Distance Sampling Simulations

Overview

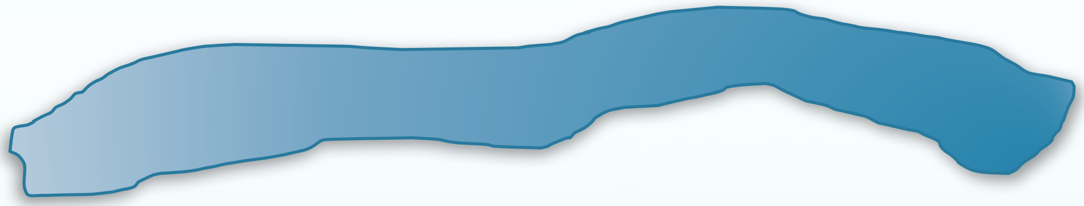
- Why simulate?
- How it works
- Automated survey design
  - Coverage probability
  - Which design?
  - Design trade-offs
- Defining the population
  - Population description
  - Detectability
- Example Simulations

Why Simulate?

- Surveys are expensive, we want to get them right! *(simulations cheap)*
- Test different survey designs
- Test survey protocols
- Investigate violation of assumptions
- Investigate analysis properties

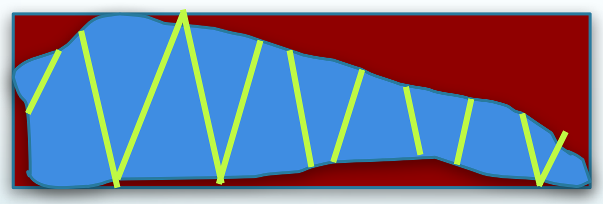
Why Simulate?

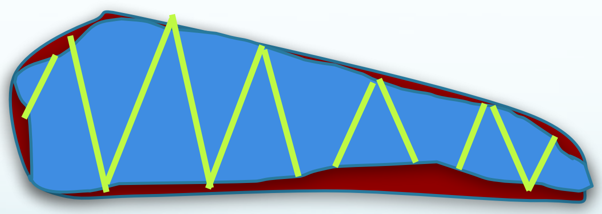
- I have a fairly long and narrow study region, are edge effects likely to be a problem?



Why Simulate?

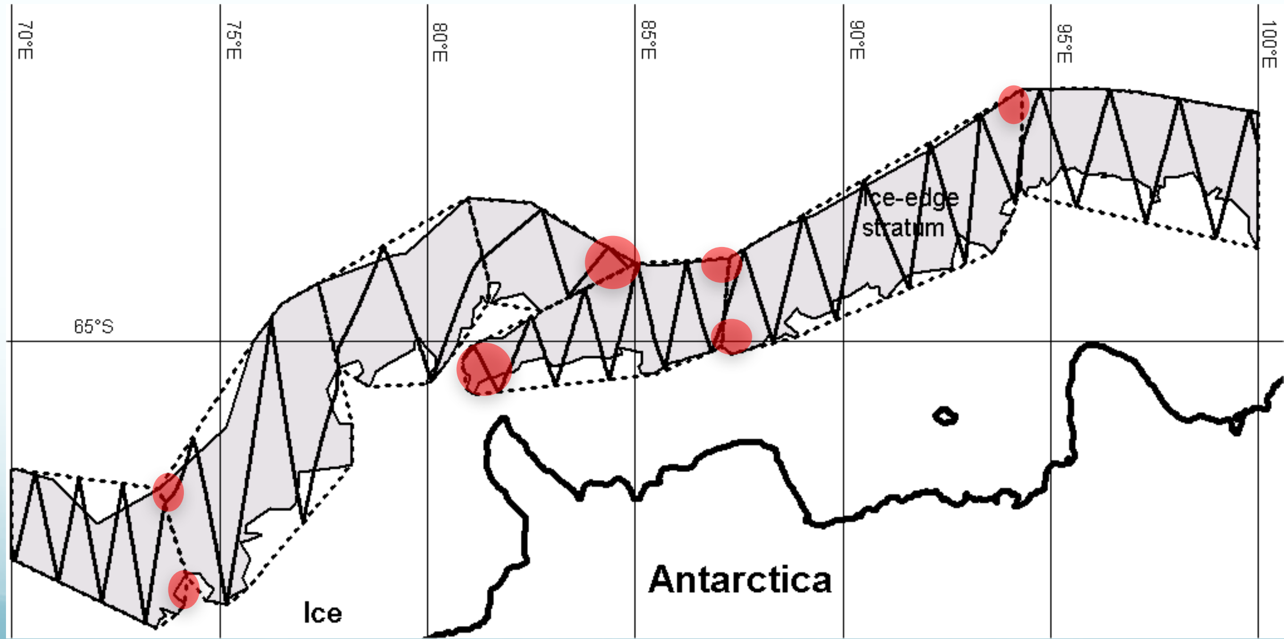
- Generating my equal spaced zig zag design in a convex hull gives better efficiency (less off effort transit time) but is this likely to introduce large amounts of bias due to non uniform coverage probability?





Why Simulate?

- What is the potential bias in this stratification technique?



Why Simulate?

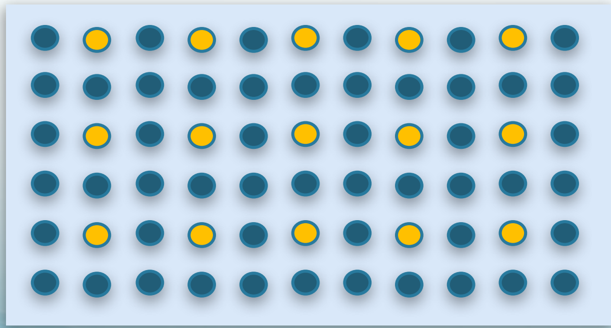
- From pilot study trials I know that there can be multiplicative error on recorded distances
- This error has a ~15% CV when collecting data in 3 bins or ~30% CV when attempting to collect exact distances… which is preferable (if we cannot improve accuracy or correct the measurements)?

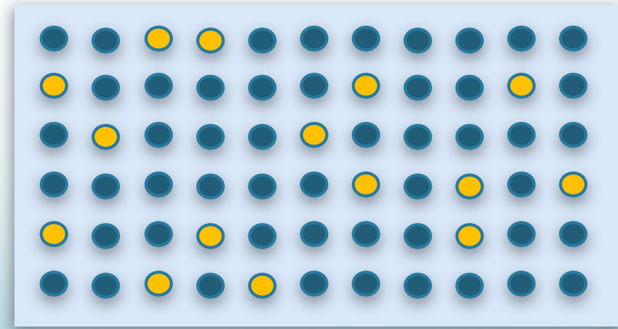
Why Simulate?

- We suspect that the current survey design is less than ideal and may be introducing bias but people are reluctant to change…
- Simulate the current situation to get an idea of how bad things could be
- Simulate a new design to show how things could be improved

Why Simulate?

