

ISEC – Montpellier – July, 4th 2014

Hierarchical Bayesian species distribution models with the **hSDM** R Package



Ghislain Vieilledent^{*,1} Cory Merow² Jérôme Guélat³
 Andrew M. Latimer⁴ Marc Kéry³
 Alan E. Gelfand⁵ Adam M. Wilson⁶ Frédéric Mortier¹
 and John A. Silander Jr.²

[1] **Cirad France**, [2] **University of Connecticut USA**, [3] **Swiss Ornithological Institute Switzerland**,
[4] **University of California USA**, [5] **Duke University USA**, [6] **Yale University USA**

1 Introduction

- Species distribution models
- Issues : imperfect detection and spatial correlation
- Available softwares

2 hSDM R package

- Package main characteristics
- Parameter inference

3 Examples

- N-mixture iCAR model
- Binomial iCAR model with large data-set

4 Recommendations

- “statistical machismo” ?
- hSDM and applied research in ecology

1 Introduction

- Species distribution models
- Issues : imperfect detection and spatial correlation
- Available softwares

2 hSDM R package

- Package main characteristics
- Parameter inference

3 Examples

- N-mixture iCAR model
- Binomial iCAR model with large data-set

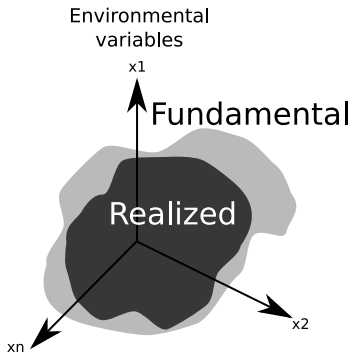
4 Recommendations

- “statistical machismo” ?
- hSDM and applied research in ecology

SDM definition

Objectives

- Identifying the suitable habitat for species persistence
- Reference : species niche (Hutchinson 1957)
- Representing this habitat spatially (maps)
- Applications : conservation biology



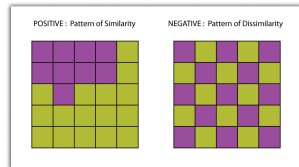
Imperfect detection

- Species is not observed perfectly (“Now you see me, now you don’t”)
- Detection probability < 1
- Treating observations as the “true” species distribution might lead to completely wrong habitat models
- See **Lahoz-Monfort et al.** 2014 *Global Ecology and Biogeography*



Spatial autocorrelation

- Most species present geographical patchiness
- Positive spatial correlation
- Causes :
 - Exogeneous environmental factors (climate, soil)
 - Endogeneous biotic processes (dispersal, migration)
- Ignoring spatial correlation may lead to biased conclusions about ecological relationships
- See **Lichstein et al. 2002**
Ecological Monographs



Site-occupancy and N-mixture models

Site-occupancy model (occurrence data)

Ecological process :

$$Z_i \sim \text{Bernoulli}(\theta_i)$$

$$\text{logit}(\theta_i) = X_i \beta$$

Observation process :

$$y_{it} \sim \text{Bernoulli}(\delta_{it} \times Z_i)$$

$$\text{logit}(\delta_{it}) = W_{it} \gamma$$

N-mixture model (count data)

Ecological process :

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = X_i \beta$$

Observation process :

$$y_{it} \sim \text{Binomial}(N_i, \delta_{it})$$

$$\text{logit}(\delta_{it}) = W_{it} \gamma$$

Repeated observations at particular sites (multiple visits)

CAR process

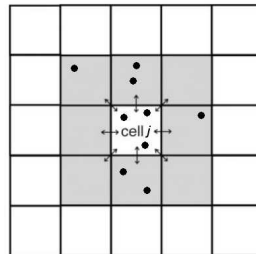
intrinsic CAR

$$p(\rho_j | \rho_{j'}) \sim \text{Normal}(\mu_j, V_\rho / n_j)$$

μ_j : mean of $\rho_{j'}$ in the neighborhood of j .

V_ρ : variance of the spatial random effects.

n_j : number of neighbors for cell j .



