Introduction to Hierarchical Bayesian Analysis for Ecological Data using WinBUGS

June 29-30th 2014, Montpellier

Eric PARENT

AgroParisTech

UMR Applied Mathematics and Informatics

eric.parent@agroparistech.fr

Etienne RIVOT

Agrocampus Ouest, University of Brittany

UMR Ecology and Ecosystem Health, Fisheries Ecology group

etienne.rivot@agrocampus-ouest.fr

The success story of HBM and BUGS

The success story of HBM

- **HBM have received considerable attention in the last 15 years**
- **Are becoming more and more popular, in all field of applied statistics, in particular in Statistical Ecology**

Clark, J. S. (2005). Why environmental scientists are becoming Bayesians? *Ecology Letters*, *8*, 2–14.

Clark, J. S., & Gelfand, A. E. (2006). A future for models and data in environmental sciences. *Trends in Ecology & Evolution*, *21*(7), 375–380.

Cressie, N. A. C., Calder, C. A., Clark, J. S., Ver Hoeff, J. M., & Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modelling. *Ecological Applications*, *19(3)*, 553–570.

Hobbs, N. T., & Ogle, K. (2011). Introducing data–model assimilation to students of ecology. *Ecological Applications*, *21*(5), 1537–1545.

The success story of HBM

- **The fast and widespread development is intimately linked to the apparition of flexible (and free !) softwares**

  Monte Carlo sampling in posterior distributions

" The genie can not be put back into the bottle. The Bayesian machine, together with MCMC, is arguably the most powerful mechanism never created for processing data and knowledge."

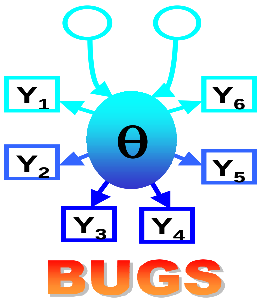
Berger, J. O. (2000). Bayesian analysis: a look at today and thoughts of tomorrow. *Journal of the American Statistical Association*, *95*(452), 1269–1276.

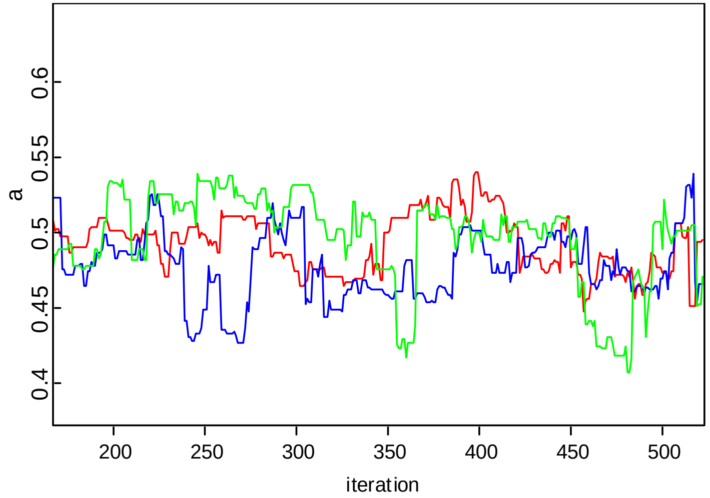
Bayesian computation : a statistical revolution.

Brooks, S. P. (2003).
Phil. Trans. R. Soc. Lond., A, 361, 2681–2697.

The success story of HBM

The “BUGS-boom” generation



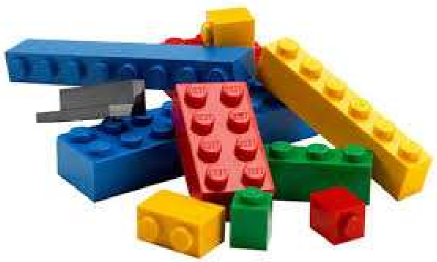


Objectives for the workshop

Objectives

- Introduce to HBM and illustrate the flexibility for statistical ecology
- Learning by the examples, not a lot of theory
- Practice HBM with classical toolboxes (BUGS, JAGS)
- Write your own code
- Start with simple models

  … and complexify little by little



Program

1 Introduction to HBM

2 A first simple model with BUGS (+JAGS / STAN)

Linear regression

3 …

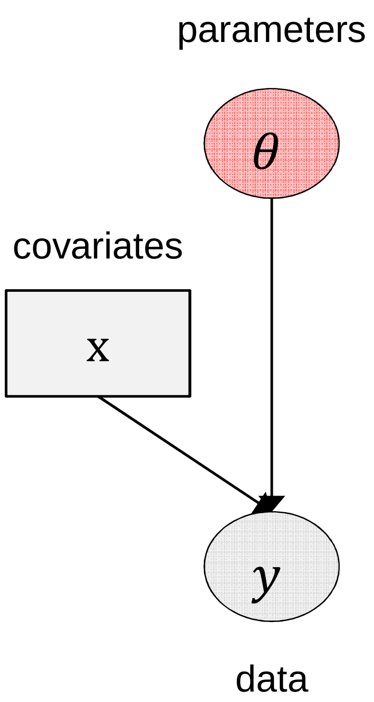
4 A hierarchical model for mark-recapture data