- I want to do an acoustic survey with two types of detectors.
  - The first records distances as per standard distance sampling requirements (standard detectors).
  - The second only records the presence of a sound (simple nodes).
- How many standard nodes do I need and how should I distribute them?

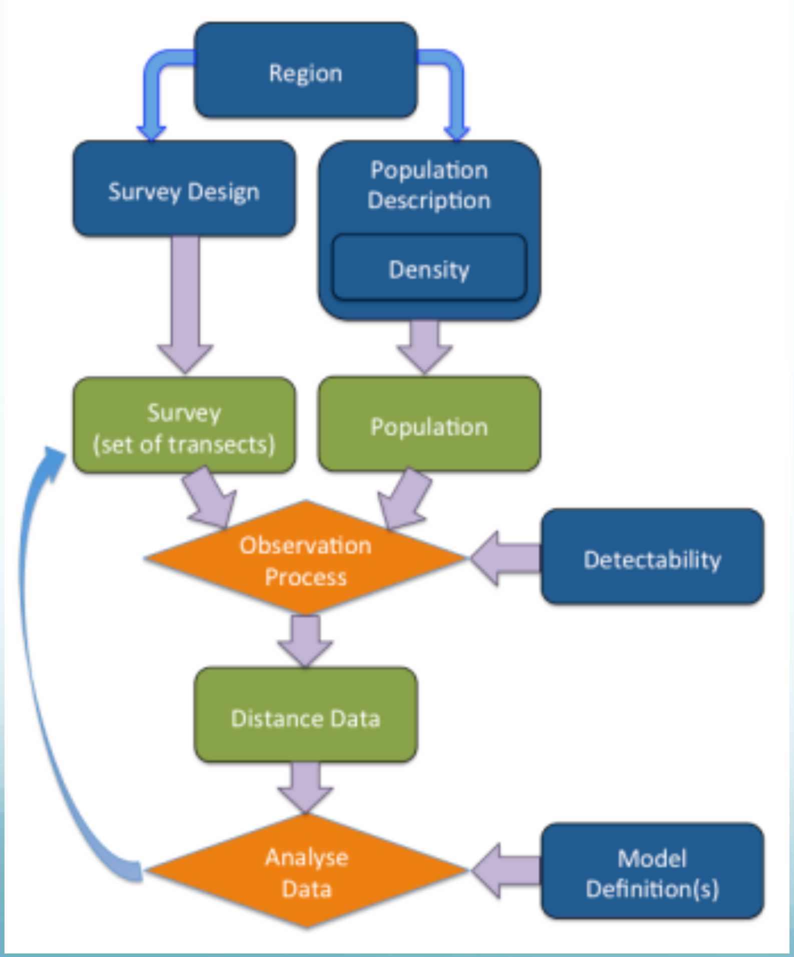




Why Simulate?

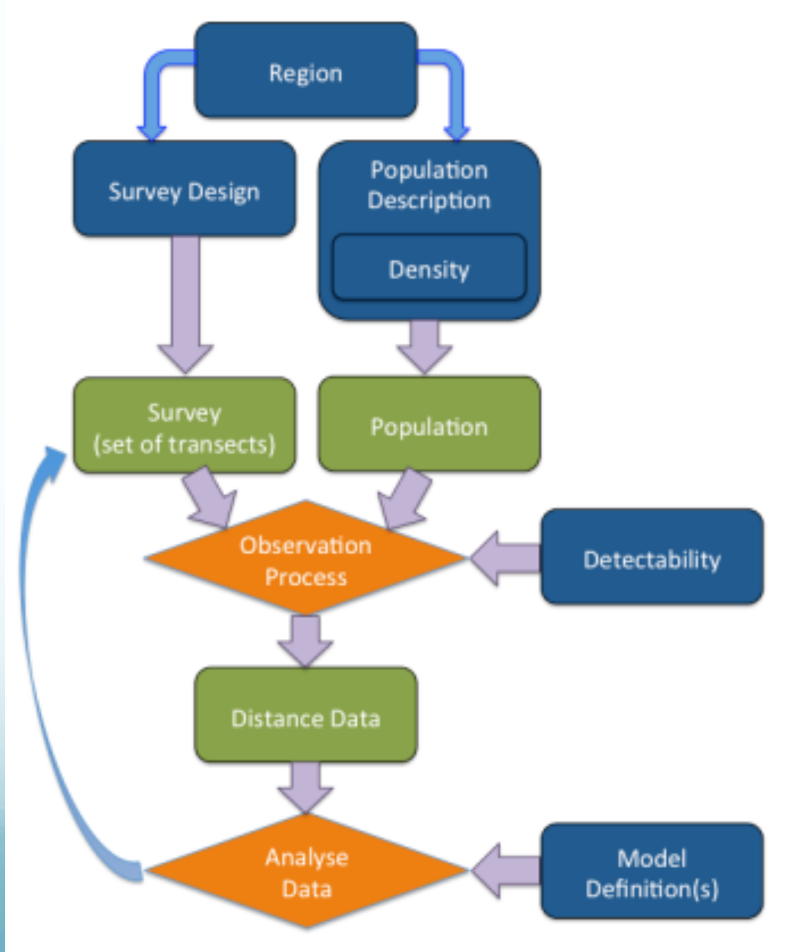
- I would like to use my data to generate both design (standard distance sampling) and model based (density surface model) estimates of density... which design will work best for my study?
- Hopefully coming soon to DSsim...
- Some example simulations can be found here: https://github.com/DistanceDevelopment/DSsim/wiki

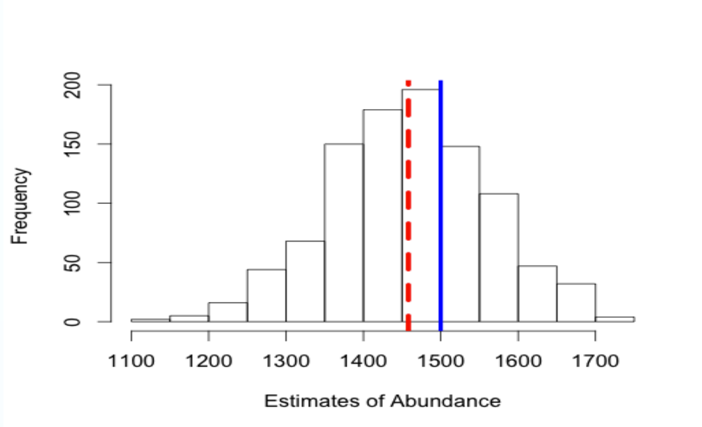
How it works



- Blue rectangles indicate information supplied by the user.
- Green rectangles are objects created by DSsim in the simulation process.
- Orange diamonds indicate the processes carried out by DSsim.

