oSCR - an R package for SCR analyses

Chris Sutherland
University of Massachusetts — Amherst
csutherland@umass.edu

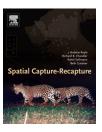
Analyzing SPARCnet data

What is oSCR?

oscr o opensource SCR

- built upon various functions in scrbook
- an editable and evolving code base
- focus on increased transparency and accessibility
- a community of contributors
 - contributing ideas
 - contributing problems
 - contributing solutions
 - contributing code





Where is oSCR?

oSCR - getting the package

- hosted on github
- download directly from github for most up-to-date version

```
install.packages("githubinstall")
install.packages("devtools")
library(githubinstall)
library(devtools)
#install oSCR directly from github
install_github("jaroyle/oSCR")
#load oSCR
library(oSCR)
```



- 1. Format the sampling data
 - spatial encounter history data
 - trapping information data
 - □ etc...



- 1. Format the sampling data
- 2. Format the spatial extent data
 - extent of the state space
 - spatial covariates for density
 - state space constraints



- 1. Format the sampling data
- 2. Format the spatial extent data
- 3. Analyze the data data model fitting



- 1. Format the sampling data
- 2. Format the spatial extent data
- 3. Analyze the data data model fitting
- 4. Post processing model output to do inference



Every analysis in oSCR involves the following steps:

- 1. Format the sampling data
- 2. Format the spatial extent data
- 3. Analyze the data data model fitting
- 4. Post processing model output to do inference

Each step has helpful oSCR helper functions

• repeatable, transparent & sharable workflow



Reminder: Spatially explicit observation model

$$y_{ijk}|s_i \sim \mathsf{Bernoulli}(p[x_j,s_i])$$

$$p[x_j, s_i] = p_0 \times \exp\left(-\frac{\operatorname{dist}(x_j, s_i)^2}{2\sigma^2}\right)$$

- data requirements:
 - y_{ijk} observations (binary or frequencies)
 - \square x_i trap locations
- in oSCR the data are organized in an scrFrame

Making the scrFrame from SPARCnet-style data

- 1. the encounter data file edf (a single data frame):
 - a row per detection
 - several key columns with capture information
 - session ID (required)
 - individual ID (required)
 - occasion ID (required)
 - trap ID (required)
 - sex (optional)

- 1. the encounter data file edf (a single data frame):
- a row per detection
- several key columns with capture information

```
# the rbs edf:
rbs.edf <- read.csv("rbsNY.csv",h=T)
head(rbs.edf)
Site Ind Occasion Board Sex
1 P1A BxxYP1A 1 A2 M
2 P1A YxYBP1A 1 A3 M
3 P1A xxBYP1A 1 A6 M
4 P1A YYxBP1A 1 A8 M
5 P1A xBBYP1A 1 B10 U
6 P1A xBxXP1A 1 B5 U
```

- 2. the trap deployment data file tdf (a list):
 - a list containing a data frame for each session
 - a row per trap
 - several key columns with trap information
 - detector name (required)
 - X coordinates (required)
 - ☐ Y coordinates (required)
 - other data typically stored but not required

- 2. the trap deployment data file tdf:
 - a list containing a data frame for each session
 - a row per trap
 - several key columns with trap information

```
# the rbs edf:

tdf1 <- read.csv("tdfNY1.csv",h=T)

tdf2 <- read.csv("tdfNY2.csv",h=T)

tdf3 <- read.csv("tdfNY3.csv",h=T)

tdf4 <- read.csv("tdfNY4.csv",h=T)

head(tdf1)

board x y X1 X2 X3 X4 X5 X6 X7

1 A1 0 0 1 1 1 1 1 1 1 1

2 A2 0 1 1 1 1 1 1 1 1

3 A3 0 2 1 1 1 1 1 1 1 1

4 A4 0 3 1 1 1 1 1 1 1

5 A5 0 4 1 1 1 1 1 1

6 A6 0 5 1 1 1 1 1 1 1
```

- 2. the trap deployment data file tdf:
 - a list containing a data frame for each session
 - a row per trap
 - several key columns with trap information

- 2. the trap deployment data file tdf:
- a list containing a data frame for each session
- a row per trap
- several key columns with trap information

```
# the rbs edf:

tdf1 <- read.csv("tdfNY1.csv",h=T)

tdf2 <- read.csv("tdfNY2.csv",h=T)

tdf3 <- read.csv("tdfNY3.csv",h=T)

tdf4 <- read.csv("tdfNY4.csv",h=T)

head(tdf3)

board x y X1 X2 X3 X4 X5 X6

1 A1 0 0 1 1 1 1 1 1

2 A2 0 1 1 1 1 1 1 1

3 A3 0 2 1 1 1 1 1 1

4 A4 0 3 1 1 1 1 1 1

5 A5 0 4 1 1 1 1 1 1

6 A6 0 5 1 1 1 1 1 1
```