CAR process

Site-occupancy model (occurrence data)

Ecological process :

$$Z_i \sim \text{Bernoulli}(\theta_i)$$

$$\text{logit}(\theta_i) = X_i\beta + \rho_{j(i)}$$

Observation process :

$$y_{it} \sim \text{Bernoulli}(\delta_{it} \times Z_i)$$

$$\text{logit}(\delta_{it}) = W_{it}\gamma$$

N-mixture model (count data)

Ecological process :

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = X_i\beta + \rho_{j(i)}$$

Observation process :

$$y_{it} \sim \text{Binomial}(N_i, \delta_{it})$$

$$\text{logit}(\delta_{it}) = W_{it}\gamma$$

Available softwares

Softwares for mixture models

PRESENCE, MARK, E-SURGE, unmarked, stocc, JAGS, stan, WinBUGS, OpenBUGS

Softwares for spatial autocorrelation

OpenBUGS, WinBUGS, BayesX, stocc, CARBayes, R-INLA, spatcounts, OpenBUGS, spdep, CARramps, spBayes

Mixture models + spatial autocorrelation

Very few softwares available and limitations : **OpenBUGS, WinBUGS** (might be slow), **stocc** (probit and binary data only).

1 Introduction

- Species distribution models
- Issues : imperfect detection and spatial correlation
- Available softwares

2 hSDM R package

- Package main characteristics
- Parameter inference

3 Examples

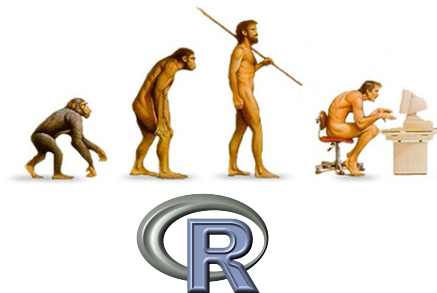
- N-mixture iCAR model
- Binomial iCAR model with large data-set

4 Recommendations

- “statistical machismo” ?
- hSDM and applied research in ecology

Package main characteristics

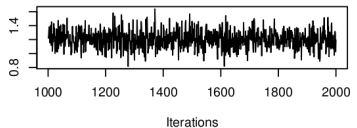
- **hSDM** : R “user-friendly” package
- Mixture models : **site-occupancy**, **N-mixture**, but also **ZIB** and **ZIP** models
- Including a spatial autocorrelation process (**iCAR**)
- Web-site : <http://hSDM.sf.net>
- Vignette with several examples on simulated and real data-sets



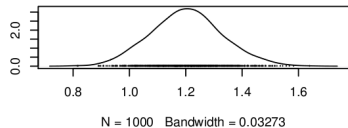
Parameter inference

- hierarchical Bayesian models
- MCMC methods (no approximation of the posterior)
- adaptive Metropolis within Gibbs (**“efficient”**)
- written in pure **C** code (**“fast”**)
- source code available through git on Sourceforge

Trace of beta0



Density of beta0



1 Introduction

- Species distribution models
- Issues : imperfect detection and spatial correlation
- Available softwares

2 hSDM R package

- Package main characteristics
- Parameter inference

3 Examples

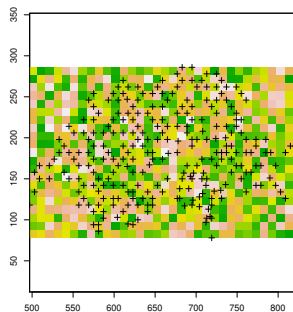
- N-mixture iCAR model
- Binomial iCAR model with large data-set

