which the scientist does not know). These each correspond to 200 samples from u(x) = exp(βr(x)), a fact that God knows but the scientist does not.

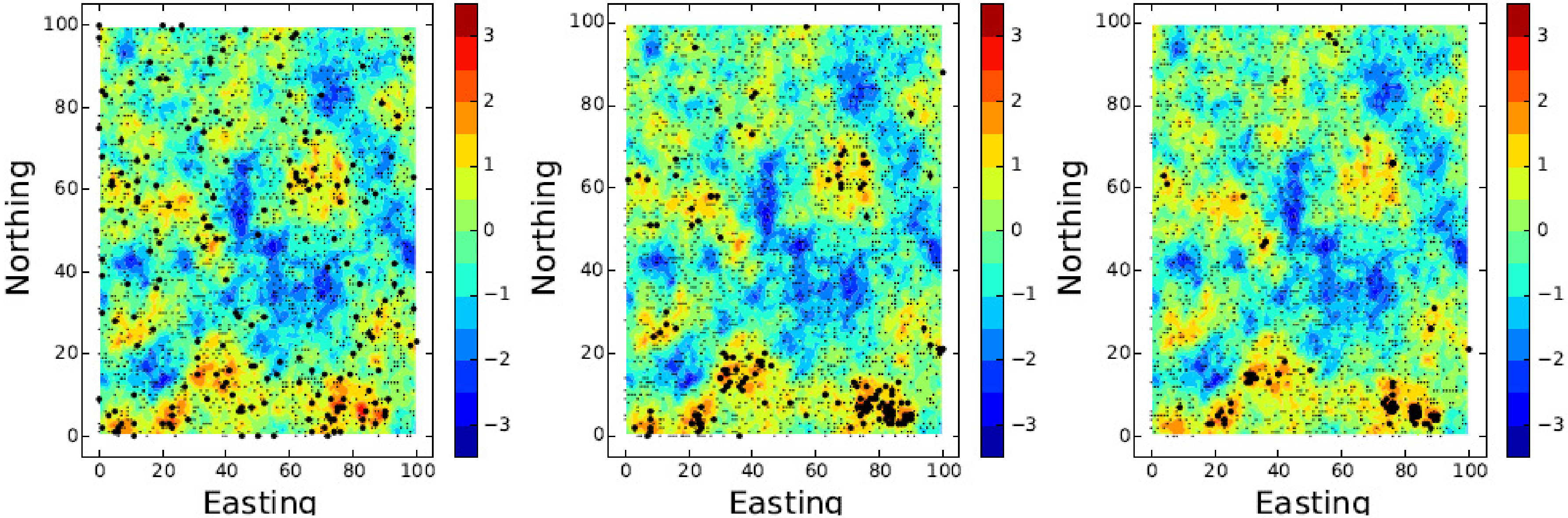


Fig. 2. Simulated data for resource selection analysis. The colour contours for each panel show resource quality for a simulated landscape, giving a function $r(\mathbf{x})$ of the resource quality at each point, $\mathbf{x}$. The large dots in each panel show simulated animal positions, obtained by sampling from a population distribution proportional to $\exp(\beta r(\mathbf{x}))$. In the left panel, $\beta = 1$; in the central panel, $\beta = 3$; in the right panel, $\beta = 5$. The small dots on each panel give control positions sampled uniformly at random from the landscape.

RSA in R

For each of the 200 `real' animal locations (Cases), the scientist samples 10 random points from the terrain, thus 2000 random samples (Controls). The scientist then summarises the complete set of 2,200 data points in a table, which is placed in a CSV (Comma Separated Value) file. The three tables for the panels of Fig. 2 are given in sim_rsf_rf1_beta1_1.csv, sim_rsf_rf1_beta1_1.csv sv, sim_rsf_rf1_beta1_1.csv for beta equal to 1,3,5, respectively. This format can be fed directly into R for use with clogit().

```
require(`survival')

rsfex1 = read.csv("sim rsf rf1 beta1 1.csv ")

beta1sim = clogit(Observed ~ resource + strata(strata), data = rsfex1)

# Observed is the first column in the table from the csv file. Each entry has value 1 if
# the row is a Case and 0 if it is a Control.
# resource is the fifth column, giving the value of the resources.
# strata(strata) tells the code which Controls are coupled to each Case.
# There are 200 Cases, labelled 1 to 200.
# The fourth column gives a number that denotes the Case to which each Control is
# associated.
```

RSA in R

Typing beta1sim gives us the model output:

beta1sim

```
Call:
clogit(rsfex1$Observed ~ rsfex1$resource + strata(rsfex1$strata),
data = rsfex1)

                  coef    exp(coef)  se(coef)  z     p
 rsfex1$resource  0.9869  2.6830     0.0919    10.7  <2e-16

Likelihood ratio test=136 on 1 df, p=0
n= 2200, number of events= 200
```

The first thing to look at is the number under exp(coef). This is the inferred value of
β. Second, the number under se(coef) gives the standard deviation of the inferred
β. Hence we have inferred that β = 0.9869 +/- 0.0919. For a simulated value = 1
the inference is pretty good.

repeat this now, then try for beta = 3 and 5, add a quadratic function,
then on real elk data.