ECOGRAPHY

Software notes

oSCR: a spatial capture–recapture R package for inference about spatial ecological processes

EDITOR'S CHOICE

Chris Sutherland, J. Andrew Royle and Daniel W. Linden

C. Sutherland ✉ (csutherland@umass.edu), Univ. of Massachusetts, Amherst, MA, USA. – J. A. Royle, US Geological Survey Patuxent Wildlife Research Center, Laurel, MD, USA. – D. W. Linden, NOAA National Marine Fisheries Service, Gloucester, MA, 01930, USA.

Spatial capture–recapture (SCR) methods have become widely applied in ecology. The immediate adoption of SCR is due to the fact that it resolves some major criticisms of traditional capture–recapture methods related to heterogeneity in detectabililty, and the emergence of new technologies (e.g. camera traps, non-invasive genetics) that have vastly improved our ability to collection spatially explicit observation data on individuals. However, the utility of SCR methods reaches far beyond simply convenience and data availability. SCR presents a formal statistical framework that can be used to test explicit hypotheses about core elements of population and landscape ecology, and has profound implications for how we study animal populations. In this software note, we describe the technical basis and analytical workflow of `oSCR`, an `R` package for analyzing spatial encounter history data using a multi-session sex-structured likelihood. The impetus for developing `oSCR` was to create an accessible and transparent analysis tool that allows users to conveniently and intuitively formulate statistical models that map directly to fundamental processes of interest in spatial population ecology (e.g. space use, resource selection, density and connectivity). We have placed an emphasis on creating a transparent and accessible code base that is coupled with a logical workflow that we hope stimulates active participation in further technical developments.

Keywords: capture-recapture, density, landscape ecology, SCR, SECR, spatial ecology

Background

Spatial capture–recapture (SCR) was originally developed as an extension of traditional (i.e. non-spatial) mark–recapture to accommodate live trapping of animals (Efford 2004) but is now acknowledged as a profound advancement in spatial statistical modeling of animal populations (Borchers et al. 2016, Royle et al. 2018). SCR represents a unifying framework for investigating many aspects of spatial population ecology using data collected by the repeated encountering or capturing of individual animals (i.e. encounter histories) across space. With over a decade of development and application, ecologists have used SCR to examine a variety of topics including landscape and network connectivity, demography, resource selection and movement and dispersal (reviewed by Royle et al. 2018).

NSO
NORDIC SOCIETY OIKOS

www.ecography.org

As with other hierarchical models in ecology (Royle and Dorazio 2008), the formal linkage between ecological processes and observational processes in an SCR model has allowed for better inferences on the former and more robust accommodation of the latter. In particular, SCR considers the collection of individuals in a population as a latent spatial point process of activity centers represented by coordinates distributed within some region of interest. A probability model for the activity centers can be used to estimate variation in density across space as it relates to environmental attributes or summarized to provide the density/abundance for a given area and time frame. Conditional on the latent point process, the probability of observing or encountering an individual at a sampling device is a function of the distance between the individual's activity center and the location of the sampling device. This distance function approximates the average scale of individual movement about an activity center and is often associated with home range features (Royle et al. 2014). By adding spatial context to the ecological and observational processes, SCR overcomes several deficiencies of traditional mark–recapture related to defining the sampling region and it expands the scope of potential inferences about both individual and population level attributes. Any survey methods that produce spatial encounters of individuals can provide the necessary data for fitting SCR models and exploring ecological hypotheses about the spatial structure of animal populations.

The rapid adoption of SCR has been supported by new technologies that ease the collection of spatial encounter information and by analytical resources that make model fitting for such data more accessible to a wider audience. Non-invasive sampling techniques where individuals can be encountered and identities determined without physical capture allow researchers to efficiently sample large landscapes, particularly useful for wide-ranging species (Long 2008). The most common techniques are camera trapping, in which remote cameras provide individual encounters for species with unique natural markings; and non-invasive genetic sampling (NGS), where genetic material collected through active (e.g. hair snare) or passive (e.g. scat) methods yields DNA for identification of individuals for any species. The collection of marks, whether by photograph or genotype, can occur at fixed locations in space or by search–encounter approaches where observers use systematic spatial surveys and record unique locations of individual encounters. This flexibility in data collection enables the application of SCR methods to population estimation for a wide range of species.

Regardless of the technique used to collect spatial encounter histories, many variations of an SCR model can be fit in freely available software programs across a spectrum of user proficiency and potential for customization. Programs like DENSITY (Efford et al. 2004) and its R package descendent `secr` (Efford 2019) are designed to allow both novice and advanced users to fit a wide collection of SCR models using maximum likelihood estimation, though custom specifications are not possible. Intermediate users can write custom SCR likelihood specifications with the BUGS language (Lunn et al. 2009) to estimate parameters with a Bayesian approach by Markov chain Monte Carlo (MCMC), as first illustrated in Royle and Young (2008). A major advantage of the BUGS approach is the ease with which users can develop their own code (Kéry 2010) or adapt code already published (Royle et al. 2014). A disadvantage is that complex BUGS models can be notoriously slow, limiting the ability to easily explore multiple model structures or accommodate complex spatio–temporal dynamics. It is noteworthy, however, that these challenges have been addressed somewhat with the recent development of NIMBLE (de Valpine et al. 2017). Advanced users can freely construct custom code to address specific modeling problems, though high specificity can result in limited utility beyond the original purpose (and, thus, low adoption/application by other users). The tremendous benefit of open source software is that some initial efforts at generality can provide a framework upon which additional optional complexity can be built by interested collaborators, facilitated by community-shared development platforms (e.g. GitHub). With this notion in mind, we present the `oSCR` package for fitting SCR models in `R`.

We started developing the oSCR project in 2014 to achieve two key objectives: First, we wanted a platform for fitting basic SCR models that accommodated non-Euclidean distance models in the form of least-cost path (Royle et al. 2013b, Sutherland et al. 2014) and handled full sex-specificity using a full likelihood approach in which fully or partially observed sex data are included in the likelihood (Royle et al. 2015, Fuller et al. 2016). Second, we required a fully open and accessible implementation of the package so that a user moderately proficient in computing and statistics could understand, verify, and extend its capabilities and ensure timely development to meet rapidly developing analytical needs. Thus, `oSCR` is implemented completely in native R, meaning that anyone proficient in `R` can contribute to the project. From a conceptual standpoint, we aimed to provide a platform for using individual encounter history data to address the three core elements of spatial population ecology (Royle et al. 2018): 1) modeling spatial variation in density using inhomogeneous Poisson point process models (Borchers and Efford 2008); 2) modeling landscape connectivity using a model which allows for effective distance to be parameterized in terms of landscape structure, with the parameters of that 'cost-distance' model estimated from observed data; and 3) integration of explicit resource selection models and telemetry data into spatial capture–recapture models (Royle et al. 2015, Linden et al. 2018). These concepts define the core functionality of `oSCR`. In this note, we describe the technical basis and analytical workflow of `oSCR` (Table 1, 2).