5 A hierarchical model for stock-recruitment data

An introduction to HBM

Statistical inferences
-
Maximum likelihood

Maximum likelihood



Modelling

Define the likelihood $p(y/\theta,x)$

Inference

Explore how $p(y/\theta,x)$ varies in the space of θ

Point estimates

Find the parameters values θ^ that maximize $p(y/\theta,x)$

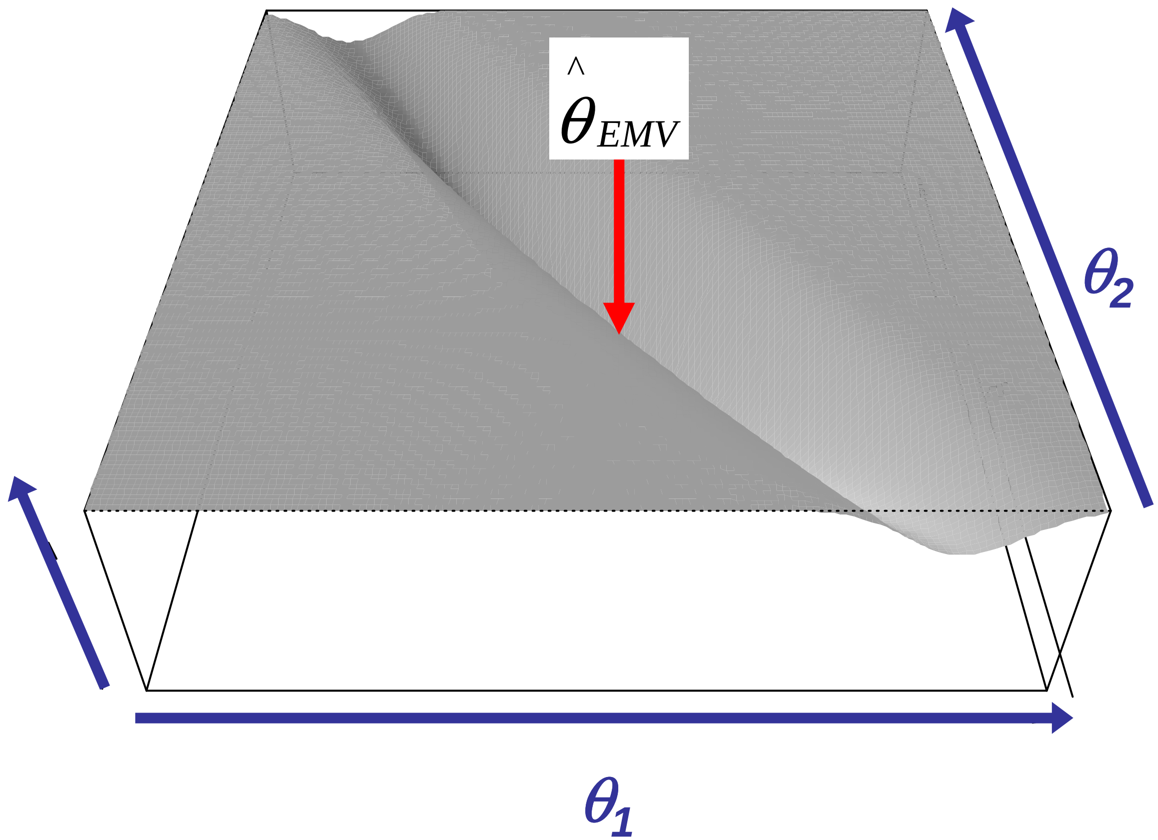
Uncertainty

Explore how $p(y/\theta,x)$ varies in the vicinity of θ^

Maximum likelihood

Objective function

$-\log(p(y/\theta,x))$



Maximum likelihood

Use asymptotic theorems to assess the uncertainty about point estimates

$$\hat{\theta}_{EMV} \xrightarrow[n \to +\infty]{\overset{Loi}{\sim}} N(\theta, \Sigma_{\theta_{EMV}})$$

$$\hat{\Sigma}_{\theta_{EMV}} = \left(I(\hat{\theta}_{EMV}) \right)^{-1}$$

$$\underbrace{I(\theta)}_{Fisher\,Information\,Matrix} = E \underbrace{\left[\frac{\partial^2 - \log(L(y,\theta))}{\partial\theta_i\,\partial\theta_j} \right]}_{H(\theta) = Hessian\,matrix}$$

Statistical inferences

-

Bayesian

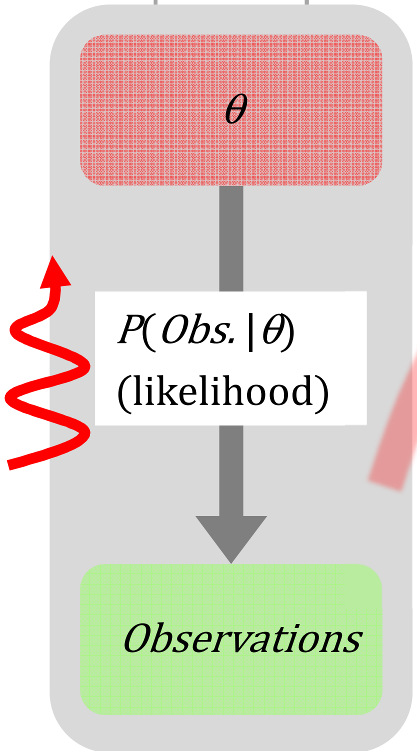
Classical / Bayesian inferences (NON Hierarchical models)

Maximum likelihood

Maximum likelihood estimates

{ }

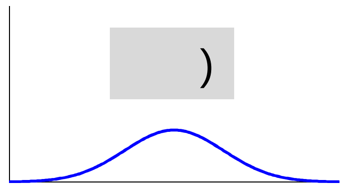
Uncertainty
Fisher Information Matrix



Bayesian

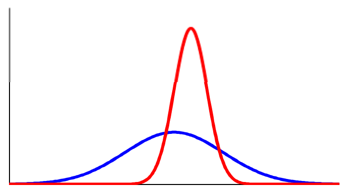
Prior distribution

)



Updating = learning from the data

Posterior distribution



(|) ( )

Hierarchical models

What are Hierarchical Models ?

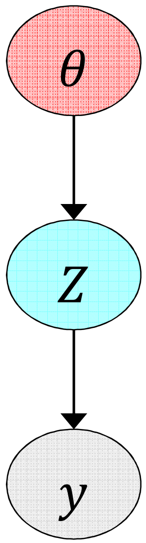
- **HM = Multilevel stochastic models**
- **Not necessarily "Bayesian" !**
- **Useful when the relationship between the data and the parameters can not be written directly as** $p(y\,|\,\theta)$

  → Need « intermediate layer(s) » in the model

  - **to explicitly separate out different sources of stochasticity (e.g. process errors, measurement errors)**
  - **to capture some complex dependence in a hierarchy of scales (among individuals, in space, in time …)**

What are Hierarchical Models ?

HM = Multilevel stochastic models



Parameters

All intermediate process of interest

$p(Z\,|\,\theta)$ *Z = « latent » or « hidden » variables*

Data

$p(y\,|\,Z,\theta)$

What are Hierarchical Models ?