- 2. the trap deployment data file tdf:
 - a list containing a data frame for each session
 - a row per trap
 - several key columns with trap information

Data are naturally entered in two basic spreadsheets:

- 1. the encounter data file edf (a single data frame):
- 2. the trap deployment data file tdf (a list):

Some important points about this data:

- the detector names in edf and tdf MUST match
 - same names
 - □ same class (integer/character/factor)

Making the scrFrame using data2oscr

The data2oscr() function is a very useful helper function

- inputs are 'traditional' data formats
- returns several data objects
 - an scrFrame
 - data formatted for Bayesian analysis

Making the scrFrame using data2oscr

The data2oscr() function is a very useful helper function

```
# create general SCR data objects
data <- data2oscr(edf.
                              # the edf
               tdf.
                   # the tdf
               sess.col, # session col NUMBER (edf)
              id.col, # ind ID col NUMBER (edf)
               occ.col, # occasion col NUMBER (edf)
              trap.col, # detector col NUMBER (edf)
               sex.col, # sex col NUMBER
               sex.nacode, # character for unknown sex?
               Κ.
                         # number of occassions
               ntraps)
                              # the number of traps
# extract the scrFrame
sf <- data$scrFrame
```

Making the scrFrame using data2oscr

```
rbs.sf <- rbs$scrFrame
```

Inspecting the scrFrame

Summary functions for the capture data in the scrFrame

- type object name for a numerical summary: sf
 - □ number of individuals, captures & spatial recaptures
 - mean maximum distance moved (MMDM)
- plot spatial captures: plot(sf)
 - spatial average of locations
 - traps captured in

Inspecting the scrFrame

```
rbs.sf #print a summary

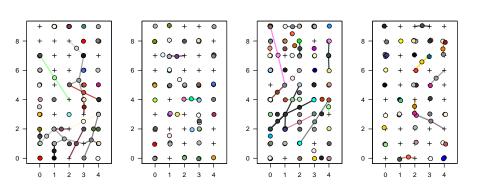
S1 S2 S3 S4
n individuals 77 61 107 54
n traps 50 50 50 50
n occasions 7 5 6 4

S1 S2 S3 S4
avg caps 1.92 1.48 1.73 1.37
avg spatial caps 1.30 1.16 1.27 1.13
mmdm 2.10 1.05 1.68 1.29

Pooled MMDM: 1.66
```

Exploring the scrFrame

plot(rbs.sf) #plot a summary



Reminder: Spatially explicit density model

Spatially explicit density model (homogeneous):

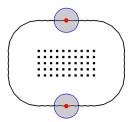
describes how activity centers are distributed in space

$$Pr(s_i) \propto exp(\beta)$$

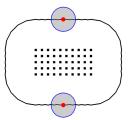
 $s_i \in S$

- ullet ${\cal S}$ is the state space
- ullet S is a discretized representation of space
- s_i is a *pixel* centeroid & possible activity center location

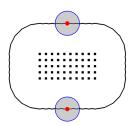
- ullet ${\cal S}$ is part of the model!
 - defines where individuals can live
 - defines the population of interest
 - includes unsampled parts of the landscape



- ullet ${\cal S}$ is part of the model!
 - defines where individuals can live
 - defines the population of interest
 - includes unsampled parts of the landscape
- should represent activity centers of all detectable individuals
 - \Box a buffer of at least $2\hat{\sigma}$ around traps
 - ensures activity centers of detectable inds. are represented



- S is part of the model!
 - defines where individuals can live
 - defines the population of interest
 - includes unsampled parts of the landscape
- should represent activity centers of all detectable individuals
 - \Box a buffer of at least $2\hat{\sigma}$ around traps
 - ensures activity centers of detectable inds. are represented
- discrete approximation space
 - \square pixel centroids \rightarrow activity centers
 - $\ \square$ resolution should be $\leq \hat{\sigma}$



ssDF - the state space data object

ssDF: the state space data object (state space data frame):

- a list containing a data frame
- at least the coordinates of the discrete state space
 - must be named X and Y (upper case)
 - each coordinate represents a pixel centroid
- can add named columns of pixel-specific covariate values
 - used to model spatial variation in density
 - continuous or categorical
 - coordinates should be same units as traps
 - non-habitat can be removed!

ssDF - a state space data object

Create the ssDF object Using make.ssDF():

ssDF - a state space data object

Create the ssDF object Using make.ssDF():

A note about specifying values:

- ullet buffer should be $\geq 2\hat{\sigma}$
- ullet res should be $<\hat{\sigma}$
- use $\frac{1}{2}$ mmdm as approximation of $\hat{\sigma}$
- always test sensitivity of parameter values to ssDF definition