Methods and features

The modeling framework implemented in the `oSCR` package is that of a multi-session sex-structured spatial capture–recapture (MSSS-SCR) model. The two key

Table 1. A description of a typical `oSCR` workflow, including the relevant `oSCR` functions. A full documented version of the workflow is provided as a Supplementary vignette Appendix 1.

Workflow step	Process	In `oSCR`
Set up	Read in TDF and EDF objects. The EDF and TDF files should be read into R. This is not an `oSCR` process.	–
Formatting	Process and link EDF and TDF files to create the main `oSCR` data object, an `scrFrame`.	`data2oscr()`
	Plotting an `scrFrame` produces visualisations of the individual encounters and detector locations.	`plot()`
	Create a state space object with the desired characteristics (i.e. resolution and buffer).	`make.ssDF()`
	Plotting an `ssDF` produces a visualisation of the state space and, if an `scrFrame` is supplied, trap locations.	`plot()`
Model fitting	Use standard R formula notation to specify, and fit, models for density, baseline detection, and sigma.	`oSCR.fit()`
Model selection	Create an `oSCR`-specific list of fitted objects that can be manipulated to produce standard AIC-based model ranking and summaries.	`fitList.oSCR()`
	Produce a ranked model table and model×parameter coefficient matrix.	`modSel.oSCR()`
Prediction	Produce predicted values from a fitted density, detection or sigma model. Predictions and 95% confidence intervals are on the 'real' scale.	`get.real()`

features of the construction of the MSSS-SCR likelihood are (a) it accommodates sex structure in the form of a categorical individual covariate (Royle et al. 2015) and an additional parameter which is the population probability that an individual belongs to the baseline group (e.g. 'female'), and (b) it accommodates discrete groups or sessions where N_g is the population size of group g (e.g. N_{female}). We feel that these two features, along with the additional essential elements of the SCR model, address a large majority of applications. However, we also note that encounter data does not have to be stratified by independent sessions, nor is it essential that information about sex

is available. Rather, we developed the multi-session sex-structured architecture because it offers a very flexible and general likelihood structure.

The basic SCR model and likelihood

SCR models assume that a population of N individuals each have an associated spatial location which represents the home range or activity center of the individual, represented by the geographic coordinates $\mathbf{s}_i = (s_{X,i}, s_{Y,i})$. The collection of such activity centers $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$ is regarded as a realization of a statistical spatial point process (Illian et al. 2008).

Table 2. A description of the various models fitted, and data structures needed, in `oSCR`. The model formulation and data structures follow the model description provided in Royle et al. (2018).

Par	Formulation	Description and data structures
p_0	$g(p_0) = \alpha_0 + \sum_k \alpha_k X_k$	*Baseline encounter model* (p ~ X). This model describes the probability (g=logit) or rate (g=log) of detection at distance=0 and can vary by sex, session, and both constant or time-varying trap-level covariates. Detections are stored as Encounter Data Files (EDF) documenting individual, session, trap, and occasion ID, and if sex if recorded. Trap information such as coordinates and covariates are stored as session-specific Trap Deployment Files (TDF). `data2oscr()` links the EDF and TDFs by trap location to generate the `scrFrame`.
σ	$\log(\sigma) = \gamma_0 + \sum_k \gamma_k X_k$	*Spatial decay model* (sig ~ X). This model represents an explicit source of heterogeneity related to the distance between individual activity centers and traps. The parameter can vary by sex, session, or as a function of session-specific covariates added to the `sigCovs` (a session-by-covariate data frame) in the `scrFrame`.
$d(s)$	$\log(d(s)) = \beta_0 + \sum_k \beta_k X_k$	*Spatial point process model* (D ~ X). The default homogeneous point process model assumes density in each pixel of the state space (`ssDF`, see Table 1) is the same. Spatially heterogeneous density surfaces can be modelled using an inhomogeneous point process model that relates variation in density to spatially referenced (pixel-specific) covariates. Spatial covariates can be added directly to the `ssDF` as named columns that can by referenced in the model formula in the typical way (e.g. D ~ X for covariate X).
δ	$\text{cost}(\upsilon_a, \upsilon_{a+1}) = \dfrac{e^{\delta z(\upsilon_a)} + e^{\delta z(\upsilon_{a+1})}}{2}$	*Cost function model* (asu ~ X). Estimation of resistance parameters (δ) provides a direct measure of the strength of species–landscape interactions allowing flexible (asymmetric and non-stationary) detection functions that depend on activity center location and the surrounding landscape. The cost function model is not restricted to a single landscape covariate and follows standard model syntax. To fit these models, a second spatial data frame, the `costDF`, must be provided. The `costDF` must have the same spatial extent, and at least the same resolution (but preferably finer) as the `ssDF`, and the argument distmet="ecol" passed to `oSCR.fit()`. Covariates of interest are added as named columns in the `costDF`.

The simplest point process assumes that the activity centers are distributed uniformly in space, denoted by

$$\mathbf{s}_i \sim \text{Uniform}(\mathcal{S})$$

where $\mathcal{S}$ is the state space of the point process (this is the area for which density and abundance will be estimated). This can be thought of as a prior distribution for possible locations of activity centers of individuals that might be captured by the sampling effort. The point process is perhaps the most important element of SCR models as it accommodates many elements of spatial population ecology. Moreover, the state space of the point process is a critical element of the SCR model, as it induces an explicit linkage between density and population size as well as an explicit definition of the effective area sampled (Royle et al. 2018). The uniformity assumption corresponds to a type of point process model called a homogeneous point process in which the density of points is spatially constant. Under this model, estimation is of either the total number of points N or the expected value $E(D) = N/A$ where D is density and A is the area of the state space. One important generalization of this model allows for the possibility that density depends on one or more spatially referenced covariates. For example, if $z(\mathbf{s})$ is a covariate (e.g. habitat structure) at location s, then the inhomogeneous point process model posits that

$$\text{Pr}(\mathbf{s}) \propto \exp(\beta_1 z(\mathbf{s}))$$

where the parameter β_1, to be estimated, affects the relative density of points along a gradient of $z(\mathbf{s})$.