- **A wide class of models**
  - **Random effects models, mixed effects models**
  - **Observation error models**
  - **State-space models**
  - **« Integrated models »**

Examples of Hierarchical models

Random effect models

- **Most often associated with the hierarchical structure of the data**

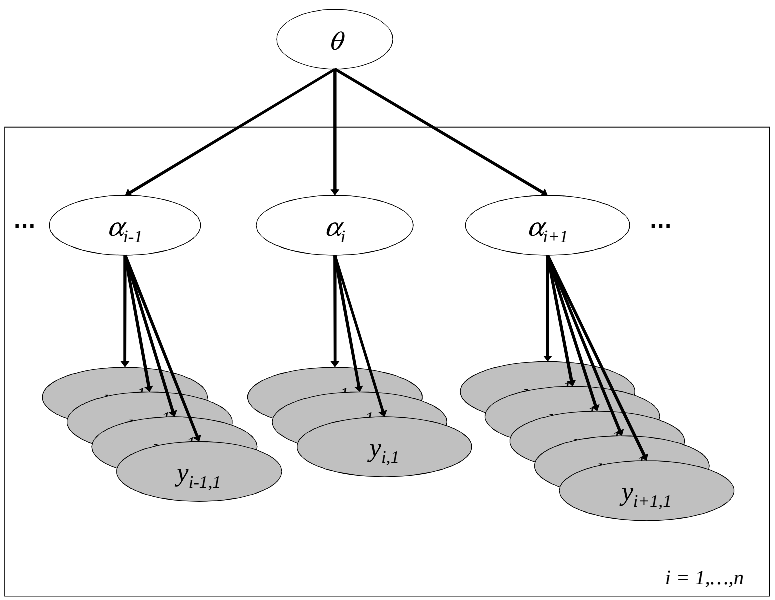
Examples **Children in schools, schools in towns …**
Populations in regions, regions in country …

Repeated measure on individuals (individual heterogeneity)

$$Y_{i,j} = \mu + \alpha_i + \varepsilon_{i,j}$$

Random eff. $\alpha_i \sim N(0, \sigma_\alpha)$

$$\varepsilon_{i,j} \sim N(0, \sigma_i)$$



schools

children

Advantages of Random *versus* fixed effects ?

- **Shrinkage effect**

Estimates of effects of unit *i*, $\hat{\alpha}_i$ are <u>weighted</u> by the number of individuals in units *i* and by their inter-individual variability within unit *i*

Will be shrinked toward the overall mean of the $\hat{\alpha}_i$

The less informative the data, the higher the shrinkage effect

The superiority of shrinked estimates for prediction is well established

Models with observation errors

- **Observation errors**

  - can occur in one or several variables in the model

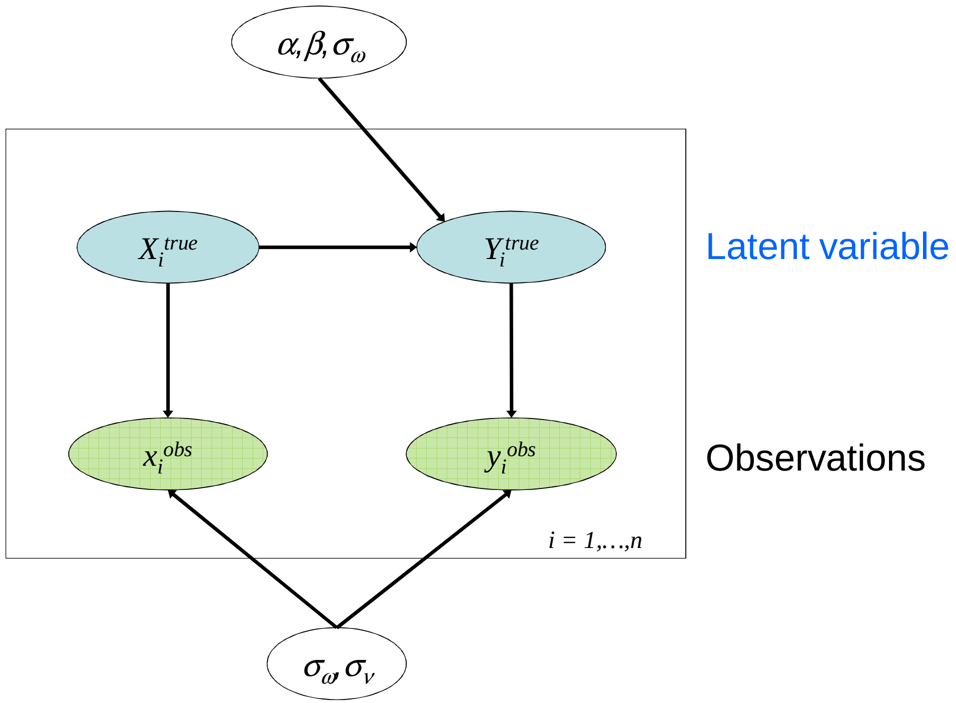
  - can be due to measurement errors, sampling errors …

- **Variable not directly observed are LATENT VARIABLES**

$$Y_i^{true} = \alpha \cdot X_i^{true} + \beta + \underbrace{\varepsilon_i}_{process\ \ errors}$$

$$y_i^{obs} = Y_i^{true} + \underbrace{\omega_i}_{obs\ \ error\ \ on\ \ Y}$$

$$x_i^{obs} = X_i^{true} + \underbrace{\nu_i}_{obs\ \ error\ \ on\ \ X}$$



State-space models

- **A wide class of dynamic models with latent variables (non observed variables)**

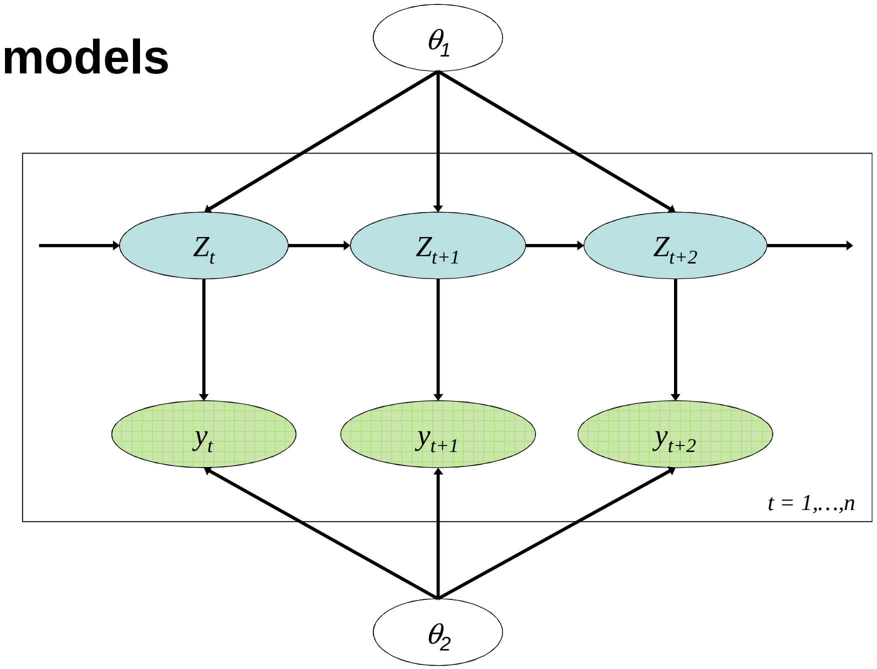
Examples

- **Population dynamics**
- **Hidden Markov Models**
- **Multi-state mark-recapture models**

