

Presentation 2:  
Introduction to *PointedSDMs*  
'ISEC 2024 – Swansea'

Philip Mostert

Norwegian University of Science and Technology

13 July 2024

# Introduction

- ▶ Previous presentation introduced *ISDMs*.
- ▶ In this presentation we will present an *R* package to estimate these models.

# Introduction to integrated modelling

- ▶ Species distribution modelling has many applications in ecology and is now a mature discipline.
- ▶ Previously, the approach used was to favour one dataset, and either discard the other datasets or use them in secondary analyses.
- ▶ However integrated species distribution models (ISDMs) have emerged – a method to model multiple datasets in a common statistical framework.

# Process model

The underlying process model is a log-Gaussian cox process (LGCP) with a spatially varying intensity function,  $\lambda(s) = \exp\{\eta(s)\}$ , where:

$$\eta(s) = \alpha_0 + \sum_{k=1}^p \beta_k X_k(s) + \zeta(s),$$

where:

- ▶  $\alpha_0$  is an intercept term,
- ▶  $\beta_k$  is the  $k^{th}$  coefficient for the  $k^{th}$  environmental covariate, which varies in space (denoted by the index,  $s$ ),
- ▶  $\zeta(s)$  is a zero-mean Gaussian random field determined by a Matèrn covariance function.

# Observation models

- Observation models are chosen based on the underlying sampling protocols of a given dataset.

Data type	Distribution
Present absent	Binomial (with a <i>cloglog link</i> )
Present only	Thinned Poisson
Count data	Poisson
Marks/traits	Variety

## *PointedSDMs R package*

- ▶ Package is designed to simplify the construction of ISDMs using a joint-likelihood framework.
- ▶ Wrapper around *R-INLA* and *inlabru* packages.
- ▶ Manuscript available in *Methods in Ecology and Evolution*: Mostert and O'Hara (2023)

# Installation of the package

- ▶ Install from either *CRAN* or *Github*.

```
install.packages("PointedSDMs")
```

*# or*

```
devtools::install_github("PhilipMostert/PointedSDMs")
```

# Package structure

- ▶ Initialize model,
- ▶ Specify model,
- ▶ Estimate model,
- ▶ Predict and plot,
- ▶ Cross-validation.



Initializing model

# startISDM

```
args(startISDM)
```

```
## function (... , spatialCovariates = NULL, Projection  
  , Mesh, IPS = NULL,  
##      Boundary = NULL, pointCovariates = NULL, Offset  
  = NULL, pointsIntercept = TRUE,  
##      pointsSpatial = "copy", responseCounts = "  
  counts", responsePA = "present",  
##      trialsPA = NULL, temporalName = NULL, Formulas  
  = list(covariateFormula = NULL,  
##      biasFormula = NULL))  
## NULL
```

## startSpecies

```
args(startSpecies)
```

```
## function (... , spatialCovariates = NULL, Projection
, Mesh, speciesSpatial = "replicate",
##   speciesIntercept = TRUE, speciesEnvironment =
TRUE, speciesName,
##   IPS = NULL, Boundary = NULL, pointCovariates =
NULL, Offset = NULL,
##   pointsIntercept = TRUE, pointsSpatial = "copy",
responseCounts = "counts",
##   responsePA = "present", trialsPA = NULL,
temporalName = NULL,
##   Formulas = list(covariateFormula = NULL,
biasFormula = NULL))
## NULL
```

## startMarks

```
args(startMarks)
```

```
## function (... , spatialCovariates = NULL, Projection
, Mesh, IPS = NULL,
##      Boundary = NULL, markNames = NULL, markFamily =
      NULL, marksSpatial = TRUE,
##      pointCovariates = NULL, pointsIntercept = TRUE,
      marksIntercept = TRUE,
##      Offset = NULL, pointsSpatial = "copy",
      responseCounts = "counts",
##      responsePA = "present", trialsPA = NULL,
      trialsMarks = NULL,
##      temporalName = NULL, Formulas = list(
      covariateFormula = NULL,
##      biasFormula = NULL))
## NULL
```

Specifying model

## Available slot functions

Name	Use
<code>\$plot()</code>	Plot data
<code>\$addBias()</code>	Add second spatial effect
<code>\$updateFormula()</code>	Update a likelihood's formula
<code>\$changeComponents()</code>	Update model components
<code>\$priorsFixed()</code>	Specify priors for the fixed effects
<code>\$specifySpatial()</code>	Specify the spatial effects
<code>\$changeLink()</code>	Change the link function of a process
<code>\$spatialBlock()</code>	Initiate spatial-block cross-validation
<code>\$addSamplers()</code>	Change the integration domain for a likelihood
<code>\$specifyRandom()</code>	Specify the priors for other random effects

## Information about slot functions

- ▶ Obtain documentation through the `.$help()` function.

```
obj <- startISDM(...)  
obj.$help()
```

Estimate the model



# fitISDM

```
args(fitISDM)
```

```
## function (data, options = list())  
## NULL
```

## Post-estimation

## Predict and plot

```
Predictions <- predict(model,  
  ...)  
plot(Predictions)
```

## Model evaluation

# Spatial block

```
args(blockedCV)
```

```
## function (data, options = list())  
## NULL
```

# Leave-one-out

```
args(datasetOut)
```

```
## function (model, dataset, predictions = TRUE)  
## NULL
```

Quick example

# Introduction

- ▶ We will primarily use the *Setophaga* dataset throughout this workshop
- ▶ Contains information from three datasets:
  - ▶ *eBird*
  - ▶ *Pennsylvania Breeding Bird Atlas*
  - ▶ *North American Breeding Bird Survey*



## startISDM

```
caerulescensModel <- startISDM(eBird,  
  BBS, BBA, Boundary = PA,  
  Projection = proj,  
  Mesh = mesh, responsePA = "NPres",  
  responseCounts = "Counts",  
  spatialCovariates = covariates,  
  Formulas = list(covariateFormula = ~elevation +  
    I(elevation^2) +  
    canopy + I(canopy^2)))
```

## Specify Model

```
caerulescensModel$specifySpatial(sharedSpatial = TRUE,  
  prior.sigma = c(1,  
    0.1), prior.range = c(15,  
    0.1))
```

```
caerulescensModel$addBias(datasetNames = "eBird")  
caerulescensModel$specifySpatial(Bias = TRUE,  
  prior.sigma = c(1,  
    0.1), prior.range = c(15,  
    0.1))
```

```
caerulescensModel$priorsFixed(Effect = "Intercept",  
  mean.linear = 0,  
  prec.linear = 0.1)
```

## fitISDM, predict and plot

```
caerulescensEst <- fitISDM(data = caerulescensModel,  
  options = modelOptions)  
  
caerulescensPredictions <- predict(caerulescensEst,  
  data = fm_pixels(mesh = mesh,  
    mask = PA), spatial = TRUE,  
  n.samples = 1000)  
  
plot(caerulescensPredictions,  
  variable = c("mean",  
    "sd"))
```

## Next steps

- ▶ Questions and discussions.
- ▶ Pull *Github* repository:  
*PhilipMostert/PointedSDMs\_Workshop*.
- ▶ We will through *Vignette 1: Basic model* together.

# References

- Mostert, Philip S, and Robert B O'Hara. 2023. "PointedSDMs: An R Package to Help Facilitate the Construction of Integrated Species Distribution Models." *Methods in Ecology and Evolution* 14 (5): 1200–1207. <https://doi.org/10.1111/2041-210X.14091>.