The spatial point process describes the spatial structure of the population to be sampled. With the introduction of this point process, the capture–recapture framework can be extended directly to include a model for the probability of detection of an individual. The probability of detection at trap $\mathbf{x}_j$ for $j = 1, 2, \ldots, J$ can be described as a function of Euclidean distance between the individual's activity center, $\mathbf{s}$, and $\mathbf{x}$. A common formulation for describing this conditional-on-$\mathbf{s}$ detection probability is the half-normal encounter probability model:

$$p(\mathbf{x}, \mathbf{s}) = p_0 \exp\left(-\|\mathbf{x} - \mathbf{s}\|^2 / (2\sigma^2)\right)$$

This model introduces two key parameters in the SCR model: the baseline encounter probability p_0, which is the probability that an individual with activity center $\mathbf{s}$ is detected at location $\mathbf{x} = \mathbf{s}$, and the parameter σ, which controls the rate of decrease in detection probability as a function of the distance between $\mathbf{x}$ and $\mathbf{s}$. Intuitively, σ relates to the extent of space used by an individual of the species under study. Many functional forms of the detection probability model are possible (Efford 2018), and they may involve different parameters that have slightly different meanings than p_0 and σ of

the half-normal model, but these details are minor aspects of the SCR modeling framework. oSCR implements two detection functions – the half-normal model and the two-parameter hazard model where σ is, again, the spatial scale parameter but p_0 is the detection rate, as well as variations of these that involve non-Euclidean distance. Others could be easily added by the user if required.

Maximum likelihood estimation in SCR models is based on the marginal likelihood in which we remove the latent variable $\mathbf{s}$ from the conditional-on-$\mathbf{s}$ likelihood by averaging (or marginalizing) over the possible values of $\mathbf{s}$. Therefore, we first identify the conditional-on-$\mathbf{s}$ likelihood. The model for encounter observation y_{ijk} for individual i, at trap j, on occasion k, conditional on $\mathbf{s}_i$, is

$$y_{ijk} \mid \mathbf{s}_i \sim \text{Bern}(p(\mathbf{x}, \mathbf{s}_i; \theta)) \tag{1}$$

where we have indicated the dependence of encounter probability, p, on $\mathbf{s}$ and parameters $\theta = (p_0, \sigma)$ explicitly. The joint distribution of the data for individual i is the product of $J \times K$ such terms (i.e. contributions from each of J traps and K occasions):

$$[\mathbf{y}_i \mid \mathbf{s}_i, \theta] = \prod_{k=1}^{K} \prod_{j=1}^{J} \text{Bern}(p(\mathbf{x}_j, \mathbf{s}_i; \theta))$$

We note this assumes that encounter of individual i in each trap is independent of encounters in every other trap, conditional on $\mathbf{s}_i$. This is the fundamental property of models implemented in oSCR. The marginal likelihood is computed by removing $\mathbf{s}_i$ by integration from the conditional-on-$\mathbf{s}$ likelihood, so we compute the marginal probability mass function (pmf) for each encounter history $\mathbf{y}_i$:

$$[\mathbf{y}_i \mid \theta] = \int_{\mathcal{S}} [\mathbf{y}_i \mid \mathbf{s}, \theta][\mathbf{s}] d\mathbf{s}$$

In the default SCR model, we assume the activity centers $\mathbf{s}$ are uniformly distributed throughout the state space, which is to say: $[\mathbf{s}] = 1/A(\mathcal{S})$ where $A(\mathcal{S})$ is the area of the prescribed state space $\mathcal{S}$. But, in general, oSCR accommodates spatial variation in density as noted below. In oSCR we compute this marginal likelihood by evaluating the integrand on a regular grid of points of equal area and computing the average of the integrand over that grid of points. This grid of points defines the state space object (see Methods and features) which must be specified by the user. Let $u = 1, 2, \ldots, G$ index a grid of G points, $\mathbf{s}_u$, where the area of grid cells is constant. In this case, the marginal pmf of $\mathbf{y}_i$, is approximated by

$$[\mathbf{y}_i \mid \theta] = \frac{1}{G} \sum_{u=1}^{G} [\mathbf{y}_i \mid \mathbf{s}_u, \theta] \tag{2}$$

The joint likelihood for the data from n observed individuals, assuming independence of encounters among and between individuals, is the product of n such terms and, in addition, a contribution of the $n_0 = N - n$ uncaptured individuals.

Obviously each of these all-zero encounter histories will have the same marginal pmf contribution in the likelihood given above, and we denote that marginal probability by π_0. The only question is, how many? We therefore include the number of such all-zero encounter histories (that is, the number of individuals not encountered) as an unknown parameter of the model. With n_0 unknown, we have to be sure to include a combinatorial term to account for the fact that, of the n observed individuals, there are $\binom{N}{n}$ ways to realize a sample of size n. Therefore, the joint likelihood has this form:

$$\mathcal{L}(\theta, n_0 \mid \mathbf{y}) = \frac{N!}{n! n_0!} \left\{ \prod_{i=1}^{n} [\mathbf{y}_i \mid \theta] \right\} \pi_0^{n_0}. \tag{3}$$

This is discussed in Borchers and Efford (2008) as the conditional-on-N form of the likelihood, and referred to by Royle et al. (2013a) as the 'binomial-form' due to its origin as arising from an assumption that n is a binomial outcome from a population of size N (and also it is 'binomial looking').

The binomial-form of the likelihood just described can be analyzed directly in oSCR (option `DorN == "N"`) and, in this case, $\log(n_0)$ is estimated along with parameters p_0 and σ and other covariates of the models as described in Table 2. Important variations of this model accommodated in oSCR modify the binomial observation model to allow for multiple detections during an occasion – a Poisson observation model (`encmod = "P"`), and a multinomial or multi-catch observation model (`encmod = "M"`). For these observation models, there are no additional considerations in building the likelihood, only in replacing the Bernoulli observation model with an appropriate alternative.

Poisson-integrated likelihood

For likelihood analysis it is convenient to remove N from the likelihood by putting a Poisson prior distribution on it. That is, we assume $N \sim \text{Poisson}(D \times A(\mathcal{S}))$ where $A(\mathcal{S})$ is the area of the state space and D is the density of activity centers. Under this prior, we can remove N from the binomial-form of the likelihood, again by summation over the possible values of N. We call this the Poisson-integrated likelihood which is implemented in oSCR (as the default with option `DorN = "D"`) and also is the default likelihood form in secr (Efford 2018). To compute the Poisson-integrated likelihood we do a further level of marginalization over the Poisson prior distribution:

$$\sum_{n_0=0}^{\infty} \frac{N!}{n! n_0!} \left\{ \prod_i [\mathbf{y}_i \mid \alpha] \right\} \pi_0^{n_0} \frac{\exp(-\Lambda) \Lambda^N}{N!}$$

where $\Lambda = D \times A(\mathcal{S})$. Carrying out the summation produces:

$$\mathcal{L}(\alpha, \Lambda \mid \mathbf{y}) = \left\{ \prod_i [\mathbf{y}_i \mid \alpha] \right\} \Lambda^n \exp(-\Lambda(1 - \pi_0)) \tag{4}$$

We emphasize there are two marginalizations involved in the formation of this likelihood: 1) the integration to remove the latent variables $\mathbf{s}$; and, 2) summation to remove the parameter N.