Process eq. $Z_t \sim f\left(Z_{t-1}, \theta_1, \varepsilon\right)$

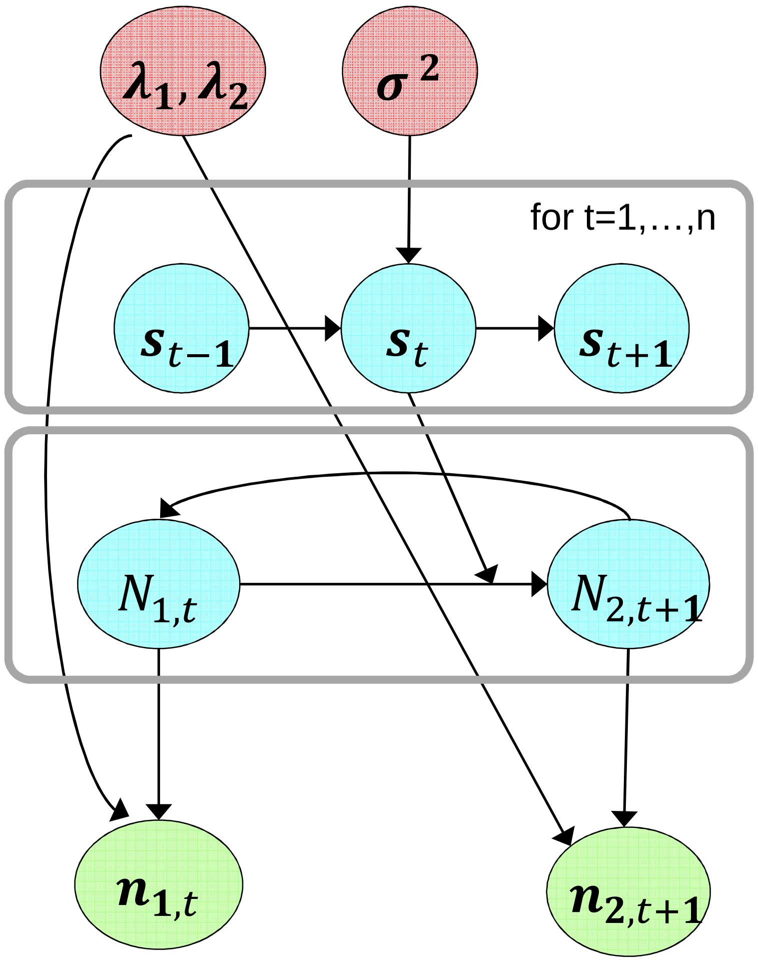
Observation eq. $y_t^{obs} \sim g(Z_t, \theta_2, \omega)$

The time dependence is accounted for in the process layer, and not in the observation layer



A simple life cycle model

Graphic models may help !



Parameters

States

$P(States \mid Parameters)$

$s_{t+1} = s_t + \varepsilon_t \sim N(0, \sigma^2)$ ➞ $P(s_{t+1} | s_t, \sigma^2)$

$N_{2,t+1} \sim Binom(N_{1,t}, s_t)$ ➞ $P(N_{2,t+1} | N_{1,t}, s_t)$

Observations

$P(Observations \mid States, Param.)$

$n_{1,t} \sim Pois(N_{1,t} \cdot \lambda_1)$ ➞ $P(n_{1,t} | N_{1,t}, \lambda_1)$

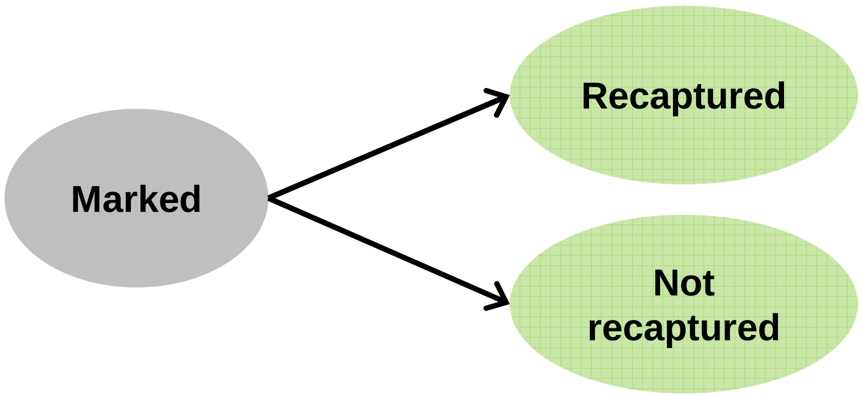
$n_{2,t+1} \sim Pois(N_{2,t+1} \cdot \lambda_2)$ ➞ $P(n_{2,t+1} | N_{2,t+1}, \lambda_2)$