How it works





Assess:

- Bias
- Precision
- CI coverage

Across different designs/scenarios

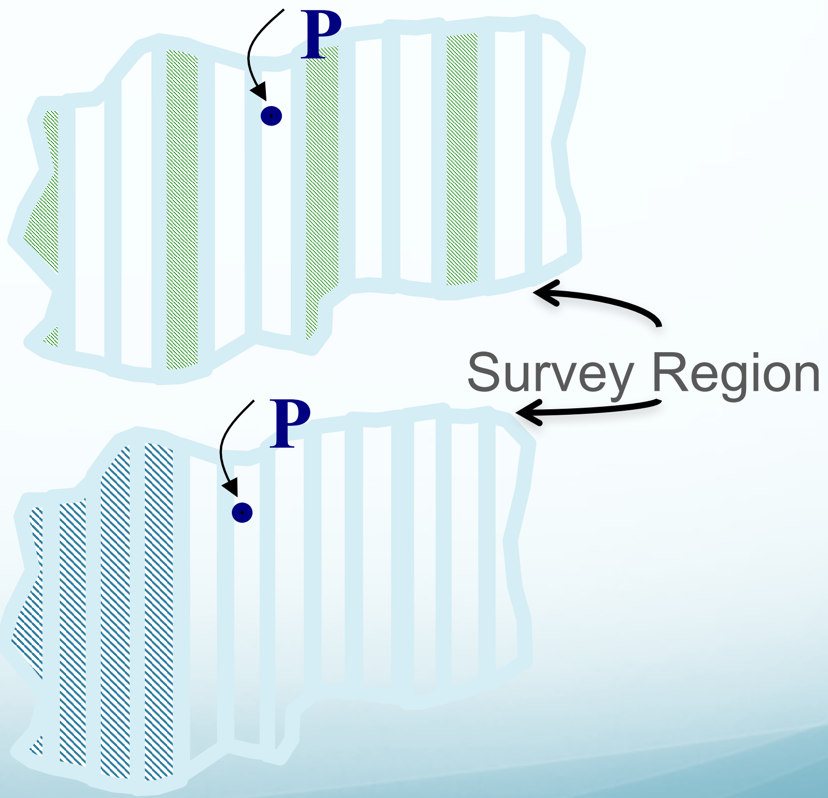
Automated Survey Design

- Generate random sets of transects according to an algorithm
  - Assess design properties
  - Generate multiple transect sets for simulations

Automated Survey Design

- Coverage Probability
  - Uniform coverage probability, π = 1/3
  - Even coverage for any given realisation

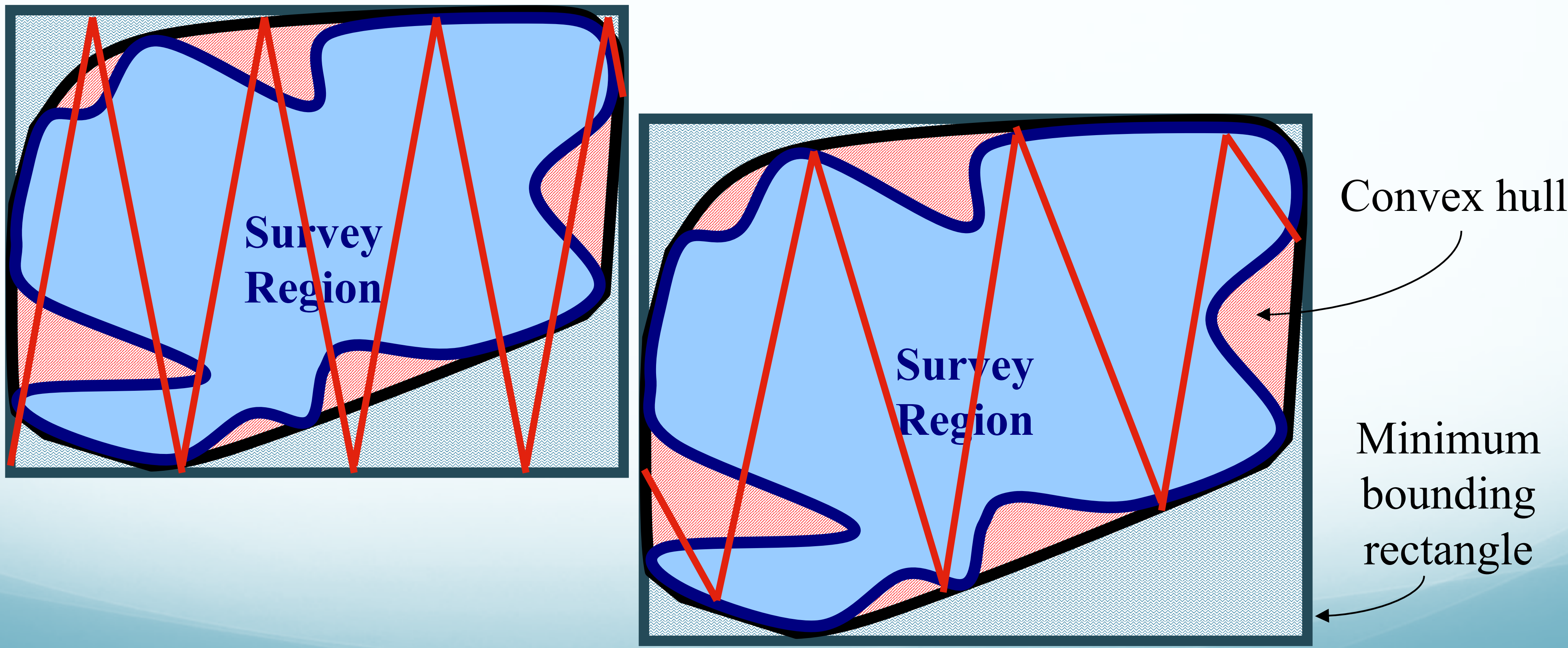
  - Uniform coverage probability, π = 1/3
  - Uneven coverage for any given realisation



Which Design?

- **Uniformity** of coverage probability
- **Even-ness** of coverage within any given realisation
- **Overlap** of samplers
- **Cost** of travel between samplers
- **Efficiency** when density varies within the region

Design Trade-Offs



Population Definition

- True population size?
- Occur as individuals or clusters?
- Covariates which will affect detectability?
- How is the population distributed within the study region?
  - Ideally have a previously fitted density surface
    Otherwise test over a range of plausible distributions

Detectability

- Distance needs:
  - shape and scale parameters on the natural scale
  - covariate parameters on the log scale

Detectability

- Golftees project

exp(0.268179) = 1.307581

Detection Fct/Global/Parameter Estimates (MCDS)

```
Effort        :    210.0000
# samples     :      1
Width         :    4.000000
# observations:   162

Model
   Half-normal key, k(y) = Exp(-y**2/(2*s**2))

   s = A(1) * Exp(fcn(A(2)) + fcn(A(3)))
   Parameter A(1) is the intercept of the scale parameter s.
   Parameter A(2) is the coefficient of covariate CLUSTER SIZE.
   Parameter A(3) is the coefficient of level 0 of factor covariate SEX.


             Point        Standard     Percent Coef.       95 Percent
 Parameter   Estimate       Error      of Variation    Confidence Interval
 ---------  -----------  -----------  --------------  ----------------------
   A( 1)      2.622        0.8370
   A( 2)    0.9294E-01   0.8172E-01
   A( 3)    -0.6951       0.2937
   f(0)     0.36330      0.17850E-01       4.91      0.32972       0.40030
   p        0.68814      0.33810E-01       4.91      0.62454       0.75821
   ESW       2.7525      0.13524           4.91       2.4981        3.0329
 ---------  -----------  -----------  --------------  ----------------------
```

