spatial modelling for ecological surveys – contributions from and to point process modelling

Janine Illian

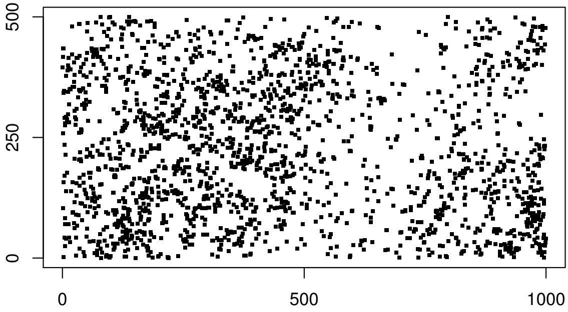
CREEM
Centre for Research into Ecological and Environmental Modelling,
University of St Andrews, Scotland, UK

December 7, 2017

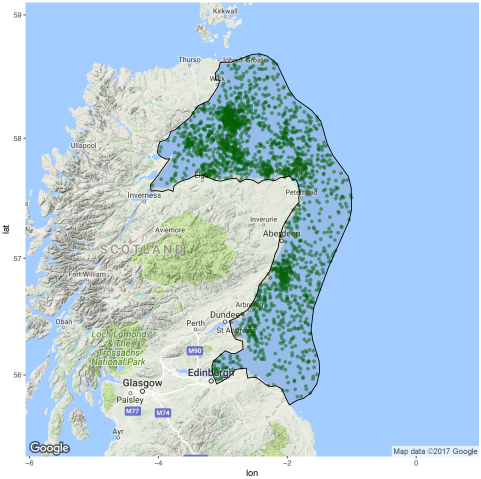
joint work with: David Borchers, Fabian Bachl, Yuan Yuan, Håvard Rue, Finn Lindgren, Daniel Simpson, Laura Williamson and others

an example:

Oenocarpus mapoura observed in a 50-ha study plot on Barro Colorado Island, Panama

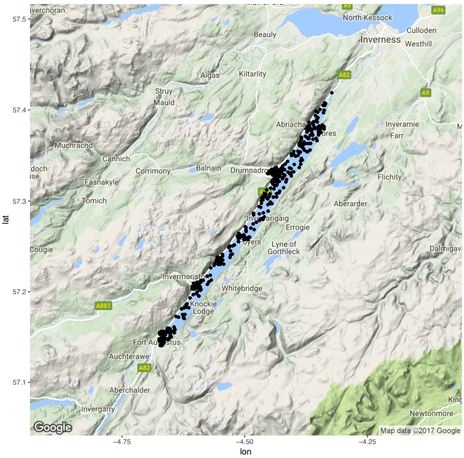


some more examples:



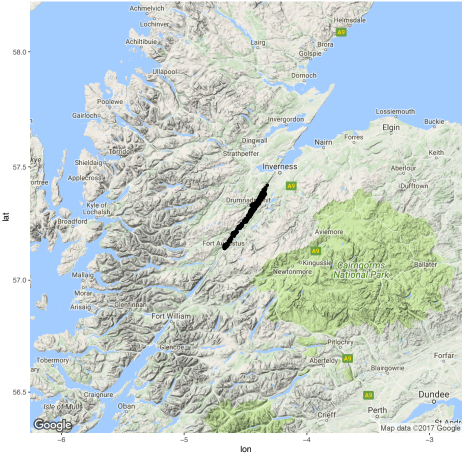
Locations of harbour porpoise sightings off the East Coast of Scotland.

some more examples:



Locations of reported sightings of the Loch Ness Monster, Loch Ness, Scotland.

some more examples:



Locations of reported sightings of the Loch Ness Monster, Loch Ness, Scotland.

spatial point processes in ecology

ecology – main interest:

- interactions among individual organisms and environment
- individuals exist – and interact – in space and time
- spatially explicit data increasingly available

spatial point processes in ecology

ecology – main interest:

- interactions among individual organisms and environment
- individuals exist – and interact – in space and time
- spatially explicit data increasingly available

$\Rightarrow$ data: spatial (spatio-temporal) point patterns

spatial point processes in ecology

ecology – main interest:

- interactions among individual organisms and environment
- individuals exist – and interact – in space and time
- spatially explicit data increasingly available

$\Rightarrow$ data: spatial (spatio-temporal) point patterns

$\Rightarrow$ **spatial point process methodology** should be highly relevant!

however...