Mark-Recapture model 1/3

- **Estimation of survival rates**

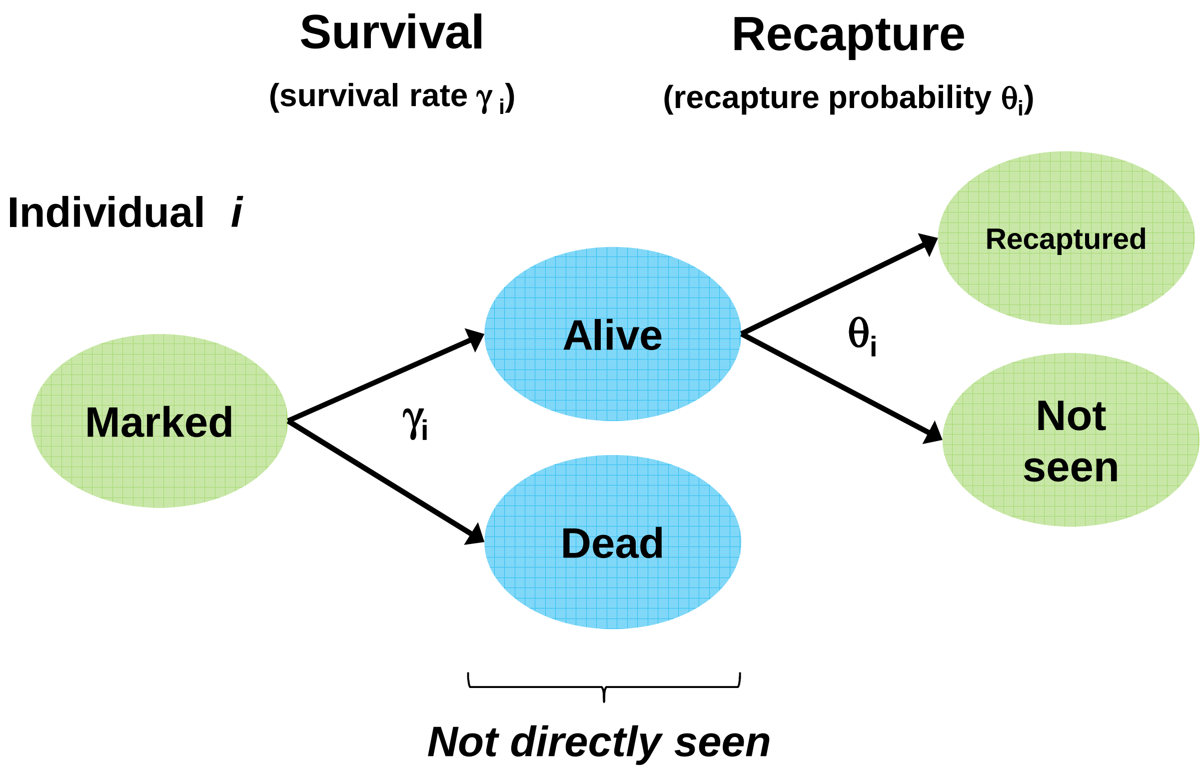
  - How to separate out the stochasticity in the survival process and in the observation process?

  - Variability in survival/recapture rates between individuals ?



Mark-Recapture model 2/3

A joint probability model for process and observations



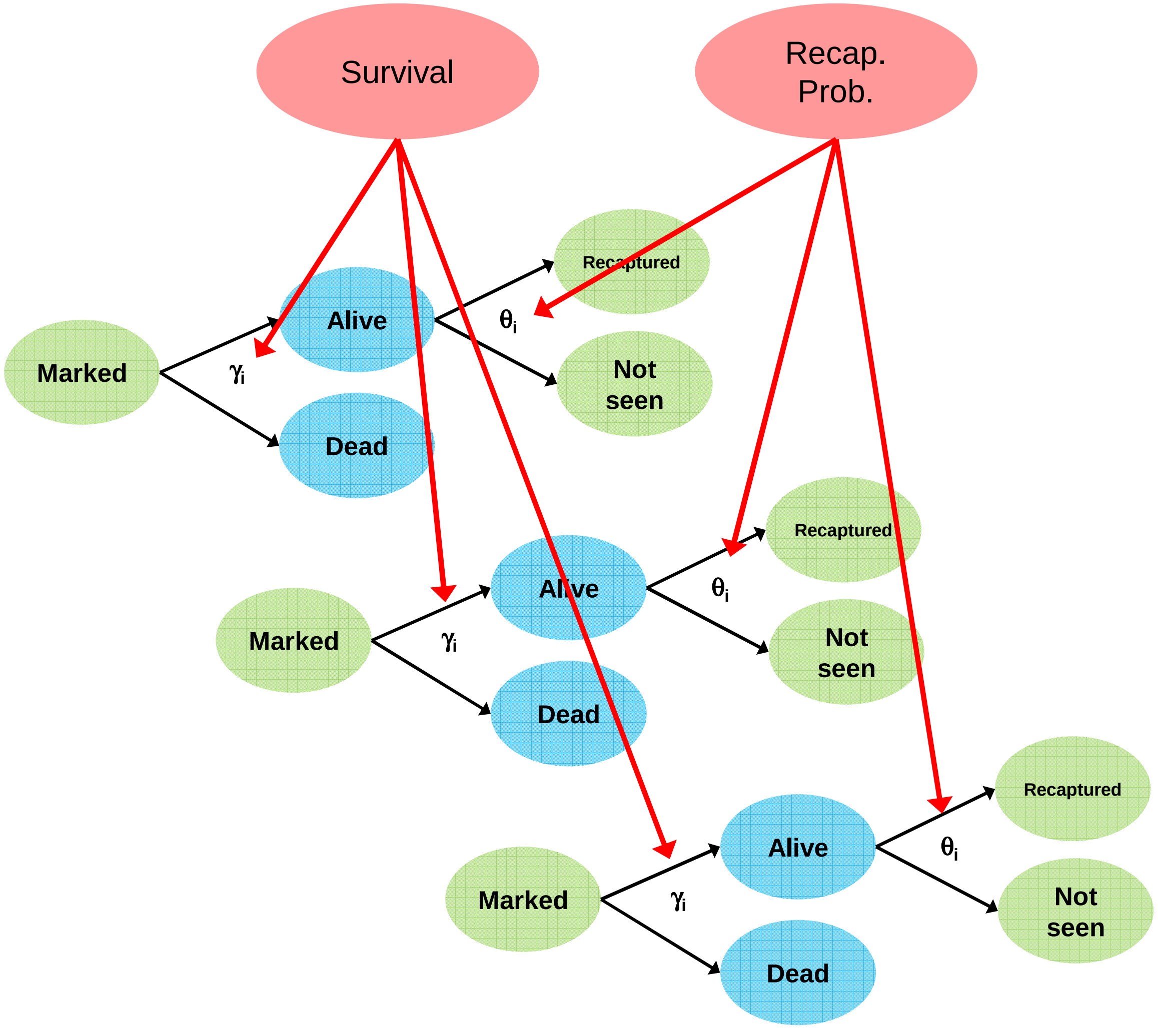
Mark-Recapture model 3/3

State-space model

Individuals are not observed at some recapture event

Random effects

Heterogeneity among individuals



Inferences
on
Bayesian Hierarchical Models

Statistical inferences on Hierarchical models

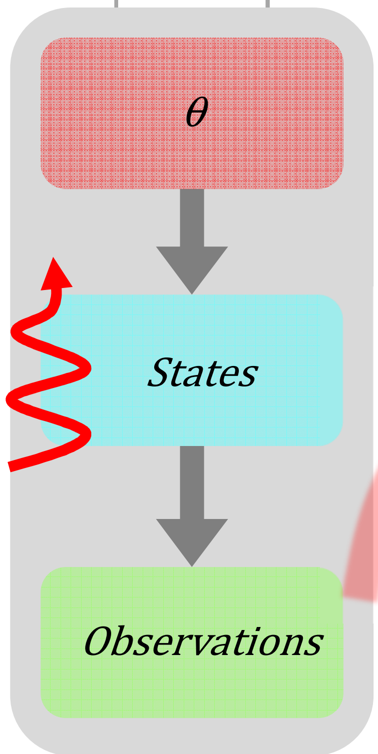
Maximum likelihood	Bayesian

Maximum likelihood estimates

Latent states must be integrated out

$\{\ (\quad |\ )\}=$

$\left\{ \quad (\quad\quad |\ ) \right\}$



Joint prior for (parameters,states)