A more general expression for the Poisson-integrated likelihood recognizes that the Poisson assumption for N implies that, on a discrete state space, the number of activity centers in a state-space pixel, $N(\mathbf{s})$, is also Poisson with mean $D(\mathbf{s})A(\mathbf{s})$ where $A(\mathbf{s})$ is the area of the state-space pixel having center coordinate $\mathbf{s}$ and $D(\mathbf{s})$ is the pixel density. This is the property of compound additivity which defines a Poisson point process and allows for specifying explicit models for the distribution of individuals across homogeneous or heterogeneous landscapes. In particular, if we suppose that $D(\mathbf{s})$ depends on measurable covariates at the scale of the state-space resolution (i.e. pixel size), e.g.

$$\log(D(\mathbf{s})) = \beta_0 + \beta_1 z(\mathbf{s})$$

then the conditional intensity function for the n observed activity centers can be expressed as:

$$\text{Pr}(\mathbf{s}) = \frac{D(\mathbf{s})}{\sum_u D(\mathbf{s}_u)}$$

Then, the marginalization operation of each observed encounter history is

$$[\mathbf{y}_i \mid \theta, \beta] = \sum_{u=1}^{G} [\mathbf{y}_i \mid \mathbf{s}_u, \theta] \text{Pr}(\mathbf{s}_u \mid \beta) \tag{5}$$

and $\Lambda = D \times A(\mathcal{S})$ is replaced with $\sum_{u=1}^{G} D(\mathbf{s}_u)A(\mathbf{s}_u)$ in Eq. 4.

These considerations of formulating the binomial-form of the likelihood and then translating it to the Poisson-integrated form define a general class of SCR models which can accommodate Bernoulli, Poisson or multinomial encounters and allow for the modeling of covariates on the various elemental parameters p_0, σ and $D(\mathbf{s})$. The scope of these models is described in Table 2. We now define two additional important structural elements that define the core likelihood structure implemented in oSCR.

Class or sex-structured likelihood

In most studies we record some information about the sex of individuals which might be used in developing sex-structured models in which one or more parameters depend on sex. Moreover, it is common in practice to have some missing values of sex due to DNA samples not amplifying, poor photo quality, or other considerations. To build models with sex specificity of parameters, we require a different form of the likelihood which accommodates observed sex data; such data are informative about the sex ratio of the population and, hence, relative densities of males and females (Royle et al. 2015). This new form of the likelihood is used by default if sex information is included in the core oSCR data object,

the `scrFrame`, and thus, all models fit using the same data object have comparable likelihoods, and AIC-based model comparisons can be used. The form of sex-structured model described here is implemented in the `oSCR` package.

In general, let there be n_1 captured individuals with sex data, n_2 captured individuals with missing values, and $n_0 = N - n_1 - n_2$ undetected individuals. Let c_i indicate the sex of individual i, having possible values $c \in (0, 1)$ with probabilities $(1 - \psi, \psi)$. The likelihood for individual encounter histories which have missing sex requires one additional level of marginalization. In addition to removing **s** from the conditional likelihood we now have to remove the missing sex variable from the conditional-on-sex likelihood to compute $[\mathbf{y}_i | \theta, c_i]$ by averaging the two sex-specific components:

$$\left[\mathbf{y}_i \mid \theta \right] = \left[\mathbf{y}_i \mid \theta, c_i = 0\right](1 - \psi) + \left[\mathbf{y}_i \mid \theta, c_i = 1\right]\psi$$

There are n_2 such contributions to the likelihood. In addition, each of the individuals with observed sex contribute a component of this form: $\left[\mathbf{y}_i \mid \theta, c_i\right]\psi^{c_i}(1 - \psi)^{1 - c_i}$. Therefore, the binomial-form of the likelihood when sex information is available for at least some individuals has the form:

$$\mathcal{L}\left(p, \psi, n_0 \mid \mathbf{y}\right) = \frac{N!}{n! n_0!}\left\{\prod_{i=1}^{n_1}\left[\mathbf{y}_i \mid c_i = 1, \theta\right]\psi^{c_i}(1 - \psi)^{1 - c_i}\right\}$$
$$\left\{\prod_{i=1}^{n_2}[\mathbf{y}_i \mid \theta]\right\}\pi_0^{n_0}$$

where $n = n_1 + n_2$ and π_0 is the marginal probability of the 'all-zero' encounter history, averaged over all possible groups. Remember, for the SCR model, the elemental part of the likelihood is the marginal likelihood of an individual's encounter history, averaged over possible values of **s**. For encounter histories which have missing sex data, we then have to do an additional marginalization over possible values of the sex variable producing the $[\mathbf{y}_i|\theta]$ term above.

Multi-session structure

A common situation in many studies is to collect data on distinct more-or-less independent populations. For example, a camera trapping study might take place in multiple reserves or use multiple camera trap grids spaced far apart. Or sampling for amphibians using artificial cover objects (ACOs) involves replicated ACO arrays in different habitats (Sutherland et al. 2016, Schmidt et al. 2017). Replication might occur over time as well – sampling in different seasons or years. It is important to be able to combine the data from such replicated studies into a single statistical model in order to increase in statistical power, and also to test effects which might represent important contrasts among the sampled populations. In practice, these populations are called 'sessions', groups, or strata and the model which integrates the data from different sessions is referred to as a multi-session model. The Poisson-integrated likelihood previously introduced applies directly to this multi-session structure. Let N_g be the population size for group g, then the multi-session model assumes:

$$N_g \sim \text{Poisson}(\lambda_g)$$

where the N_g are mutually independent random variables. The basic likelihood described above can then be computed by independently marginalizing over this prior distribution for N_g for the data from each group to obtain the marginal likelihood as a function of λ_g. This affords enormous flexibility in modeling because sessions can be spatial, temporal, indexed by other factors (e.g. species, age class), or combinations of multiple stratification variables. Moreover, explicit models can be formulated for λ_g. For example, suppose the stratification variable represents a unique trapping grid in a small mammal study (Converse et al. 2009):

$$\log(\lambda_g) = \beta_0 + \beta_1 \text{Trt}_g + \beta_2 \text{Habitat}_g$$

and then the data from all groups are simultaneously used to estimate parameters of the SCR detection model and the information among groups is used to estimate the parameters describing variation in λ_g.