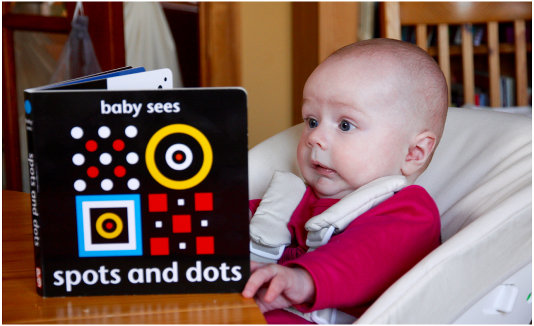
- few ecologists aware of spatial point process methodology
- e.g. models rarely used in practice

$\Rightarrow$ not part of the standard statistical toolbox

spatial point processes in ecology

WHY?

WHY?
In the end it's just a bunch of dots, isn't it?



but we have all these cool models...

with log Gaussian Cox processes and INLA+SPDE we can

but we have all these cool models...

with log Gaussian Cox processes and INLA+SPDE we can

- flexibly account for remaining autocorrelation

but we have all these cool models...

with log Gaussian Cox processes and INLA+SPDE we can

- flexibly account for remaining autocorrelation
- jointly model individuals' properties (*marks*) and spatial covariates with spatial pattern

but we have all these cool models...

with log Gaussian Cox processes and INLA+SPDE we can

- flexibly account for remaining autocorrelation
- jointly model individuals' properties (*marks*) and spatial covariates with spatial pattern
- have point pattern reflect observation process:

but we have all these cool models...

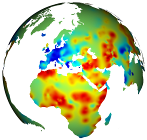
with log Gaussian Cox processes and INLA+SPDE we can

- flexibly account for remaining autocorrelation
- jointly model individuals' properties (*marks*) and spatial covariates with spatial pattern
- have point pattern reflect observation process: preferential sampling

but we have all these cool models...

with log Gaussian Cox processes and INLA+SPDE we can

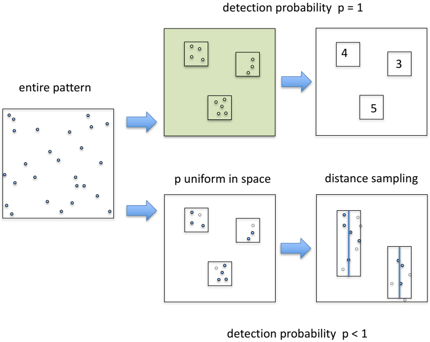
- flexibly account for remaining autocorrelation
- jointly model individuals' properties (*marks*) and spatial covariates with spatial pattern
- have point pattern reflect observation process: preferential sampling
- modelling on complex domains
  - the sphere = the earth
  - observation areas with barriers (islands, archipelagos...)



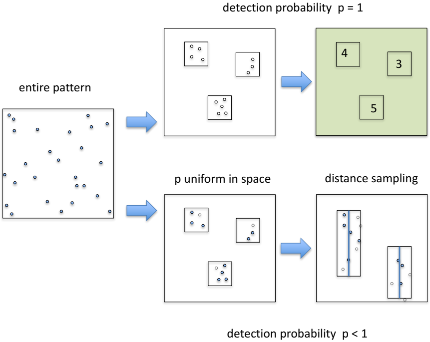
Illian et al., 2012 a+b, Simpson et al., 2016, Bakka et al., 2016, Ledo et al., 2016, Jones-Todd et al., 2017

ecological surveys...