4 Recommendations

- “statistical machismo” ?
- hSDM and applied research in ecology

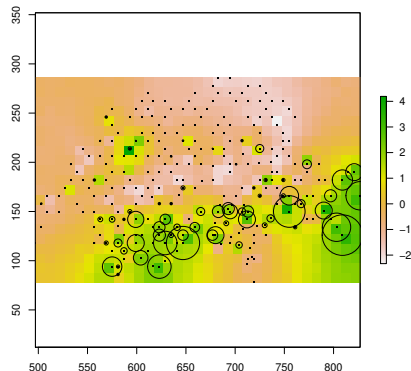
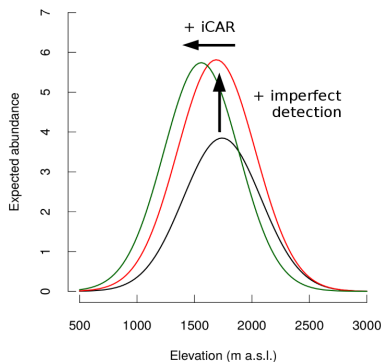
N-mixture iCAR model

- **Kéry & Royle** 2010, *Journal of Animal Ecology*
- Abundance of the Willow tit (*Poecile montanus*)
- Switzerland
- 264 sites with 2 or 3 visits
- 10×10 km cells



N-mixture iCAR model

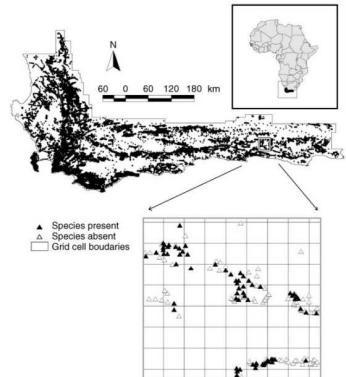
- Models of increasing complexity
- (1) Poisson, (2) N-mixture and (3) N-mixture + iCAR



Spatial random effects (raster) and
observed abundance (circles)

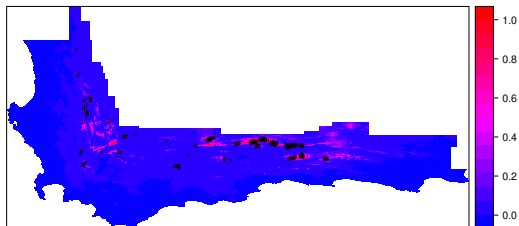
Binomial iCAR model with large data-set

- **Latimer et al. 2006, *Ecological Applications***
- Occurrence of *Protea punctata*
- Cap Floristic Region (South Africa)
- 36909 1'×1' grid cells
- Too many data to use ★**BUGS** programs

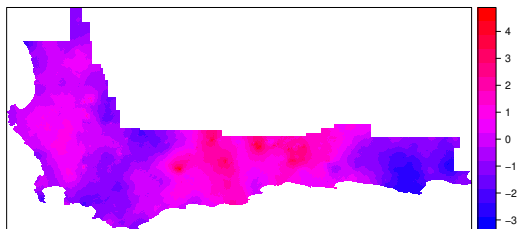


Binomial iCAR model with large data-set

Probability of
presence



Spatial random
effects



1 Introduction

- Species distribution models
- Issues : imperfect detection and spatial correlation
- Available softwares

2 hSDM R package

- Package main characteristics
- Parameter inference

3 Examples

- N-mixture iCAR model
- Binomial iCAR model with large data-set

4 Recommendations

- “statistical machismo” ?
- hSDM and applied research in ecology

“statistical machismo” ?

Following **Brian McGill's** post

- Bonferroni corrections
- Phylogenetic corrections
- Spatial regression
- Detection error
- Bayesian methods

“statistical machismo” ?

Following **Brian McGill's** post

- Bonferroni corrections
- Phylogenetic corrections
- Spatial regression
- Detection error
- Bayesian methods

Can we be accused of “statistical machismo” ?



hSDM and applied research in ecology

Some recommendations

- Complex models need **appropriate data** (quantity + structure).
- Beware of over-parametrization and identifiability problems.

Complex models and applied research

- It is true that in some cases, complex models do not significantly improve our ecological knowledge of the species.
- Conservation planning : we need to identify the level of **model complexity** we need to reach in order to take good decisions.

When complex models are necessary ?

hSDM : tool to investigate the situations (species, place) where imperfect detection and spatial correlation lead to significant different results that impact decisions.

... Thank you for attention ...