A key assumption of the multi-session model is independence of the observed encounter histories among groups and also independence of the population size random variables N_g. One implication of this is that it precludes individuals from appearing in multiple groups, which might happen if the trapping arrays are close together or if the groups represent the same population sampled over time. The effect of ignoring this non-independence should not produce bias in parameter estimates, but will likely lead to an over-statement of precision. Simulations would be required to understand the magnitude by which precision is overstated. In addition, the independence assumption is probably invalid if some of the group structure represents sub-groups within populations. For example, male and female sub-populations of multiple groups, or age classes, or species sub-groups. In such cases, we might expect population size of these sub-groups to covary.

Advanced models and features

Ecological distance model

Using a least cost path approach, SCR allows the estimation of one or more resistance parameters, δ, that quantify how movement is influenced by local landscape structure. The model replaces Euclidean distance with 'ecological' distance (Royle et al. 2013b) and therefore relaxes the restrictive assumptions of symmetrical and stationary space use regardless of the surrounding habitat (Sutherland et al. 2014, 2018, Fuller et al. 2016). Estimating parameters of the cost function requires specification of a fourth model in the `oSCR` model call, and a second state–space-like spatially referenced

data frame, a `costDF`, that has the coordinates of the pixel centroids and associated pixel specific covariates (Table 2, code supplement in Sutherland et al. 2018).

Telemetry/RSF model

The integration of telemetry data can allow for improved parameter estimation by informing the σ parameter and/or supporting a resource selection function (RSF) as part of the encounter model (Royle et al. 2013a, Tenan et al. 2017, Linden et al. 2018). Under the assumption that individual encounters at traps are a 'thinned' version of telemetry fixes (and that telemetered individuals are representative of the population targeted by trapping), certain parameters can be shared in the likelihoods for each data source. The original model formulation by Royle et al. (2013a) involved independent data, such that telemetered individuals did not appear in the spatial encounter histories generated by the trap sampling. Linden et al. (2018) described the marginal likelihood for situations where telemetered individuals were also encountered at traps, allowing each data source to be informative about the latent activity centers for relevant individuals. oSCR allows both independent and dependent telemetry data and various options for the degree of telemetry integration. In all cases, fixes need to be summarized as pixel counts within the state space (with functions provided to achieve this) and RSF integration requires an additional data frame, `rsfDF`, similar to the `costDF`, with a resolution matching the state space (Table 2, code available in Linden et al. 2018). We note that, currently, oSCR cannot model serial correlation in telemetry fixes and assumes that fix locations are independent across time.

Describing the oSCR workflow

Here we outline a typical oSCR analysis and pair the workflow with an example using redbacked salamander (RBS) data (Sutherland et al. 2016). A complete annotated workflow is provided in the Supplementary vignette Appendix 1, and here we highlight the data objects and functions required to conduct an SCR analysis in oSCR (Table 1).

Every spatial capture–recapture study gives rise to at least two data structures around which we have built the data processing step of the oSCR workflow. The first data object is a 'Trap Deployment File' (TDF) containing, at minimum, the name and coordinates of each trap or detector. In addition, the TDF can contain trap-by-occasion binary operation data (1 = operational, 0 = not operational), and, if available, trap-specific covariates that may or may not vary by occasion. If all detectors are operational in all occasions, there is no need to supply trap operation data. Trap characteristics (name, location and operation data) and covariates are separated by a column of '\'s. Because the number of occasions can vary between sessions, it is natural to have separate TDFs for each session which is required for oSCR.

The second data source is the 'Encounter Data File' (EDF) which contains the individual encounter history data. This has, at a minimum, columns containing the unique individual identifier, the name of the trap in which it was detected, the occasion during which the encounter occurred, and a numeric session identifier. If the sex of the species can be determined, this information can also be included as an additional column which may include missing values. Unlike the TDF, and because there is a session identifier, encounter data from all sessions are included in a single EDF. The names of the traps in the EDF must match the names in the associated TDFs. The sex column should contain only `'M'` for males, `'F'` for females, and an unknown identifier (we use `'U'`) for unknowns.

Once these objects have been read into R, the oSCR workflow begins with the creation of an `scrFrame`, a data object that links the EDF and TDFs, checks for errors, and formats the data for analysis. The `scrFrame` is created using the helper function `data2oscr()` and contains the following data objects:

- `caphist` – session-specific list of individual-by-trap-by-occasion encounter history arrays. Array can be binary or contain capture frequencies
- `traps` – session-specific list of trap coordinates
- `trapCovs` – session-specific list of trap-by-covariate data frames containing trap specific covariate values
- `trapOperation` – session-specific trap-by-occasion binary matrix denoting whether traps were operational
- `indCovs` – session-specific list of individual covariates. These data frames contain the sex of the individual if sex information is provided
- `sigCovs` – session-by-covariate data frame of covariates

Typing the `scrFrame` object name produces summaries of the trapping effort and detections by session (number of individuals: `n individuals`, traps: `n traps`, and occasions: `n occasions`), and of the capture histories (average number of times an individual is encounters: `avg caps`, average number of spatial locations individuals are encountered at: `avg spatial caps`, the mean maximum distance moved: `mmdm`, and the mmdm pooled across sessions). The spatial encounter history data can also be summarized graphically using `plot(scrFrame)` (Fig. 1).