Natural scale

Log scale

Detection Fct/Summary (MRDS)

```
Summary for ds object
Number of observations :  162
Distance range         :  0  -  4
AIC                    :  428.572

Detection function:
 Half-normal key function

Detection function parameters
Scale coefficient(s):
              estimate         se
(Intercept) 0.26817900 0.27140001
size        0.09314751 0.08176431
sex1        0.69600047 0.29401571

                          Estimate          SE         CV
Average p                0.6882835  0.05258548 0.07640090
N in covered region 235.3681131 21.00939868 0.08926187
```

Detectability

- In simulation:

Detectability

Detection function model: Half-Normal

☐ Define parameters for each stratum

Region	Study Area
Scale	1.31
Shape	
cluster size	0.093
sex.0	0
sex.1	0.696

(The units for the detection function are 'Meter')

=

Detectability

Detection function model: Half-Normal

☐ Define parameters for each stratum

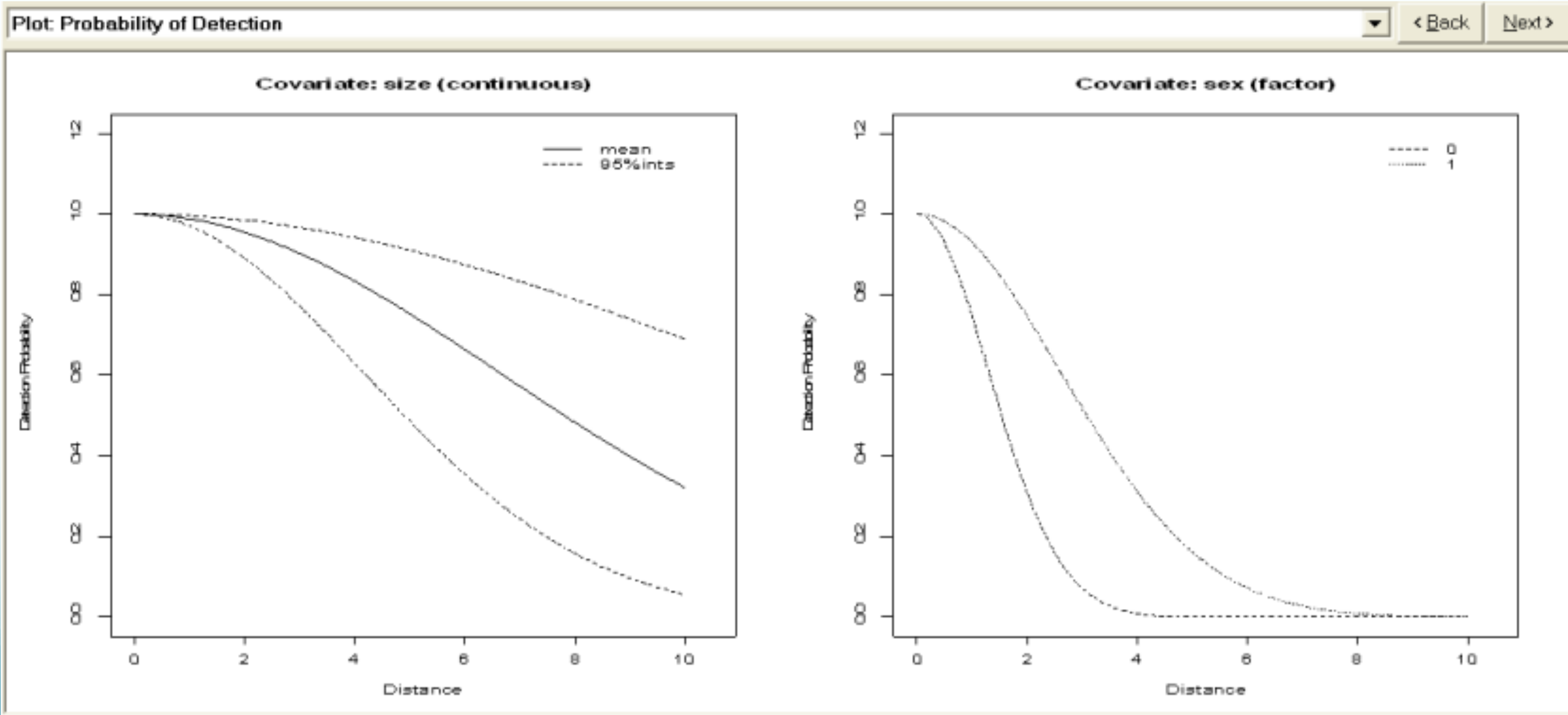
Region	Study Area
Scale	2.62
Shape	
cluster size	0.093
sex.0	-0.696
sex.1	0

(The units for the detection function are 'Meter')

exp(log(1.307581)+0.696) = 2.622633

exp(log(2.622)-0.696) = 1.307265

Detectability



Analysis

- **Data Filter** must specify a right truncation distance
- **Model Definition** must be either MRDS or MA
  - MRDS – for fitting a specific model
  - MA – for model selection (Note: MA model definitions require the creation of analyses)