Area of interest is too big to sample entirely.
thinned point process

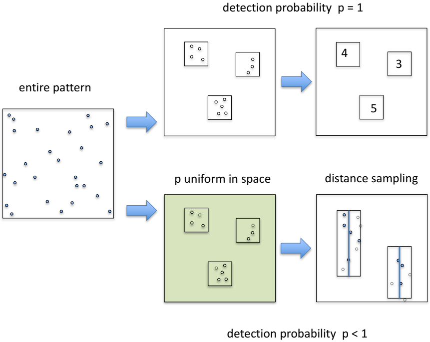


Area of interest is too big to sample entirely.



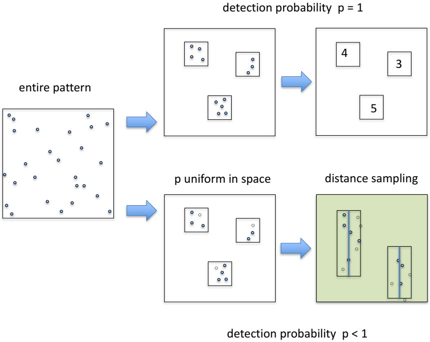
ecological surveys...

Area of interest is too big to sample entirely.

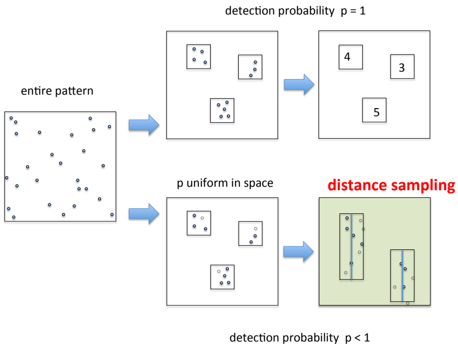


ecological surveys...

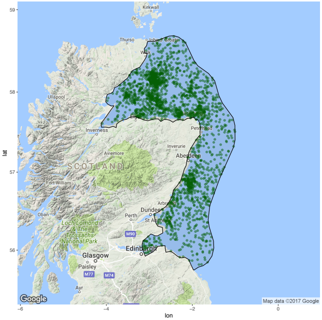
Area of interest is too big to sample entirely.



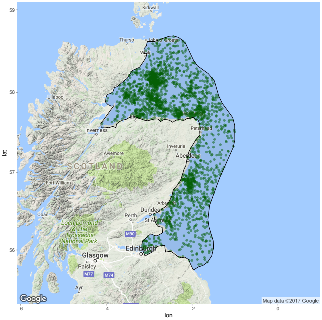
Area of interest is too big to sample entirely.

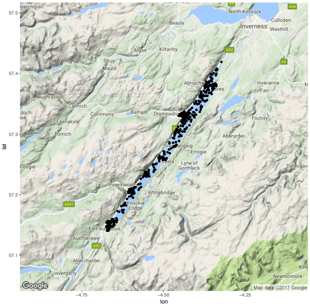


Examples... recall...

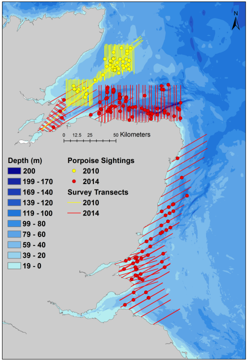


Examples... recall...





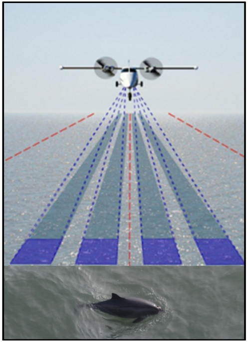
harbour porpoise study–video survey



video survey data

- conducted in August and September 2010 and 2014
- 5762 km survey effort
- 303 porpoises sighted