Salamander ssDF

Construct the rbs ssDF

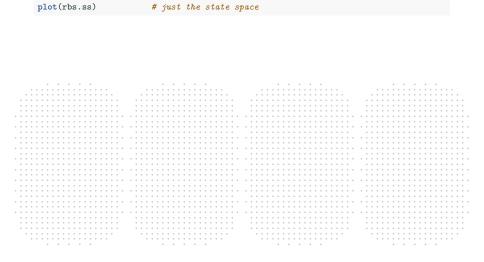
- sf is the scrFrame created earlier
- use it to construct a state space object

```
rbs.ss <- make.ssDF(scrFrame = rbs.sf,  # the rbs scrFrame buffer = 3,  # 3 m res = 0.5) # 0.5 m
```

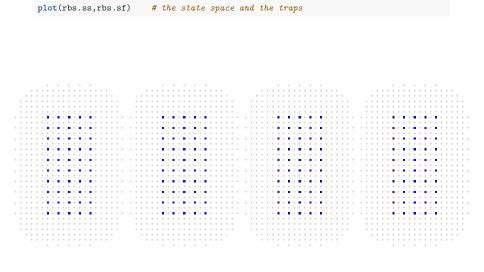
• visualize the state space object using plot() or plot.ssDF()

```
plot(rbs.ss) # just the state space
plot(rbs.ss,rbs.sf) # the state space and the traps
```

Salamander ssDF



Salamander ssDF



Salamander data preperation

```
# read in the data
rbs.edf <- read.csv("rbsNY.csv".h=T)
tdf1 <- read.csv("tdfNY1.csv".h=T)
tdf2 <- read.csv("tdfNY2.csv".h=T)
tdf3 <- read.csv("tdfNY3.csv",h=T)
tdf4 <- read.csv("tdfNY4.csv",h=T)
# create a 'encounters' data object (scrFrame)
rbs <- data2oscr(edf = rbs.edf, tdf = list(tdf1,tdf2,tdf3,tdf4),
                 sess.col = 1, id.col = 2, occ.col = 3, trap.col = 4,
                 K = c(7,5,6,4), ntraps = c(50,50,50,50)
rbs.sf <- rbs$scrFrame
# create a spatial extent object (ssDF)
rbs.ss <- make.ssDF(scrFrame = rbs.sf,
                    buffer = 3.
                    res = 0.5)
```

• next step, model fitting

oSCR.fit - the main model fitting function

To fit models in oSCR:

- use the fitting function oSCR.fit()
- must provide an scrFrame
- must provide an ssdf
- specify the model

oSCR.fit - the main model fitting function

To fit models in oSCR:

- use the fitting function oSCR.fit()
- must provide an scrFrame
- must provide an ssdf
- · specify the model

oSCR - Model fitting

model:

- a list with 3 model formulations
- list(D ~ 1, p0 ~ 1, sig ~ 1)
- D \sim : model describing variation pixel density $(D(s_i))$
- p0 ~: model describing variation in baseline encounter prob/rate (p_0)
- sig ~: model describing variation in sigma (σ)

Model SCR_0 in oSCR - Density

$$log(D(s_i)) = \beta$$

- inference about per pixel density, $D(s_i)$
- log-linear model to ensure positive densities
 - □ i.e. need to exponentiate estimate!
- intercept only model specification: D ~ 1

Model SCR₀ in oSCR - baseline encounter probability

$$p[x_j, s_i] = p_0 \times \exp\left(-\frac{\operatorname{dist}(x_j, s_i)^2}{2\sigma^2}\right)$$

$$logit(p_0) = \alpha_0$$

- inference about encounter probability, p_0
- · constant across all individuals
- logit model for probabilities (ensures 0-1 bounds)
- intercept only model specification: p0 ~ 1

Model SCR₀ in oSCR - spatial scale parameter

$$p[x_j, s_i] = p_0 \times \exp\left(-\frac{\operatorname{dist}(x_j, s_i)^2}{2\sigma^2}\right)$$

$$log(\sigma) = \gamma_0$$

- \bullet inference about the spatial scale of detection, σ
- constant across all individuals
- log-linear model to ensure positive distances
 - □ i.e. need to exponentiate estimate!
- intercept only model specification: sig ~ 1

Fitting model SCR_0 is as simple as:

So what are the ... additional arguments?

- oSCR is very flexible and has many options/setting
- check the help file ?oSCR.fit()
- BUT for salamnders with ACO, must use multicatch=TRUE

Fitting model SCR_0 is as simple as:

So what are the ... additional arguments?