Any questions so far…

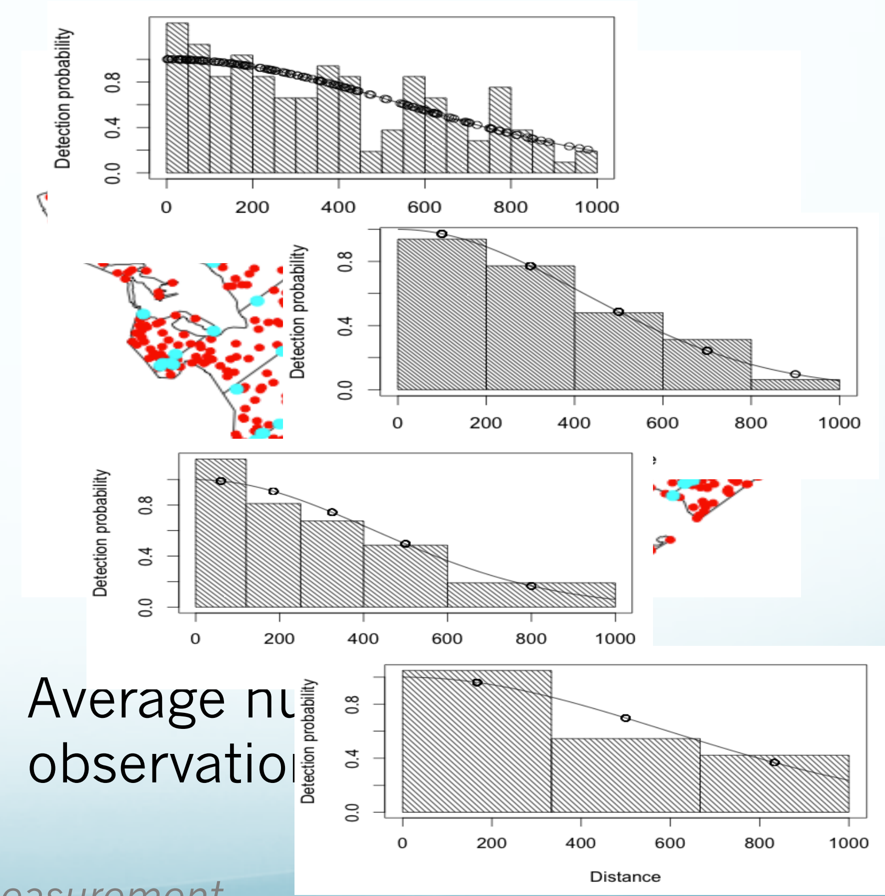
Example Simulations

- To bin or not to bin?
  - It is better to collect binned data accurately than attempt to collect exact distances and introduce measurement error!
- Testing pooling robustness in relation to truncation distance.
  - Demonstrating why you shouldn't be scared to truncate distance sampling data
- Comparison of subjective and random designs.
  - How wrong can you go with a subjective design?
  - Comparing zig zag and parallel designs.

To Bin or Not to Bin?

Simulation:

- Generated 999 datasets
- Added multiplicative measurement error
  - Distance = True Distance * R
  - R = (U + 0.5), where U~Beta(θ, θ)[1]
  - No error, ~15% CV (θ = 5), ~30% CV (θ = 1)
- Analysed them in difference ways
  - Exact distances, 5 Equal bins, 5 Unequal bins, 3 Equal bins
- Model selection on minimum AIC
  - Half-normal v Hazard rate



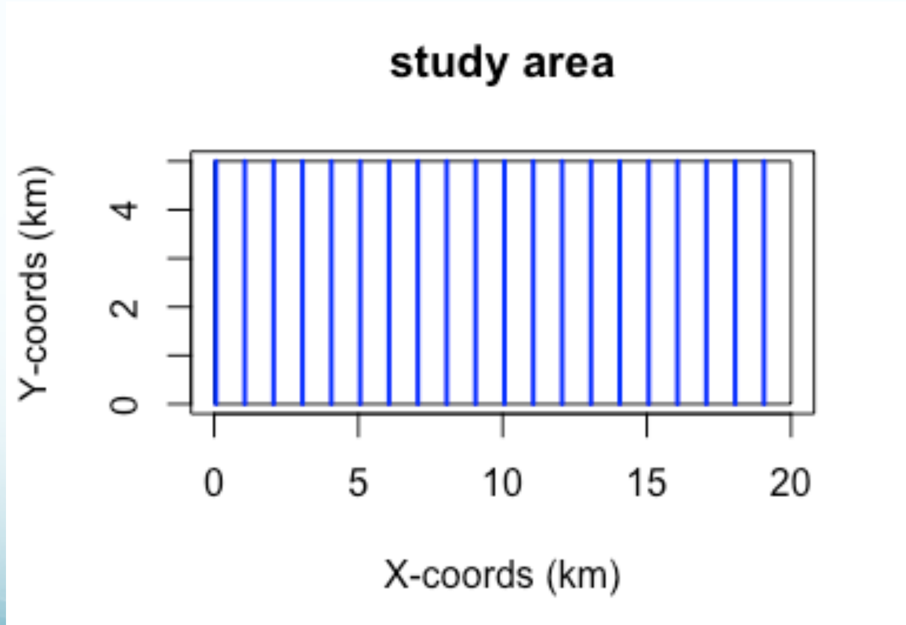
[1]*Marques T. (2004) Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models Biometrics* ***60****:757--763*

To Bin or Not to Bin Results

	Exact Distances	5 Equal Bins	5 Unequal Bins	3 Equal Bins
No Error	-1.16% bias 210 SE	-1.11% bias 217 SE	-0.16% bias 221 SE	-0.19% bias 255 SE
15% CV	0.48% bias 214 SE	o.5% bias 221 SE	1.36% bias 221 SE	1.72%bias 264 SE
30% CV	6.66% bias 237 SE	6.61% bias 250 SE	7.43% bias 262 SE	8.20% bias 338 SE

Pooling Robustness and Truncation

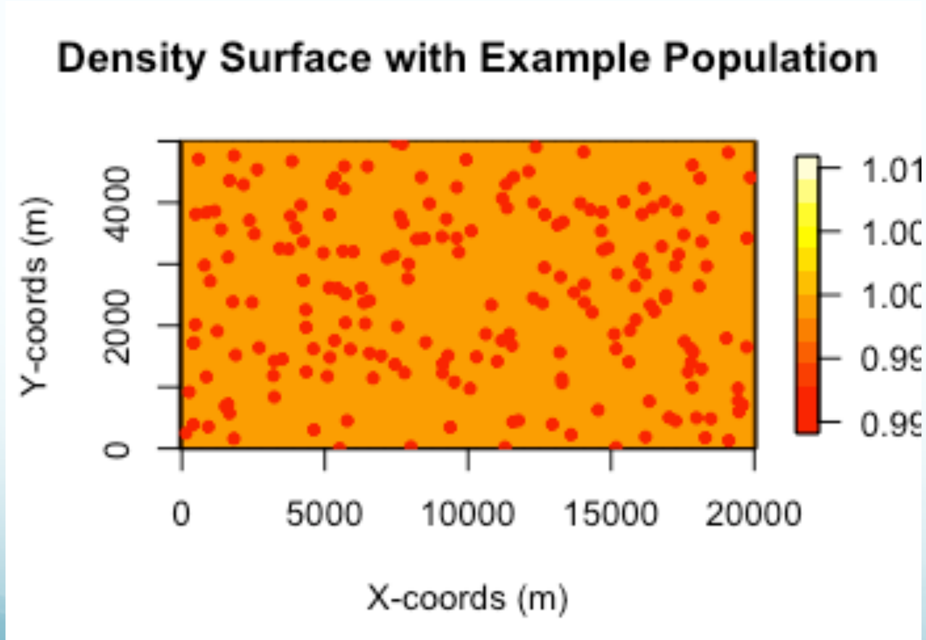
- DSsim vignette



- Rectangular study region
- Systematic parallel transects with a spacing of 1000m