harbour porpoise study–video survey

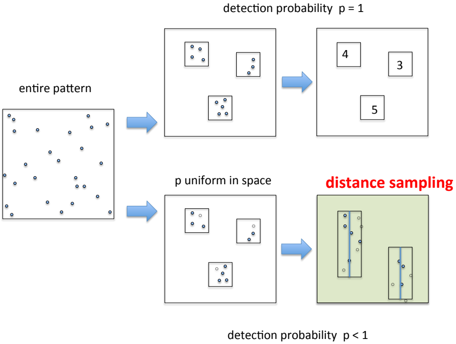


video survey data

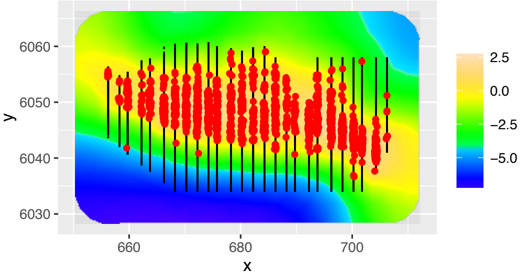
- conducted in August and September 2010 and 2014
- 5762 km survey effort
- 303 porpoises sighted

ecological surveys...

Area of interest is too big to sample entirely.

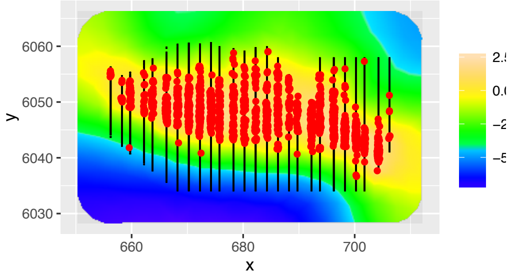


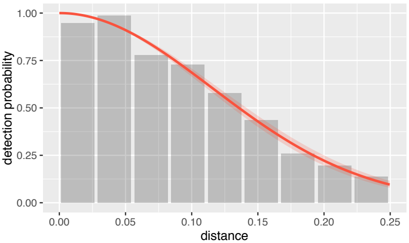
Scottish windfarm survey



distance sampling data

Scottish windfarm survey





[example data set in `inlabru`–more about this later...]

this talk

this talk

- spatial point process modelling and observation processes – in ecology
- `inlabru`

this talk

- spatial point process modelling and observation processes – in ecology
- `inlabru`
- Scottish drinks

observation processes...

ecological research – interested in individuals (in space and time)

observation processes...

ecological research – interested in individuals (in space and time)

in practice:

- need to gain information on individuals – given practical limitations

observation processes...

ecological research – interested in individuals (in space and time)

in practice:

- need to gain information on individuals – given practical limitations

$\Rightarrow$ specific observation process

observation processes...

ecological research – interested in individuals (in space and time)

in practice:

- need to gain information on individuals – given practical limitations

$\Rightarrow$ specific observation process

$\Rightarrow$ specific data structure

observation processes...

ecological research – interested in individuals (in space and time)

in practice:

- need to gain information on individuals – given practical limitations

$\Rightarrow$ specific observation process

$\Rightarrow$ specific data structure

$\Rightarrow$ specific statistical methodology

observation processes...

ecological research – interested in individuals (in space and time)

in practice:

- need to gain information on individuals – given practical limitations

$\Rightarrow$ specific observation process

$\Rightarrow$ specific data structure

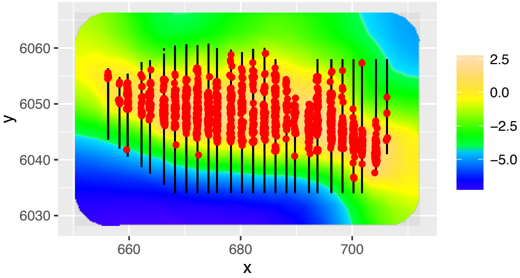
$\Rightarrow$ specific statistical methodology

here:

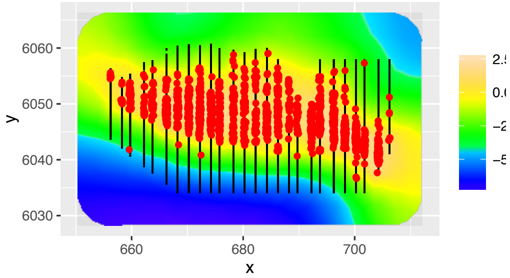
- "think" in terms of the underlying structure, the point process
- observation process is operation on the underlying data structure