$(\quad\quad)$

Updating

Joint posterior distribution

$(\quad\quad |\quad)$

Bayesian Hierarchical Model

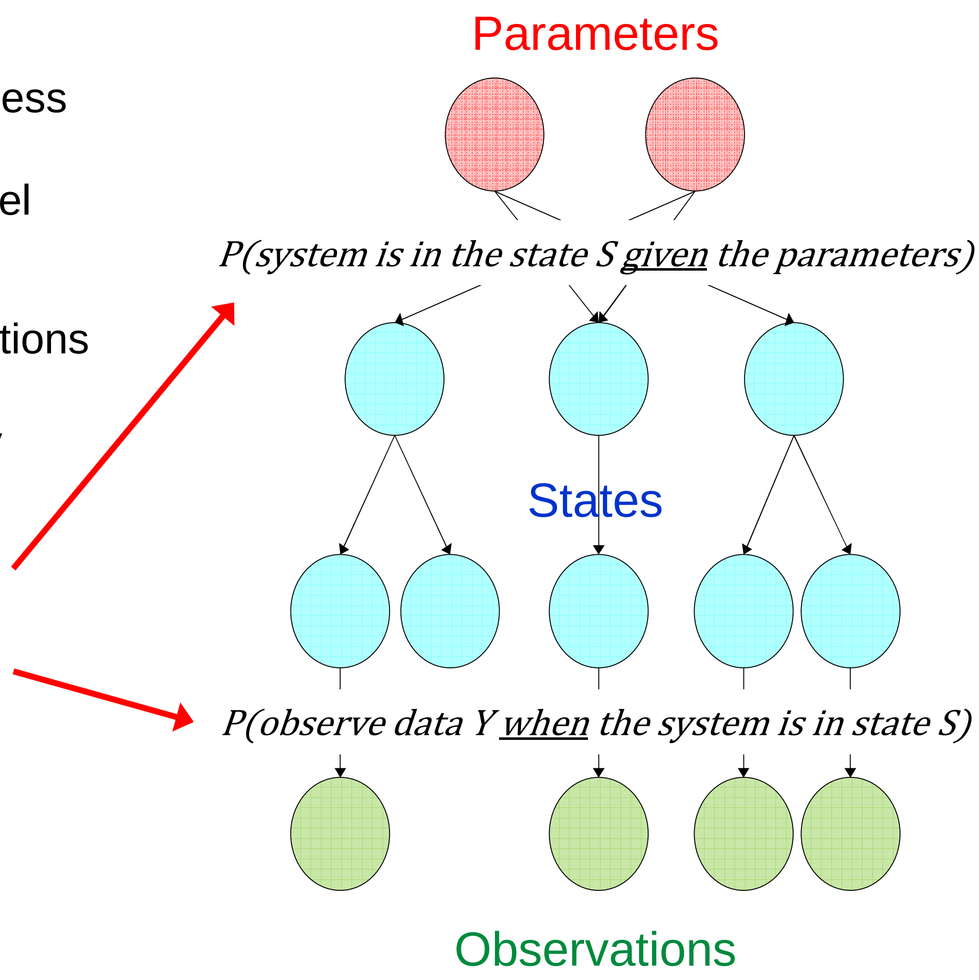
Take-home messages

Hierarchical Models

- Hierarchical Models = Multilevel stochastic models
  - Models for the ecological process and for the observations are integrated within a single model
  - Conditional probability distributions used to model *cause* → *effect* relationships with stochasticity

for the process

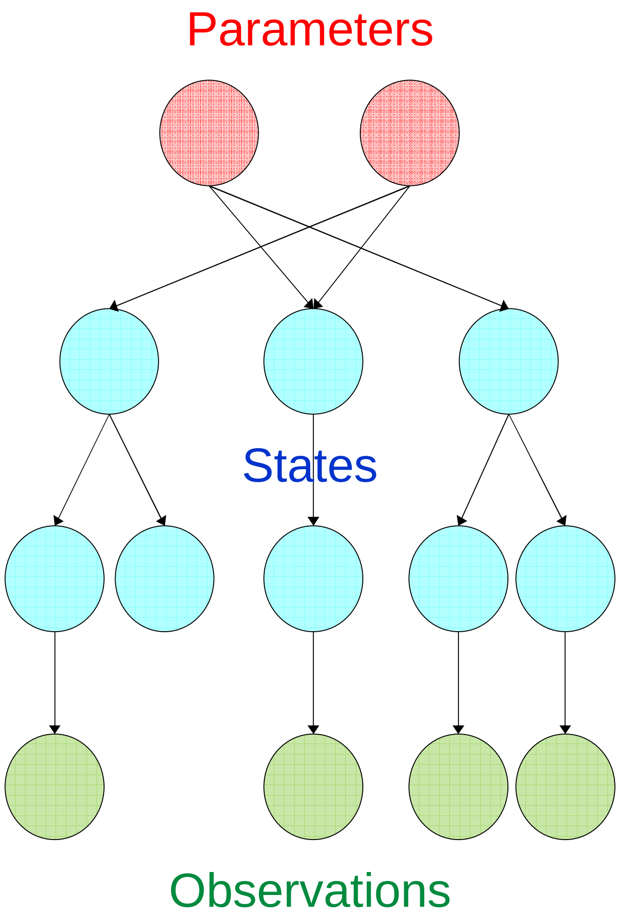
and the observations



Hierarchical Models

Clark, 2005 ; Buckland et al., 2007 ; Cressie et al., 2009 ; Kerry and Schaub, 2011 ; Parent and Rivot,2012

$$P(\theta, States, Obs.) = \underbrace{P(\theta)}_{Param.} \times \underbrace{P(States|\theta)}_{Demog.} \times \underbrace{P(Obs.|States, \theta)}_{Lik.}$$