A key, but often under-appreciated, part of an SCR analysis is the definition of the state space ($\mathcal{S}$). In oSCR, a list containing a state space for each session is required. We make the state space definition an explicit part of the workflow via the function `make.ssDF()`. The function requires an `scrFrame` (which contains the session-specific trap coordinates), a value specifying the buffering distance around the traps, and a resolution value that determines the spacing of the state–space centroids. The units of the buffer and resolution should be in the same units as the traps, which themselves should be in Cartesian coordinates (e.g. utm, not lat long). From experience, a general rule of thumb is a buffer distance of 3–4× the estimated σ from the half normal
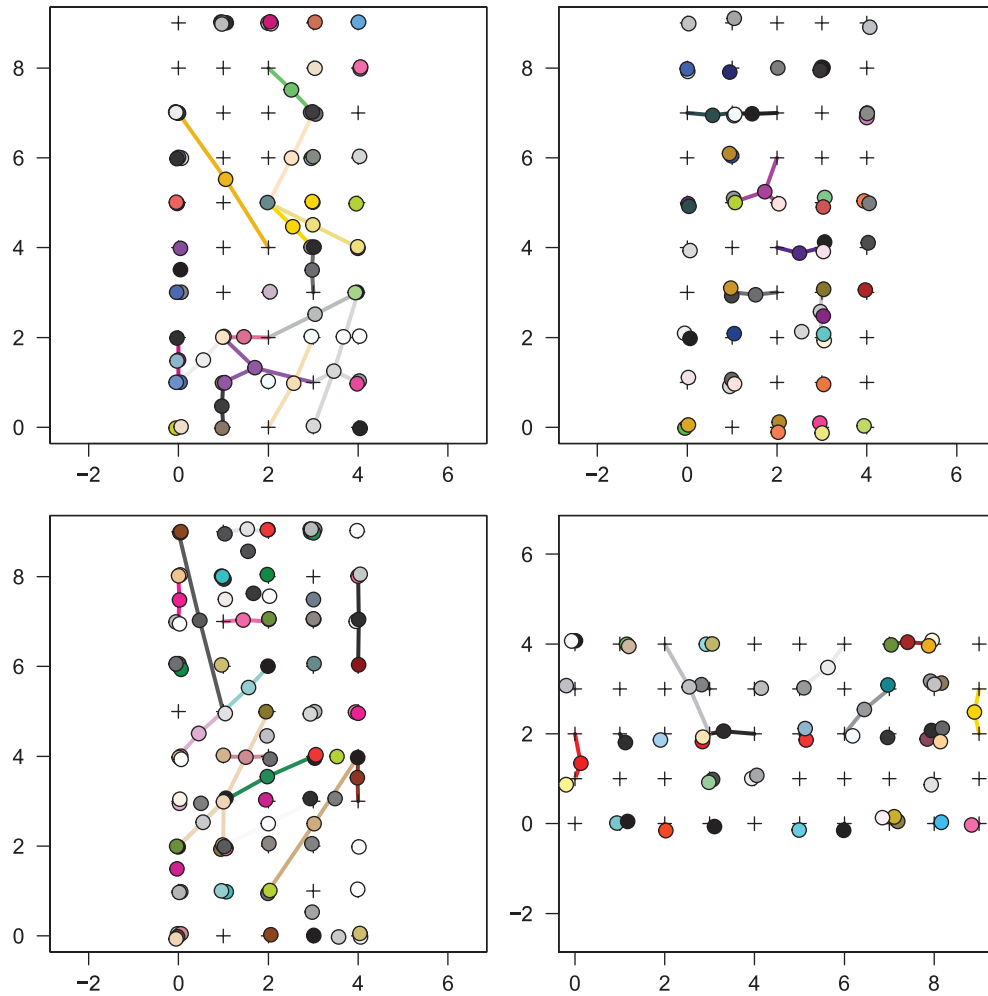
Figure 1. Visualization of the spatially explicit individual encounter histories for the four session salamander SCR data. Each filled circle represents the spatial average of all detections of an individual and lines join average locations to the traps in which individuals were captured. Each unique color (circle and lines) is a unique individual. Crosses (+) are trap locations, and circles without lines are individuals that were detected at only a single location.

encounter model, and a resolution of $0.5 \times \sigma$ (see also Efford 2018). The state space can also be summarized graphically using `plot()` (Fig. 2). Of course, σ is not known a priori, so it is recommended that a range of buffer and resolution values are tested to confirm that parameter estimates are insensitive to the choice of values.

The RBS study system is a set of four independent artificial cover board arrays set in a small woodland in Ithaca, New York. Each array consists of 50 small (25×25 cm) cover boards (detectors) arranged in a 5×10 m rectangle with each board separated by 1 m (Fig. 1, 2, Snippet 1). Because each array is independent we define each as an independent 'session' (session 1–4). Arrays were visited 7, 5, 6 and 4 times, respectively, from September to November in 2014 (Snippet 1). During each visit (or occasion), every board was checked, and unmarked salamanders were given unique individual marks by injecting visual implant elastomer (Grant 2008). Individual spatial encounter histories were generated from the collection of the location, occasion, and identity of every

individual detected. See Sutherland et al. (2016) for a full description.

With the `scrFrame` and the `ssDF` created, we can move to model fitting (step 2, Table 1). The main fitting function in `oSCR` is `oSCR.fit()`. The fitting function requires, at minimum, an `scrFrame`, an `ssDF`, and a model list that specifies the model structure in standard R model formula syntax for each of the three model components (Table 1). The list contains, in this order, a model formula for density (D~), baseline detection (p~), and the spatial scale of detection (sig~). An example of three RBS models is shown in Snippet 2. The base model has constant density and space use, and, based on what we know about RBS activity patterns, a detection model with a quadratic effect of day of year and a behavioural response. The full model has session-specific density, session- and sex-specific space use, and, in addition to a quadratic effect and behavioural response, detection varying by session and sex. For the remainder of this example, we use the model with the
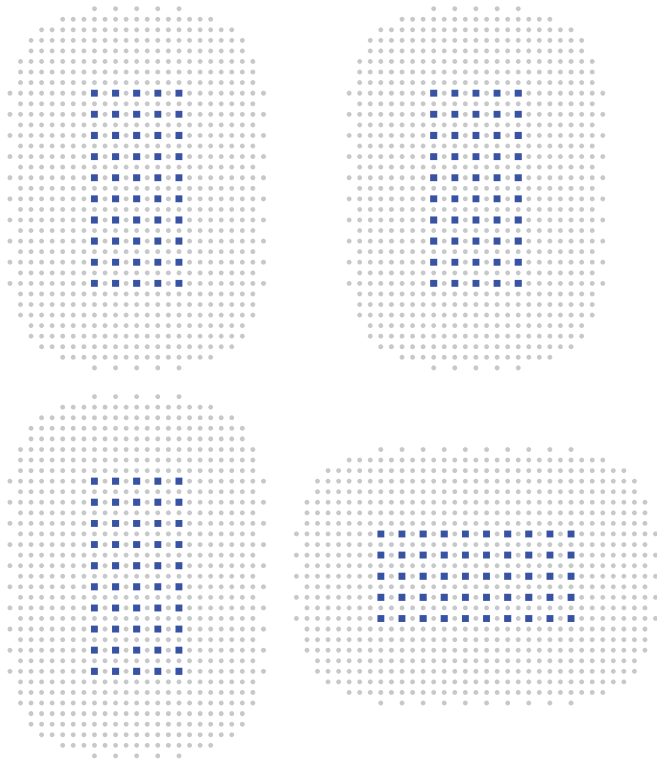
Figure 2. Visualization of the prescribed state spaces for the the four salamander sites. Blue squares are the locations of the artificial cover objects (ACOs) and grey points are the pixel centroids which are the discrete representation of sampling area. Here, ACOs are 1 m apart, the state space resolution is 0.5 m, and the buffer is 4 m.

lowest AIC as the top model (Supplementary vignette Appendix 1).