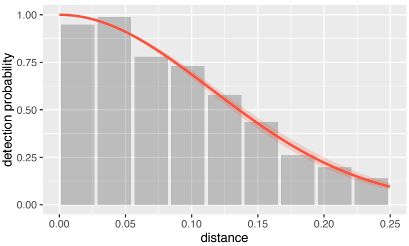
$\Rightarrow$ more general methodology and software

distance sampling data



distance sampling data

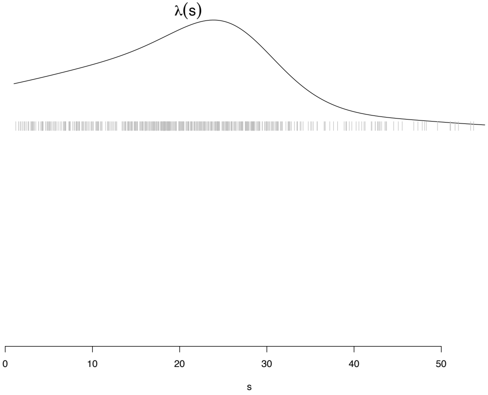




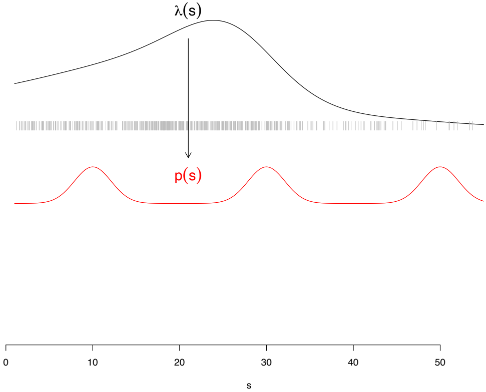
thinned point process!

distance sampling...

thinned point process

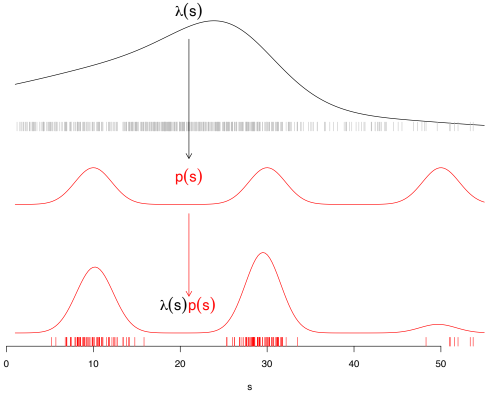


thinned point process

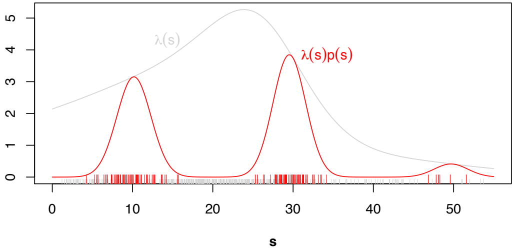


distance sampling...

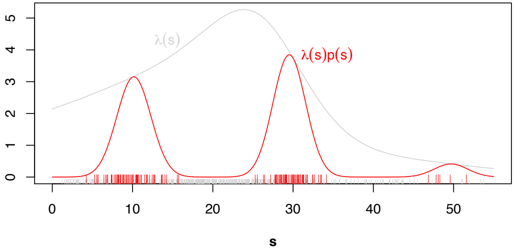
thinned point process



distance sampling...



distance sampling...



Observations are from a thinned Poisson process with intensity $\lambda(\boldsymbol{s})p(\boldsymbol{s})$

example...

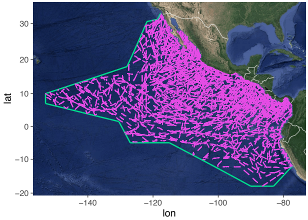
- large scale line-transect cetacean survey in the eastern tropical Pacific Ocean (ETP) between 1986 and 2007
- area of 21.353 million square kilometers ($>$ twice the size of Europe!) was surveyed (transects)
- blue whale sightings

example...