- A synthesis between mechanistic and probabilistic approaches

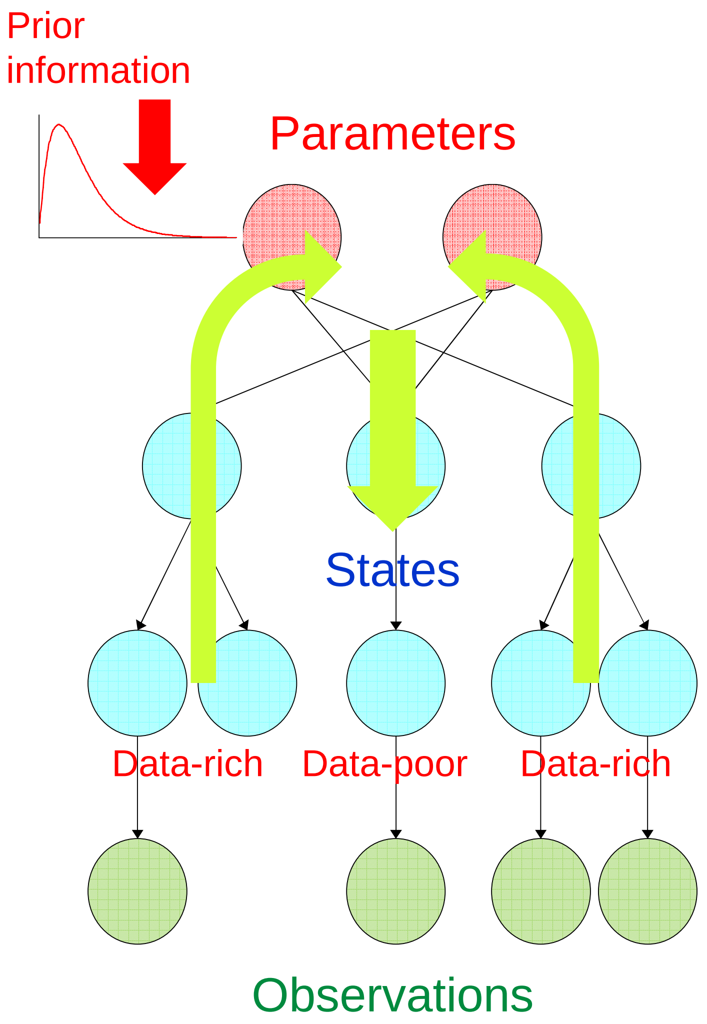
➔ Param → States → Obs. built in a mechanistic logic
→ Multiple dependencies and interactions

➔ States and observation processes are modelled in a probabilistic logic
→ Stochasticity at every level

➔ Assimilate multiple sources of observations (with errors) to extract an integrated signal

Hierarchical Bayesian Models

Clark, 2005 ; Buckland et al., 2007 ; Cressie et al., 2009 ; Kerry and Schaub, 2011 ; Parent and Rivot,2012



- Bayesian inferences offer a full probabilistic coherence
  $P(\theta, States, Obs.) \rightarrow P(\theta, States \mid Obs.)$

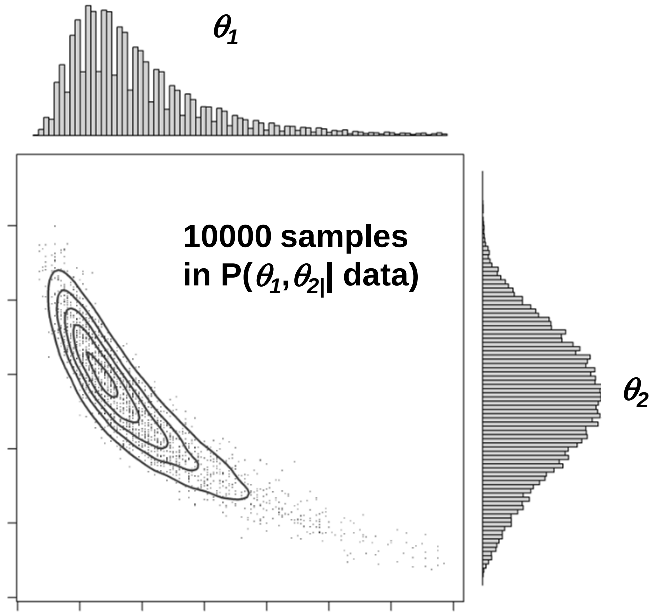
- Possibility to use informative prior
  « Standing on the shoulders of giants »
  (Hilborn and Lierman, 1998)

- Borrowing strength between similar units in the process
  « Robin Hood Approach » (Punt, 2011)

Bayesian Hierarchical Models in practice

Monte-Carlo methods

Use Monte Carlo sampling based methods
to approximate the joint posterior distribution



" The genie can not be put back into the bottle. The Bayesian machine, together with MCMC, is arguably the most powerful mechanism never created for processing data and knowledge."

Berger, J. O. (2000). Bayesian analysis: a look at today and thoughts of tomorrow. *Journal of the American Statistical Association*, *95*(452), 1269–1276.

Bayesian computation : a statistical revolution.

Brooks, S. P. (2003).
Phil. Trans. R. Soc. Lond., A, 361, 2681–2697.

Monte-Carlo methods

- **Friendly softwares exist**
  - **Write any model in a simple declarative language**
    - Prior
    - Likelihood (piece wise)
  - **Even complex non conjugate structures**

BUGS **" Bayesian inference Using Gibbs Sampler "**

- WinBUGS

- OpenBUGS

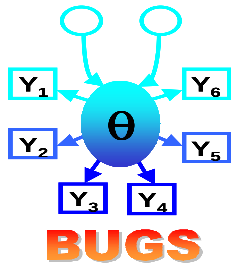
JAGS **" Just another Gibbs sampler "**

STAN **http://mc-stan.org/**

WinBUGS® / OpenBUGS

The most popular tools for Bayesian Hierarchical Modelling

WinBUGS

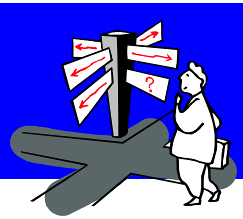


Cambridge MRC Biostatistics, Imperial College
The original project (first release in 1997)
Last version (1.4.3) released in 2007
Not further developed (but stable version still available)