The function `oSCR.fit()` has many other arguments that can be used to control the nature of the model fitting routine including, but not limited to: a choice of binomial (`encmod="B"`), Poisson (`encmod="P"`) or multinomial (`encmod="M"`) encounter models (Royle et al. 2014); asymmetric space use models (Royle et al. 2013b, Sutherland et al. 2014); resource selection functions (Royle et al. 2013a); integration of telemetry data (Tenan et al. 2017, Linden et al. 2018); and 'local-to-capture' likelihood evaluation (Milleret et al. 2018).

```
rbs.sf

                   S1 S2  S3 S4
n individuals 77 60 108 54
n traps            50 50  50 50
n occasions         7  5   6  4

                   S1   S2   S3   S4
avg caps         1.91 1.47 1.71 1.37
avg spatial caps 1.30 1.15 1.27 1.13
mmdm             2.02 1.05 1.68 1.29

Pooled MMDM:  1.66
```

Snippet 1. Summary statistics for the RBS `scrFrame`.

It is common in observational studies in ecology to develop and compare multiple competing models, typically using AIC. This is possible in `oSCR` in what we call the 'model selection' stage of the workflow. In order to formally compare models, a list of fitted models must be created using the function `fitList.oSCR()`. Ranked AIC model tables can be produced by passing the `fitList.oSCR` object to the `modSel.oSCR()` function. The model selection function returns three objects of interest: a model table ranked by AIC (lowest first), a model × parameter matrix of maximum likelihood estimates, and a model × parameter matrix of standard errors. While the model table is perhaps of most interest, the coefficient and standard error matrices are useful for model averaging. Currently there is no helper function for generating model averaged predictions, model averaged coefficients can be computed using `ma.coef()`.

Finally, making predictions from specific models can be done using `get.real()`, a function that is similar to the `predict()` functions used on `unmarked` or `glm` model objects. The function takes the model object, an argument identifying which model component to predict from, e.g. density (`type = 'dens'`), detection (`type = 'det'`), or sigma (`type = 'sig'`), and a `newdata` object that is a data frame used for making predictions. The `newdata` object should contain columns for all covariates in the model component, including a sex column if sex is included in the `scrFrame`. Figure 3 shows the predicted relationships for detection (top), density (middle row), and space use (bottom row) based on the 'top' RBS model (Supplementary vignette Appendix 1).

oSCR's place in the software landscape

In this software note, we have outlined the main features and functionality of the R package `oSCR`. We emphasize the accessible and intuitive nature of the code base and workflow and hope this stimulates active participation in further technical developments. We do, however, note that `oSCR` represents one of several options for analyzing various classes of spatial capture–recapture models, each offering unique capabilities. For example: `secr`, a comprehensive general purpose package for fitting SCR models (Efford 2019); `ascr`, a package for analyzing spatial encounters generated from acoustic data (Stevenson 2018); `openpopscr`, a package for fitting open population SCR models by maximum likelihood (Glennie et al. 2019); and its Bayesian analogue `OPenPopSCR` (Augustine 2019). We have attempted, as much as possible, to maintain consistency in data formatting and terminology with the suite of available software options to avoid confusion, and importantly, to ensure complementarity.

`oSCR` therefore represents an accessible and intuitive framework, and accompanying workflow, for analyzing spatial encounter history data that places a specific focus the inherent structure that exists both within populations (e.g. sex, or any other binary class) and between populations that

```
#RBS model definition
base_mod <- list(D~1,        p0~b+jday+jday2,              sig~1)
full_mod <- list(D~session, p0~b+jday+jday2+session+sex, sig~session+sex)
top_mod  <- list(D~session, p0~b+jday+jday2,              sig~session)

#RBS model fitting
oSCR.fit(model = base_model, scrFrame = rbs.sf, ssDF = rbs.ss)
oSCR.fit(model = full_model, scrFrame = rbs.sf, ssDF = rbs.ss)
oSCR.fit(model = top_model,  scrFrame = rbs.sf, ssDF = rbs.ss)
```

Snippet 2. Calls to oSCR.fit() for three competing RBS models.
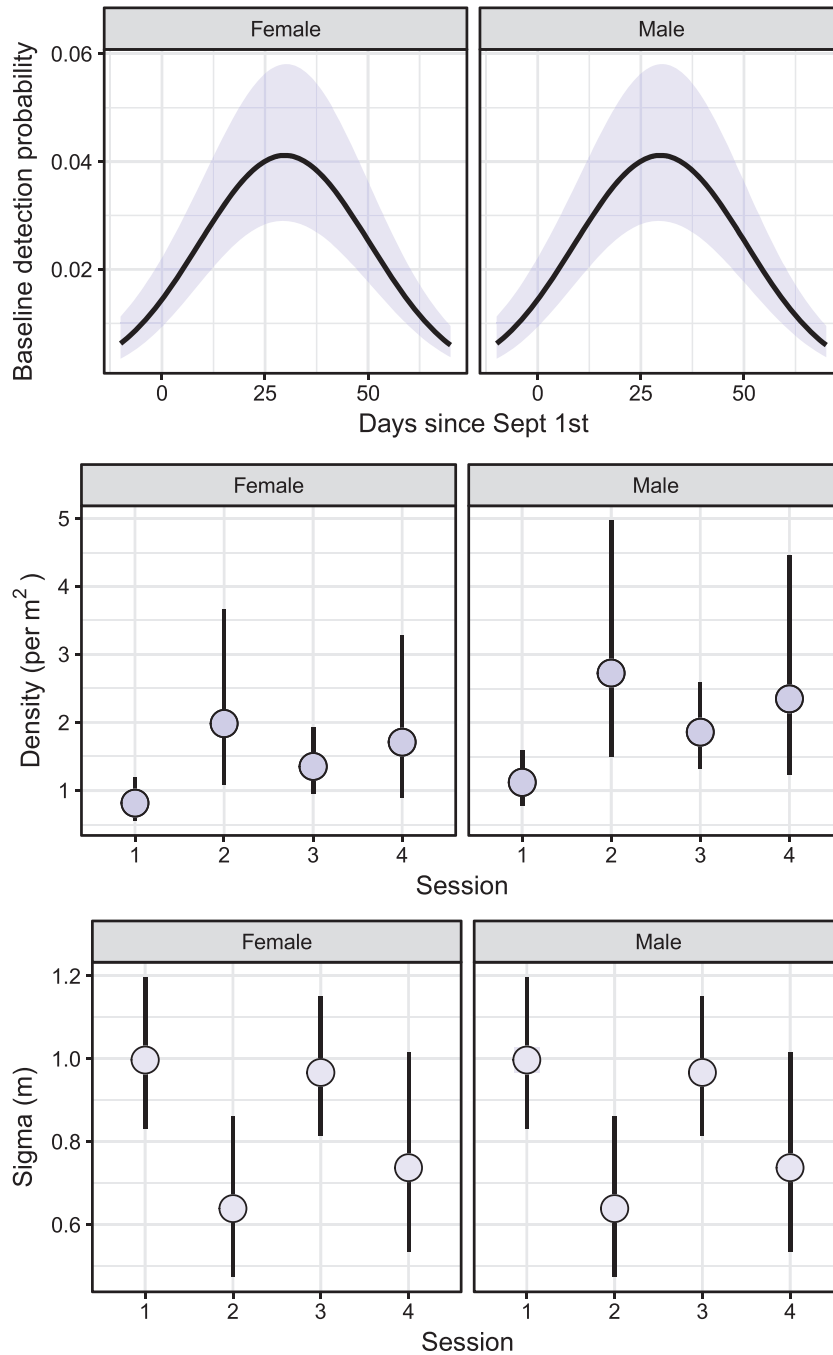
Figure 3. Predicted relationship between the probability of first detection and a quadratic effect of relative day of the year (top), session- and sex-specific density, and session- and sex-specific sigma.