- large scale line-transect cetacean survey in the eastern tropical Pacific Ocean (ETP) between 1986 and 2007
- area of 21.353 million square kilometers ($>$ twice the size of Europe!) was surveyed (transects)
- blue whale sightings

linear predictor depends on:

- (hazard rate) detection function
- (SPDE-based) model for animal intensity
- integration scheme accounts for observation process



Yuan et al. 2016, Bachl et al. in preparation

distance sampling... nice...

- spatio-temporal point process model

distance sampling... nice...

- spatio-temporal point process model
- preserving sighting locations

distance sampling... nice...

- spatio-temporal point process model
- preserving sighting locations
- models the effect of covariates continuously in space
- models spatial structure that cannot be explained by covariates

distance sampling... nice...

- spatio-temporal point process model
- preserving sighting locations
- models the effect of covariates continuously in space
- models spatial structure that cannot be explained by covariates
- elegant, integrated approach
- implemented in *inlabru*

software...

inlabru

software...

`inlabru`

- *takes observation process into account*

software...

inlabru

- *takes observation process into account*
- *makes INLA more accessible*

software...

inlabru

- *takes observation process into account*
- *makes INLA more accessible*
- *wrapper around* `R-INLA` *+ extra functionality*

inlabru – what can it do?

inlabru – what can it do?

fit log Gaussian Cox processes using INLA

`inlabru` – what can it do?

fit log Gaussian Cox processes using INLA – **conveniently**

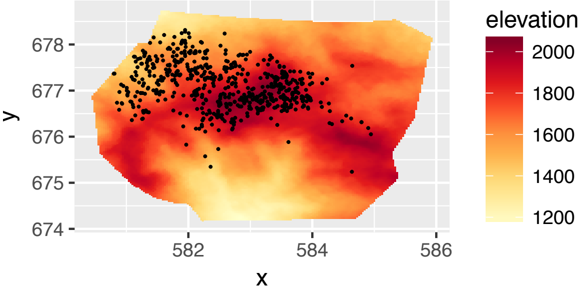
`inlabru` – what can it do?

fit log Gaussian Cox processes using INLA – **conveniently**

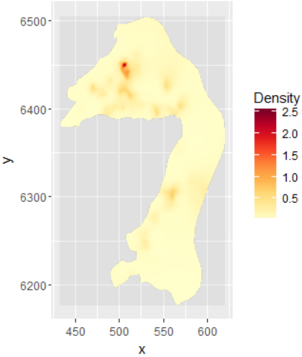
$\Rightarrow$ continuous space

fit log Gaussian Cox processes using INLA – **conveniently**

⇒ continuous space



harbour porpoise study



results

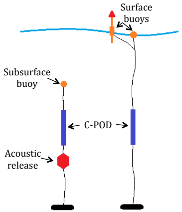
- fine scale clustering apparent
- suggests animals occur in groups

harbour porpoise study–c-pods

harbour porpoise study II

passive acoustic monitoring

- hydrophone detects cetacean vocalisation (place and time)
- harbour porpoise vocalise continuously – clicks and buzzes
- long time series data ($>$ 4 months)
- **not** point pattern data!



C-POD Mooring System

harbour porpoise study–c-pods

harbour porpoise study II

passive acoustic monitoring

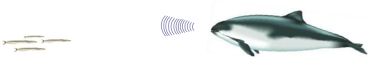
- hydrophone detects cetacean vocalisation (place and time)
- harbour porpoise vocalise continuously – clicks and buzzes
- long time series data ($>$ 4 months)
- **not** point pattern data!

harbour porpoise study–c-pods

harbour porpoise study II

passive acoustic monitoring

- hydrophone detects cetacean vocalisation (place and time)
- harbour porpoise vocalise continuously – clicks and buzzes
- long time series data ($>$ 4 months)
- **not** point pattern data!