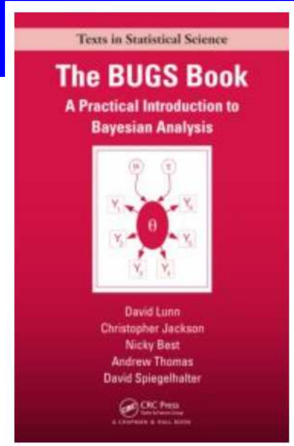
OpenBUGS

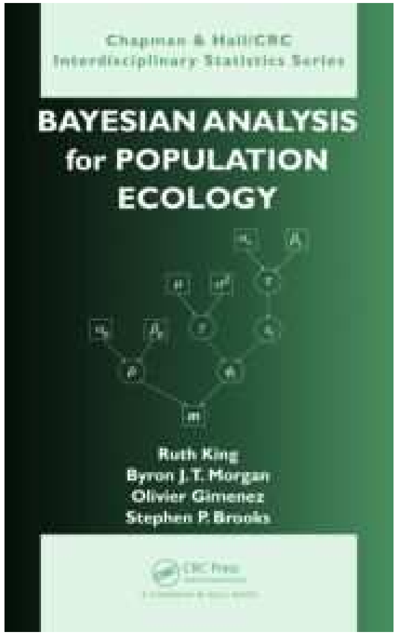
An open-source version of the software (first released 2007)
The future of the WinBUGS project
Last release OpenBUGS 3.2.3
Runs with MSWindows / Unix-Linux & Mac (using Wine)

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N. 2009. "The BUGS project: Evolution, critique and future directions". *Statistics in Medicine* **28** (25): 3049–3067.

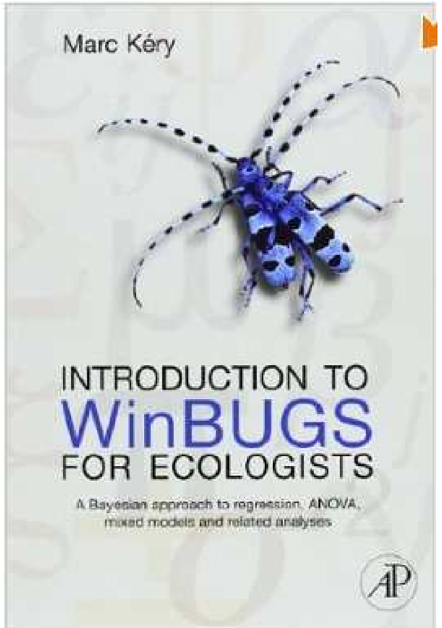


Some recommended reading

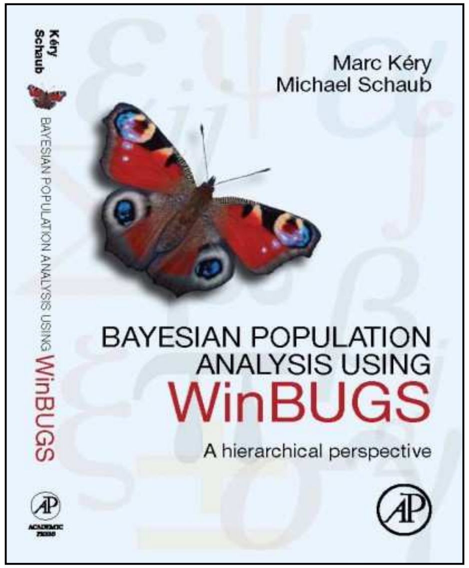




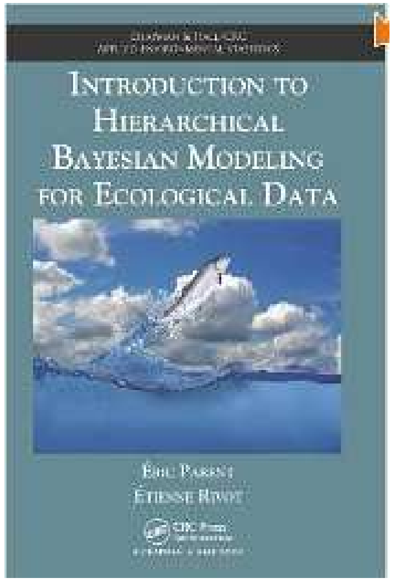
2009
Chapman & Hall / CRC



Academic Press
2010



Academic Press
2012



2012
Chapman & Hall / CRC

WinBUGS® / OpenBUGS

R Packages to run OpenBUGS within R

R2OpenBUGS

- The OpenBUGS equivalent of R2WinBUGS
- Automatically construct an OpenBUGS script to perform the analysis, launching OpenBUGS in the background to run the script, then exiting OpenBUGS

BRugs

- Allows OpenBUGS analyses to be run fully interactively from within R, without needing to launch the standalone OpenBUGS program
- Each OpenBUGS menu or script command (check model …) has its own R function
- Automatically searches standard locations for the program OpenBUGS (that must be installed separately)

Coda®, Boa®

MatBUGS

JAGS

JAGS – Just Another Gibbs Sampler

Martyn Plummer, Lyon

Initial release 2007

Use the same language that BUGS (JAGS slightly more flexible)

Written in C++ → available on any plateform
Instead, WinBUGS / OpenBUGS are written in Component Pascal
→ only available for Windows

No Windows interface for model building and MCMC sample post-processing

R interface - rjags()
- runjags()

STAN

STAN

Andrew Gelman, Columbia University, New-York

Initial release 2012

Written in C++
Models can be written in C++ or in BUGS language (and then compiled in C++)
R interface - RStan()

Motivation

Efficient sampling of complex hierarchical structure

Some large hierarchical models are intractable by Gibbs sampling (and then by BUGS/JAGS) because the variables are highly correlated

STAN uses (adaptive) Hamiltonian Monte Carlo Sampling
Requires the gradient of the log-posterior

AD - Model Builder

AD-Model Builder

First release 1993 (David FOURNIER) ; Open source since 2007

- Developed for powerful optimization
- Initially developed for Maximum Likelihood estimation
- But can also handle Bayesian models (MCMC module)

<u>Principle</u>

- Uses *Autodiff* to find point estimates of the parameters
- Gaussian estimate of the the posterior distribution (point estimates + variance-covariance derived from the Hessian)
- MCMC sampling (Metropolis-Hastings) with the Gaussian estimate as a proposal distribution

R2admb