Pooling Robustness and Truncation

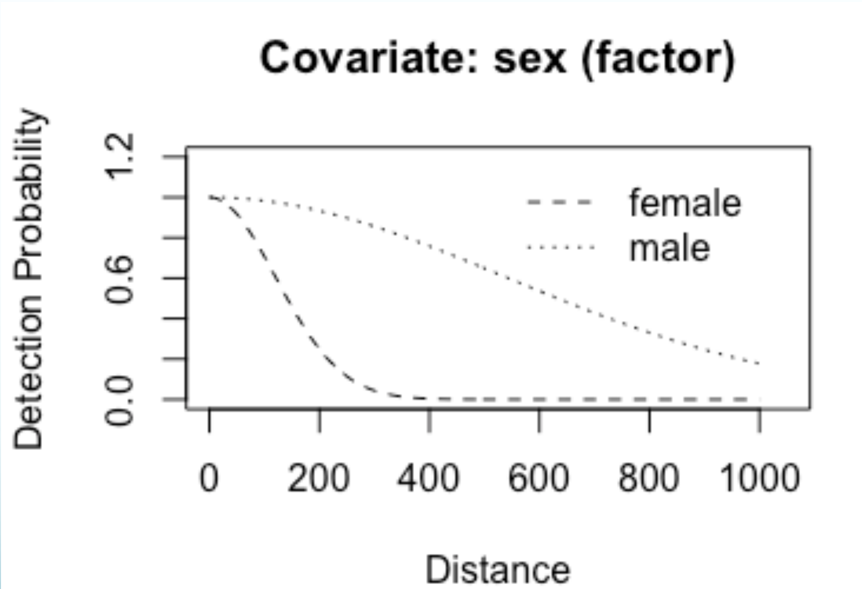
- DSsim vignette



- Uniform density surface
- Population size of 200
- 50% male, 50% female

Pooling Robustness and Truncation

- DSsim vignette



- Half-normal shape for detectability
- Scale parameter of 120 for the females
- Scale parameter of ~540 for the males

Pooling Robustness and Truncation

- DSsim vignette

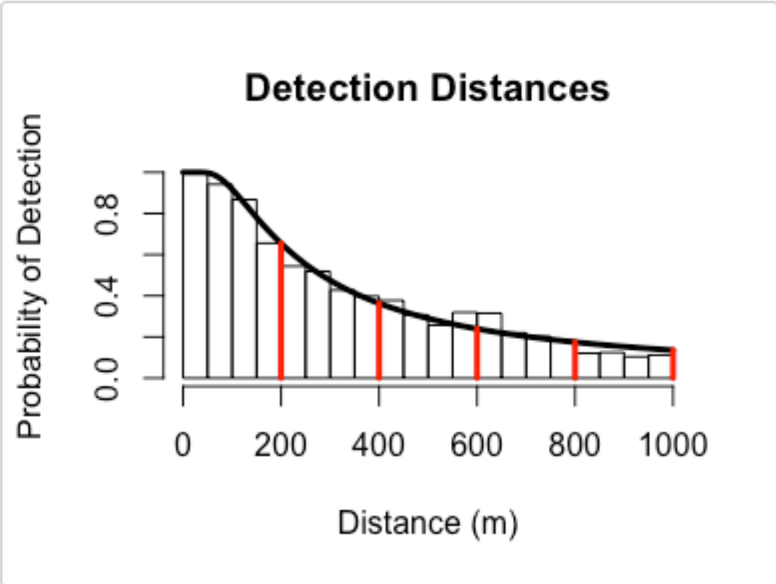
```r
# Create the covariate parameter list
cov.params <- list()
# Note the covariate parameters are supplied on the log scale
cov.params$sex = data.frame(level = c("female", "male"),
                            param = c(0, 1.5))

detect.cov <- make.detectability(key.function = "hn" ,
                                 scale.param = 120,
                                 cov.param = cov.params,
                                 truncation = 1000)
```

$\exp(\log(120)+1.5) = 537.8$

- Half-normal shape for detectability
- Scale parameter of 120 for the females
- Scale parameter of ~540 for the males

Pooling Robustness and Truncation

- DSsim vignette



Histogram of data from covariate simulation with manually selected candidate truncation distances.

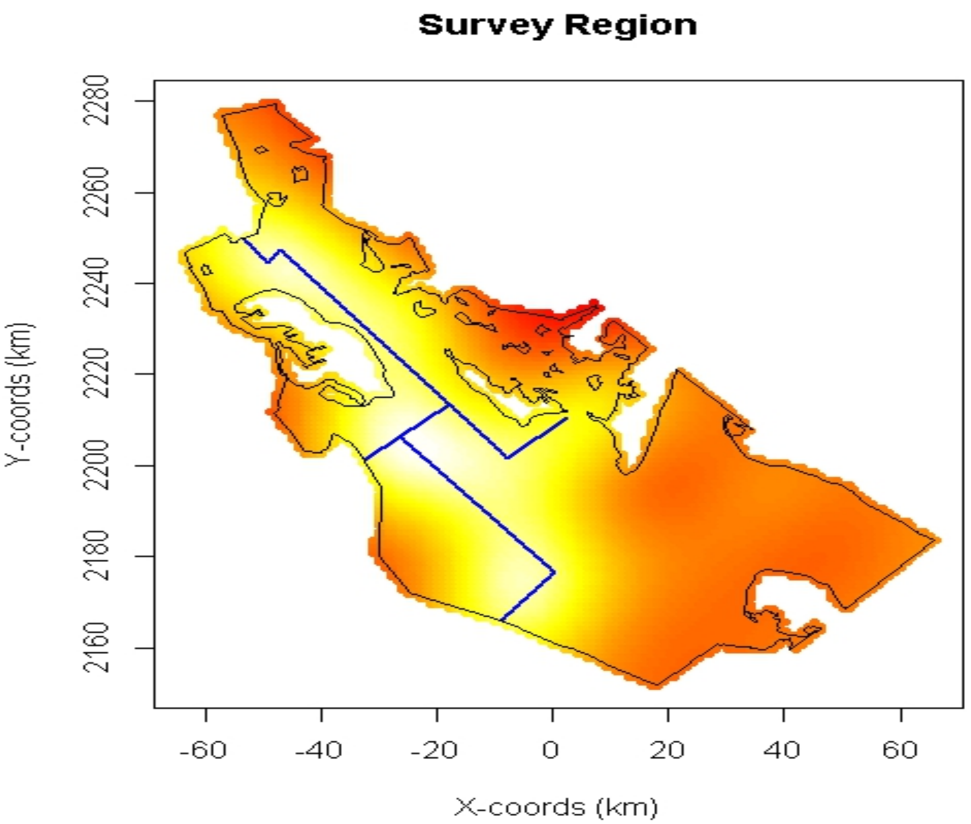
- Two types of analyses:
  - **hn v hr**
  - hn ~ sex
- Selection criteria: AIC