- oSCR is very flexible and has many options/setting
- check the help file ?oSCR.fit()
- BUT for salamnders with ACO, must use multicatch=TRUE

Let's fit SCR₀ to the salamander data

Type the name of the model for a model summary:

```
rbs.scr0

Model: D ~ 1 p0 ~ 1 sig ~ 1

Run time: 3.875 minutes

AIC: 3119.959

Summary table:

Estimate SE z P(>|z|)

p0.(Intercept) -1.730 0.120 -14.406 0

sig.(Intercept) -0.468 0.040 -11.690 0

d0.(Intercept) -0.946 0.069 -13.735 0

*Density intercept is log(individuals per pixel)

Nhat(state-space) = exp(d0.)*nrow(ssDF)

(caution is warranted when model contains density covariates)
```

- model took quite a long time to run (~6.5 mins)!
- can reduce run time using (see ?oSCR.fit())

Salamander analysis workflow

```
# read in the data
rbs.edf <- read.csv("rbsNY.csv",h=T)
tdf1 <- read.csv("tdfNY1.csv".h=T)
tdf2 <- read.csv("tdfNY2.csv",h=T)
tdf3 <- read.csv("tdfNY3.csv".h=T)
tdf4 <- read.csv("tdfNY4.csv".h=T)
# create a 'encounters' data object (scrFrame)
rbs <- data2oscr(edf = rbs.edf, tdf = list(tdf1.tdf2.tdf3.tdf4).
                 sess.col = 1, id.col = 2, occ.col = 3, trap.col = 4,
                 K = c(7,5,6,4), \text{ ntraps} = c(50,50,50,50))
rbs.sf <- rbs$scrFrame
# create a spatial extent object (ssDF)
rbs.ss <- make.ssDF(scrFrame = rbs.sf.
                    buffer = 3,
                    res = 0.5)
# fit model SCRO
rbs.scr0 <- oSCR.fit(list(D~1, p0~1, sig~1), rbs.sf, rbs.ss, trimS=4)
```

• all that's left is to interpret the output - *inference*!

Interpreting and processing oSCR output

What can you say about these results?

```
rbs.scr0

Model: D ~ 1 p0 ~ 1 sig ~ 1

Run time: 3.875 minutes

AIC: 3119.959

Summary table:

Estimate SE z P(>|z|)

p0.(Intercept) -1.730 0.120 -14.406 0

sig.(Intercept) -0.468 0.040 -11.690 0

d0.(Intercept) -0.946 0.069 -13.735 0

*Density intercept is log(individuals per pixel)

Nhat(state-space) = exp(d0.)*nrow(ssDF)

(caution is warranted when model contains density covariates)
```

Interpreting and processing oSCR output

What can you say about these results?

```
rbs.scr0

Model: D ~ 1 p0 ~ 1 sig ~ 1

Run time: 3.875 minutes

AIC: 3119.959

Summary table:

Estimate SE z P(>|z|)

p0.(Intercept) -1.730 0.120 -14.406 0

sig.(Intercept) -0.468 0.040 -11.690 0

d0.(Intercept) -0.946 0.069 -13.735 0

*Density intercept is log(individuals per pixel)

Nhat(state-space) = exp(d0.)*nrow(ssDF)

(caution is warranted when model contains density covariates)
```

- on the *linear predictor/link* scale
- need to transform onto the real scale
- use get.real() function (needs library(car))

Session (plot) specific density estimates

ullet remember, *NULL* model o no variation

Session (plot) specific encounter probability estimates

ullet remember, NULL model o no variation

Session (plot) specific sigma estimates

ullet remember, NULL model o no variation

Model SCR_{session} (the session specific model) in oSCR

Let's fit SCR_{session} to the salamander data

Model SCR_{session} (the session specific model) in oSCR

Type the name of the model for a model summary:

```
rbs.scrS
Model: D ~ session pO ~ session sig ~ session
Run time: 42,44167 minutes
AIC: 3085,642
Summary table:
              Estimate SE z P(>|z|)
p0.(Intercept) -1.795 0.200 -8.981 0.000
p0.session2 1.008 0.436 2.313 0.021
p0.session3 -0.033 0.275 -0.121 0.904
p0.session4 0.642 0.459 1.399 0.162
sig.(Intercept) -0.338 0.070 -4.832 0.000
sig.session2 -0.672 0.118 -5.708 0.000
sig.session3 -0.047 0.095 -0.493
                                   0.622
sig.session4 -0.484 0.134 -3.599
                                   0.000
d0.(Intercept) -1.143 0.127 -8.982 0.000
d.beta.session2 0.294 0.210 1.398 0.162
d.beta.session3 0.451 0.170 2.656 0.008
d.beta.session4 0.214 0.232 0.925
                                   0.355
*Density intercept is log(individuals per pixel)
 Nhat(state-space) = exp(d0.)*nrow(ssDF)
  (caution is warranted when model contains density covariates)
```

session (plot) specific density estimates

Session (plot) specific encounter probability estimates

Session (plot) specific sigma estimates