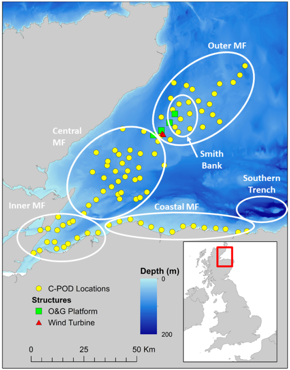


harbour porpoise study–c-pods

harbour porpoise study II

passive acoustic monitoring

- hydrophone detects cetacean vocalisation (place and time)
- harbour porpoise vocalise continuously – clicks and buzzes
- long time series data ($>$ 4 months)
- **not** point pattern data!



harbour porpoise study–c-pods

results

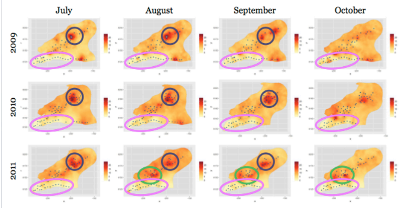
- detection positive hours – seasonal trends

⇒ changes in food availability/competition

- proportion of clicks that are buzzes

⇒ overall distribution different than that of foraging buzzes

⇒ changes in behaviour between different habitats



harbour porpoise study–c-pods

results

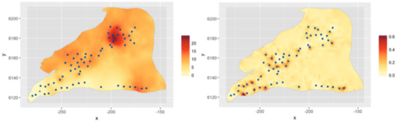
- detection positive hours – seasonal trends

⇒ changes in food availability/competition

- proportion of clicks that are buzzes

⇒ overall distribution different than that of foraging buzzes

⇒ changes in behaviour between different habitats



harbour porpoise study–c-pods

results

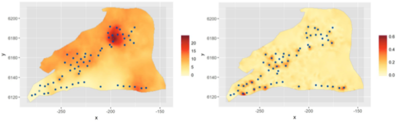
- detection positive hours – seasonal trends

$\Rightarrow$ changes in food availability/competition

- proportion of clicks that are buzzes

$\Rightarrow$ overall distribution different than that of foraging buzzes

$\Rightarrow$ changes in behaviour between different habitats



$\Rightarrow$ implications for Marine Protected Areas

distance sampling revisited...

Or: what I didn't tell you...

distance sampling revisited...

Or: what I didn't tell you...

- ETP study; other species

Or: what I didn't tell you...

- ETP study; other species
- striped dolphins – group size strongly varies among groups; size varies in space

$\Rightarrow$ larger groups are more easily detected

distance sampling revisited...

Or: what I didn't tell you...

- ETP study; other species
- striped dolphins – group size strongly varies among groups; size varies in space

⇒ larger groups are more easily detected

- also: we used a really boring (non-flexible) detection function...

distance sampling revisited...

Or: what I didn't tell you...

- ETP study; other species
- striped dolphins – group size strongly varies among groups; size varies in space

⇒ larger groups are more easily detected

- also: we used a really boring (non-flexible) detection function...

⇒ assumption that log intensity has to be linear in all latent terms no longer a good idea...

distance sampling revisited...

group size:

- detection function depends on group size (a "mark", m): $p(\boldsymbol{s}, m)$
- distribution of group sizes as function of space, $g(m|\boldsymbol{s})$
- joint point process intensity $\lambda(\boldsymbol{s})g(m|\boldsymbol{s})p(\boldsymbol{s}, m)$

distance sampling revisited...

group size:

- detection function depends on group size (a "mark", m): $p(\boldsymbol{s}, m)$
- distribution of group sizes as function of space, $g(m|\boldsymbol{s})$
- joint point process intensity $\lambda(\boldsymbol{s})g(m|\boldsymbol{s})p(\boldsymbol{s}, m)$

- **approach**: linearise non-linear terms through Taylor expansion (iterative)

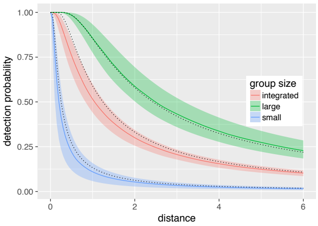
distance sampling revisited...

group size:

- detection function depends on group size (a “mark”, m): $p(\boldsymbol{s}, m)$
- distribution of group sizes as function of space, $g(m|\boldsymbol{s})$
- joint point process intensity $\lambda(\boldsymbol{s})g(m|\boldsymbol{s})p(\boldsymbol{s}, m)$

- **approach**: linearise non-linear terms through Taylor expansion (iterative)

striped dolphins:



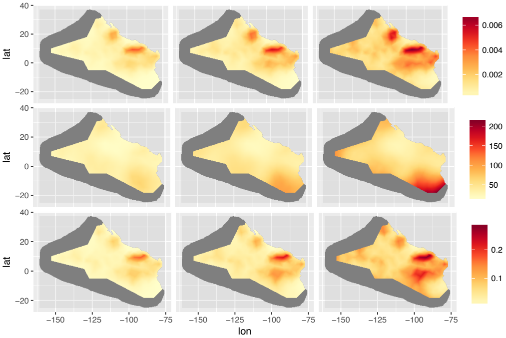
distance sampling revisited...

striped dolphins: groups size varies in space

distance sampling revisited...

striped dolphins: groups size varies in space

- dolphin group intensity (top row)
- expected group size (middle row)
- single animal intensity (bottom row)



left, middle and right column show the 2.5, 50 and 97.5 percent quantiles, respectively

ok...

for distance sampling we can now

- have flexible detection functions
- make detection dependent on marks

ok...

for distance sampling we can now

- have flexible detection functions
- make detection dependent on marks

BUT: what about if you are not interested in distance sampling...?

ok...

for distance sampling we can now

- have flexible detection functions
- make detection dependent on marks

BUT: what about if you are not interested in distance sampling...?

- ecologists
- general applied users

ok...

for distance sampling we can now

- have flexible detection functions
- make detection dependent on marks

BUT: what about if you are not interested in distance sampling...?

- ecologists
- general applied users
- INLA users
- point process people...

applied users

- convenient integrated fitting of distance sampling models

applied users

- convenient integrated fitting of distance sampling models
- user-friendly software *inlabru*

applied users

- convenient integrated fitting of distance sampling models
- user-friendly software *inlabru*
- other observation processes may be seen as different types of “thinnings”

$\Rightarrow$ unified approach, general software

applied users

- convenient integrated fitting of distance sampling models
- user-friendly software *inlabru*
- other observation processes may be seen as different types of "thinnings"

$\Rightarrow$ unified approach, general software

- can fit general spatial models (no thinning) elegantly

- can fit general spatial models (no thinning) elegantly (see next page)

spatial modellers

- can fit general spatial models (no thinning) elegantly (see next page)
- dropping linearity assumption – applicable in many contexts
- complex marked point processes

spatial modellers

- can fit general spatial models (no thinning) elegantly (see next page)
- dropping linearity assumption – applicable in many contexts
- complex marked point processes

- can interpret (univariate) function as one-dimensional LGCP

$\Rightarrow$ use *inlabru* for function estimation (detection function, pdfs, K-functions...)

inlabru, a friendlier INLA interface

R-INLA

```
A.data <- inla.spde.make.A(...)
A.pred <- inla.spde.make.A(...)
stack.data <- inla.stack(data=..., A=list(A.data, ...), effects=...)
stack.pred <- inla.stack(data=..., A=list(A.pred, ...), effects=...)
stack <- inla.stack(stack.data, stack.pred)
formula <- y ~ ... + f(field, model=spde)
result <- inla(...)
## Linear prediction:
prediction <- result$summary.fitted.values[some.indices, "mean"]
```

http://inlabru.org

```
components <- ~ ... + field(map=coordinates, model=spde)
formula <- y ~ ... + field
result <- bru(...)
result <- lgcp(...)
## Non-linear prediction (via direct posterior sampling)
prediction <- predict(..., cos(field))
## Extra: non-linear formulas and marked LGCP capabilities
formula <- y ~ field1 * exp(field2)
formula <- coordinates + size ~ field1 +
             dnorm(size, field2, sd=exp(theta), log=TRUE)
```

that Scottish drink...

that Scottish drink...

that Scottish drink...

that Scottish drink...