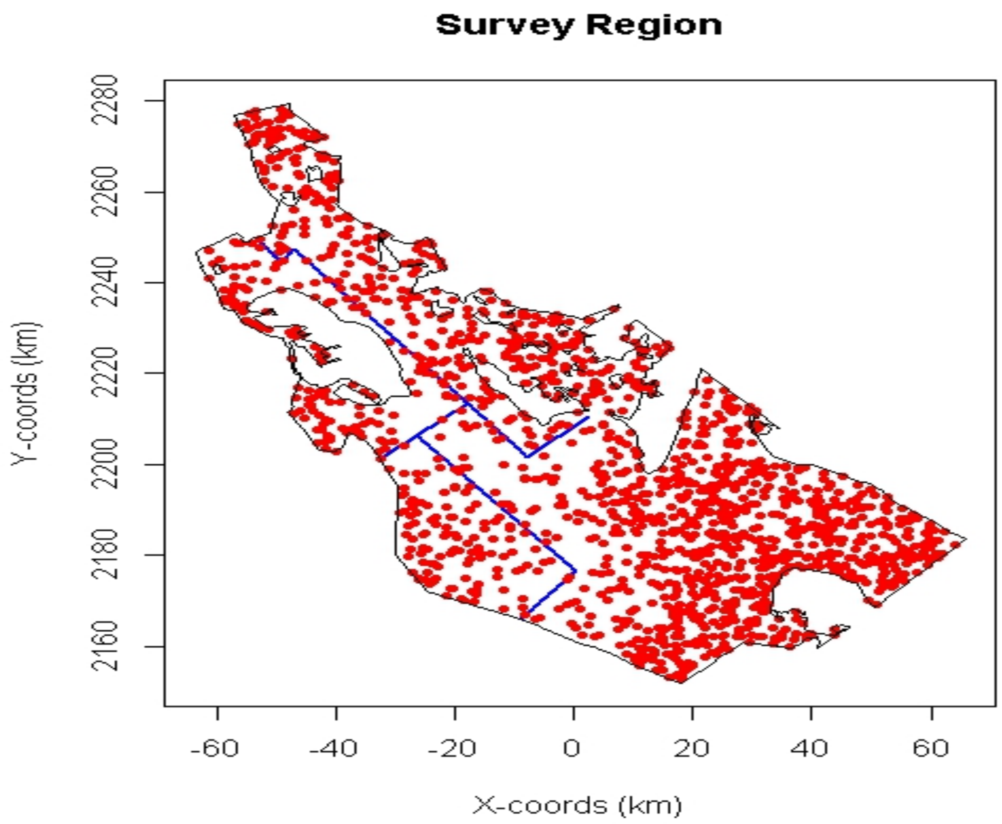
Pooling Robustness and Truncation

- Results HN v HR:

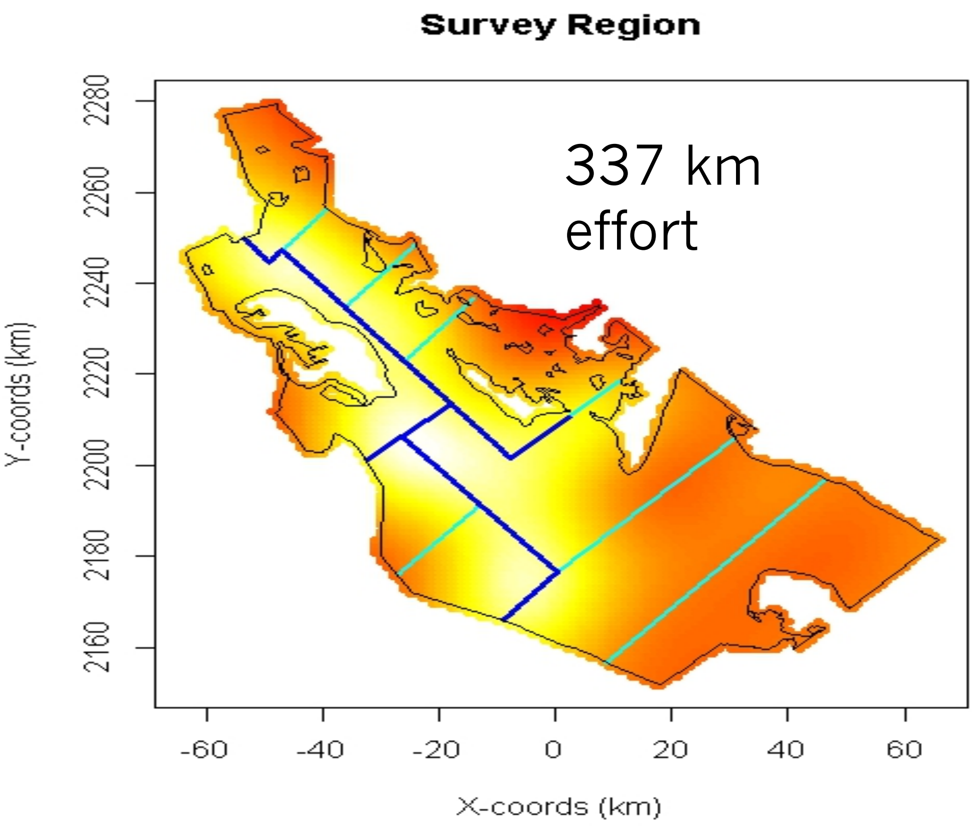
Truncation	*mean n*	*mean* $\hat{N}$	*mean se*	$SD(\hat{N})$	*%Bias*	*RMSE*	*% CI Coverage*
200	66	197	34.27	34.05	-1.32	34.13	97.5
400	102	190	31.06	34.79	-5.13	36.25	87.9
600	128	190	34.04	35.27	-5.24	36.77	81.9
800	144	190	34.31	36.61	-5.10	37.99	77.1
1000	154	184	30.93	39.49	-7.76	42.42	68.1