have some level of spatial or temporal independence (e.g. session). Importantly, oSCR is set up naturally to quantify the importance of such structure while simultaneously estimating density, detection and space use.

oSCR is free and open source, available on GitHub (<https://github.com/jaroyle/oSCR>). To cite oSCR or acknowledge its use, cite this software note including the appropriate version number:

Sutherland, C., Royle, J. A. and Linden, D. W. 2019. 'oSCR'. – Ecography 42: 1459–1469 (ver. X.X).

References

Augustine, B. 2019. Openpopscr: open population spatial capture–recapture mcmc algorithms. – <https://rdrr.io/github/benaug/OpenPopSCR/>.

Borchers, D. L. and Efford, M. G. 2008. Spatially explicit maximum likelihood methods for capture–recapture studies. – Biometrics 64: 377–385.

Borchers, D. et al. 2016. Spatial capture–recapture models. – Stat. Sci. 31: 219–232.

Converse, S. et al. 2009. Multistate models for estimation of survival and reproduction in the grey-headed albatross (*Thalassarche chrysostoma*). – Auk 126: 77–88.

de Valpine, P. et al. 2017. Programming with models: writing statistical algorithms for general model structures with nimble. – J. Comp. Graph. Stat. 26: 403–413.

Efford, M. 2019. secr: spatially explicit capture–recapture models. – R package ver. 3.2.1. <https://CRAN.R-project.org/package=secr>.

Efford, M. G. 2004. Density estimation in live trapping studies. – Oikos 106: 598–610.

Efford, M. G. et al. 2004. Density: software for analysing capture–recapture data from passive detector arrays. – Anim. Biodivers. Conserv. 27: 217–228.

Fuller, A. K. et al. 2016. Estimating population density and connectivity of American mink using spatial capture–recapture. – Ecol. Appl. 26: 1125–1135.

Glennie, R. et al. 2018. Open population maximum likelihood spatial capture–recapture. – Biometrics <https://doi.org/10.1111/biom.13078>.

Grant, E. H. C. 2008. Visual implant elastomer mark retention through metamorphosis in amphibian larvae. – J. Wildl. Manage. 72: 1247–1252.

Kéry, M. 2010. Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses. – Academic Press.

Illian, J. et al. 2008. Statistical analysis and modelling of spatial point patterns (Vol. 70). – Wiley.

Linden, D. W. et al. 2018. Integrating telemetry data into spatial capture–recapture modifies inferences on multi-scale resource selection. – Ecosphere 9: e02203.

Long, R. A. 2008. Noninvasive survey methods for carnivores. – Island Press.

Lunn, D. et al. 2009. The BUGS project: evolution, critique and future directions. – Stat. Med. 28: 3049–3067.

Milleret, C. et al. 2018. A local evaluation of the individual state-space to scale up Bayesian spatial capture–recapture. – Ecol. Evol. 9: 352–363.

Royle, J. A. and Dorazio, R. M. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. – Academic Press.

Royle, J. A. and Young, K. 2008. A hierarchical model for spatial capture–recapture data. – Ecology 89: 2281–2289.

Royle, J. A. et al. 2013a. Integrating resource selection information with spatial capture–recapture. – Methods Ecol. Evol. 4: 520–530.

Royle, J. A. et al. 2013b. Spatial capture–recapture models for jointly estimating population density and landscape connectivity. – Ecology 94: 287–294.

Royle, J. A. et al. 2014. Spatial capture–recapture. – Elsevier.

Royle, J. A. et al. 2015. Likelihood analysis of spatial capture–recapture models for stratified or class structured populations. – Ecosphere 6: art22.

Royle, J. A. et al. 2018. Unifying population and landscape ecology with spatial capture–recapture. – Ecography 41: 444–456.

Schmidt, B. R. et al. 2017. Spatial capture–recapture analysis of artificial cover board survey data reveals small scale spatial variation in slow-worm Anguis fragilis density. – R. Soc. Open Sci. 4: 170374.

Stevenson, B. 2018. ascr: an r package for acoustic spatial capture–recapture using ad model builder. – R package ver. 2.2.0. <https://github.com/b-steve/ascr>.

Sutherland, C. et al. 2014. Modelling non-Euclidean movement and landscape connectivity in highly structured ecological networks. – Methods Ecol. Evol. 6: 169–177.

Sutherland, C. et al. 2016. Spatial capture–recapture: a promising method for analyzing data collected using artificial cover objects. – Herpetologica 72: 6–12.

Sutherland, C. et al. 2018. Large-scale variation in density of an aquatic ecosystem indicator species. – Sci. Rep. 8: 8958.

Tenan, S. et al. 2017. Data integration for inference about spatial processes: a model-based approach to test and account for data inconsistency. – PLoS One 12: e0185588.

Supplementary material (available online as Appendix ecog-04551 at <www.ecography.org/appendix/ecog-04551>). Appendix 1.