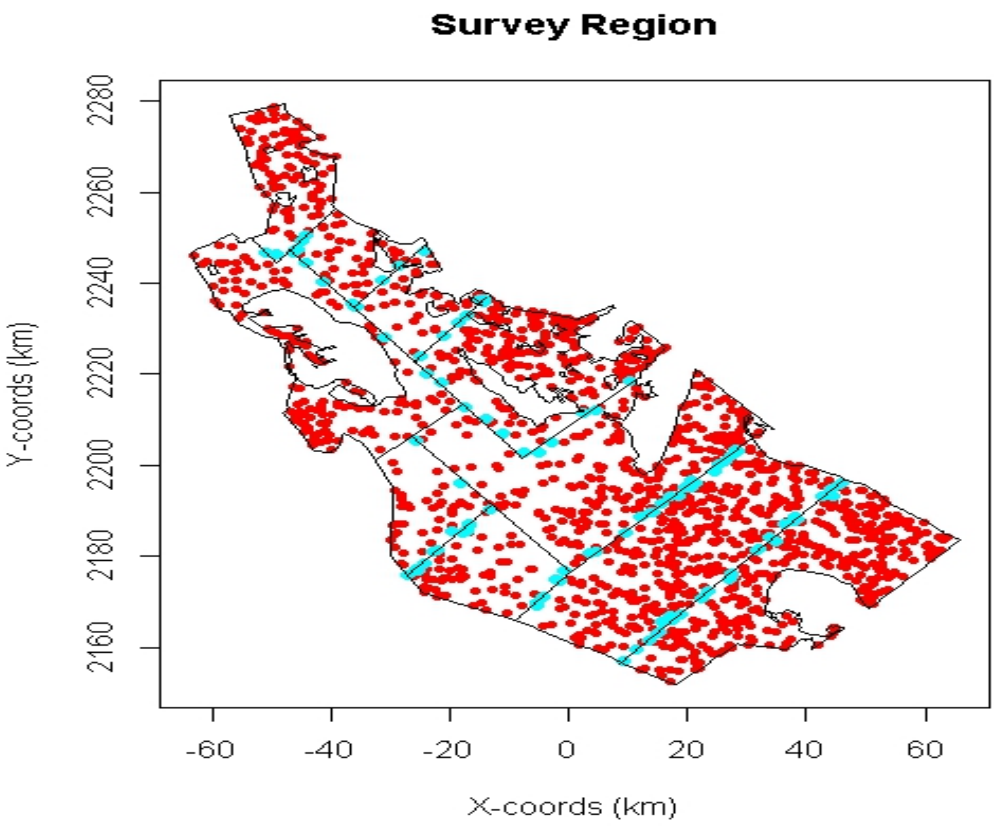
Example Simulation



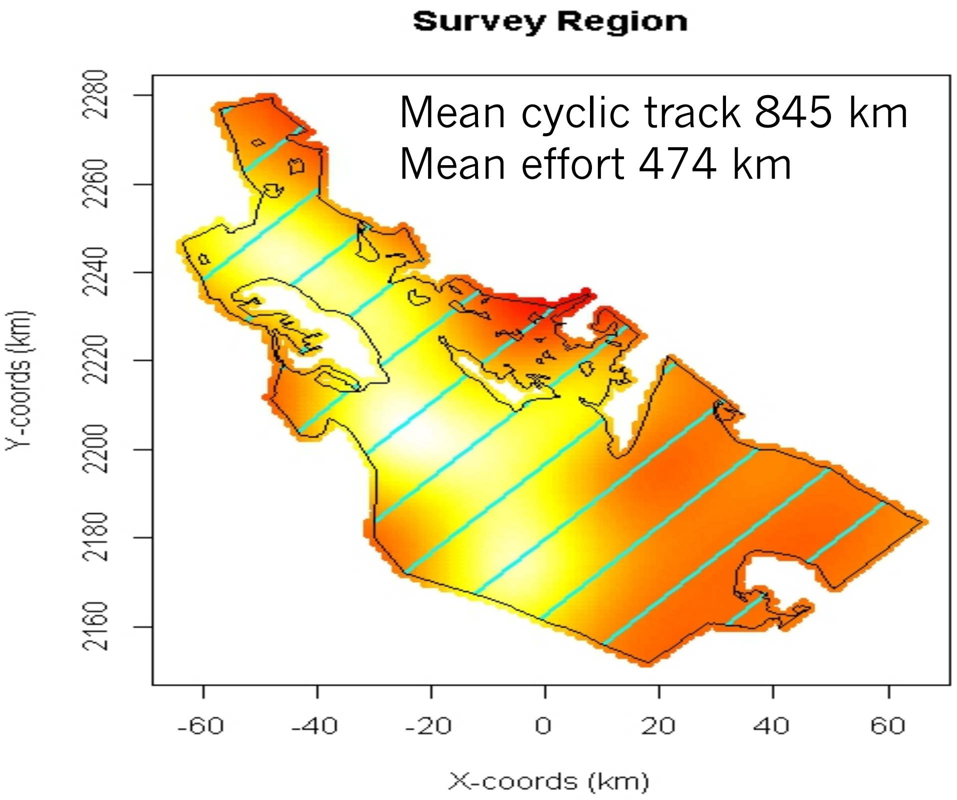


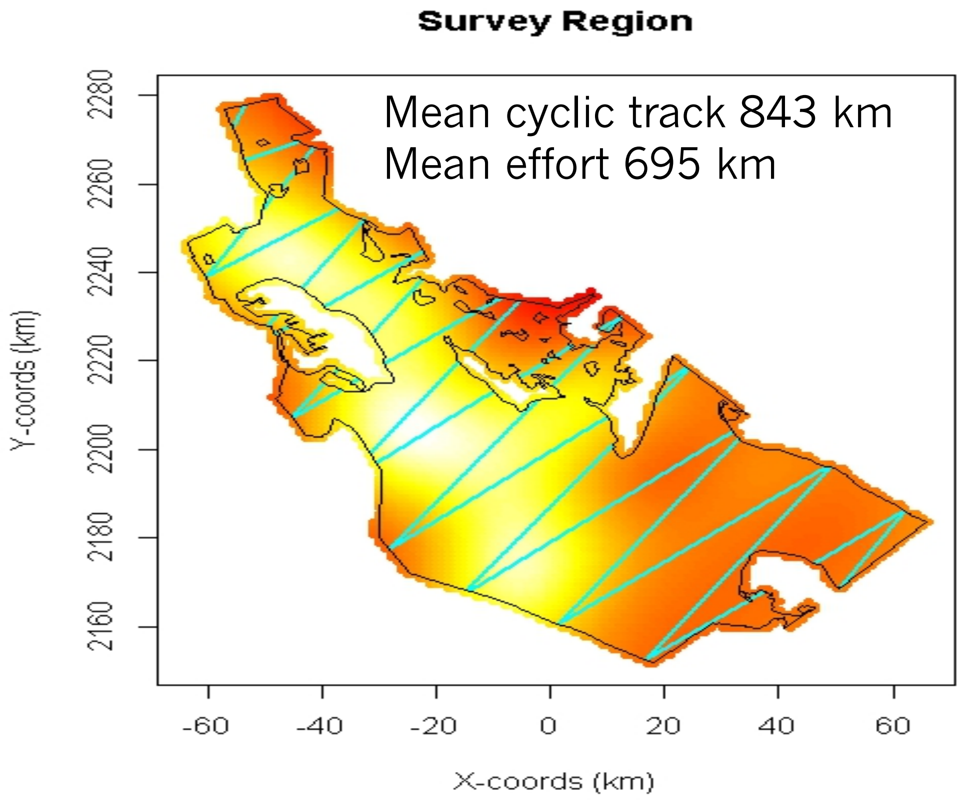
Subjective survey design





Random Designs





Coverage probability

Simulation

- Generates a realisation of the population based on a fixed N of 1500
- Generates a realisation of the design
  - Different each time for the random designs
  - The same each time for the subjective design
- Simulates the detection process
- Analyses the results
  - Half-normal
  - Hazard-rate
- Repeats a number of times

Practical

- Now attempt the DSsim practical:
  - *R version – subjective design and parallel v zig zag*
  - *Distance version – parallel v zig zag only*
- You will need the library *shapefiles.*