

# 用R和WinBUGS实现贝叶斯分级模型

## Bayesian hierarchical modeling using R and WinBUGS

李欣海

(中国科学院动物研究所)

2013年5月19日

[lixh@ioz.ac.cn](mailto:lixh@ioz.ac.cn)

<http://people.gucas.ac.cn/~LiXinhai>

<http://blog.sciencenet.cn/u/lixinhai>

<http://weibo.com/lixinhaiblog>



# 贝叶斯方法 **Bayesian method**

(Reverend Thomas Bayes 1763)

Bayesian: Probability (parameter, given data)

Frequentist: Probability (data, given parameter)

## Bayes' Theorem

The diagram illustrates Bayes' Theorem with the following components and labels:

- Likelihood**: Points to the term  $f(y|\theta)$  in the numerator.
- Prior probability**: Points to the term  $P(\theta)$  in the numerator.
- Posterior probability**: Points to the term  $P(\theta|y)$  on the left side of the equation.
- Marginal probability of y**: Points to the term  $m(y)$  in the denominator.

$$P(\theta|y) = \frac{f(y|\theta) P(\theta)}{m(y)}$$
$$m(y) = \int f(y|\theta) p(\theta) d\theta$$

## **Bayesian method is appropriate for dynamics, nonlinear, and full-of-noise ecological processes**

Systems invariably driven by endogenous dynamic processes plus demographic and environmental process noise, and are only observable with error.

The inability to make well-founded statistical inferences about biological dynamic models in the chaotic and near-chaotic regimes, ..., leaves dynamic theory without the methods of quantitative validation that are essential tools in the rest of biological science.

Here I show that this impasse can be resolved in a simple and general manner, ..., using a straightforward Markov chain Monte Carlo (MCMC) sampler (Wood 2010).

Wood SN (2010). Statistical inference for noisy nonlinear ecological dynamic systems.  
**Nature** 466: 1102-1113.

## A simple example: mortality of moths exposed to cypermethrin

(Royle and Dorazio 2008 Page 66)

The experiment was designed to test whether males and females moths suffered the same mortality when exposed to identical doses of cypermethrin (氯氰菊酯).

Data observed in a dose-response experiment involving adults of the tobacco budworm (*Heliothis virescens*, 烟青虫), a moth species whose larvae are responsible for damage to cotton crops in the United States and Central and South America (Collett, 1991, Example 3.7).

In the experiment, batches of 20 moths of each sex were exposed to a pesticide called cypermethrin for a period of 72 hours, beginning two days after the adults had emerged from pupation. Both sexes were exposed to the same range of pesticide doses: 1, 2, 4, 8, 16, and 32 g cypermethrin.

At the end of the experiment the number of dead moths in each batch were recorded.

y	sex	sexcode	dose
1	male	1	1
4	male	1	2
9	male	1	4
13	male	1	8
18	male	1	16
20	male	1	32
0	female	0	1
2	female	0	2
6	female	0	4
10	female	0	8
12	female	0	16
16	female	0	32

# Models

Let  $x_i$  denote the  $\log_2(\text{dose})$  of cypermethrin administered to the  $i$ th batch of moths that contained either males ( $z_i = 1$ ) or females ( $z_i = 0$ ). Each batch has  $N = 20$  moths. A logistic-regression model containing 3 parameters is:

$$y_i | N, p_i \sim \text{Bin}(N, p_i)$$

$$\text{logit}(p_i) = \alpha + \beta x_i + \gamma z_i$$

y	sex	sexcode	dose
1	male	1	1
4	male	1	2
9	male	1	4
13	male	1	8
18	male	1	16
20	male	1	32
0	female	0	1
2	female	0	2
6	female	0	4
10	female	0	8
12	female	0	16
16	female	0	32

where  $\alpha$  is the intercept,  $\beta$  is the effect of cypermethrin and  $\gamma$  is the effect of sex.

```
# -----data-----
N = 20 # 每组烟青虫的数量, 雌雄各6组
y = c(1,4,9,13,18,20, 0,2,6,10,12,16) # 每组死亡的烟青虫数量
sex = c(rep('male',6), rep('female',6))
dose = rep(c(1,2,4,8,16,32), 2) # 农药剂量
ldose = log(dose)/log(2) # 对数转换
sexcode = rep(0,length(sex)) # 定义性别代码为0
i = sex=='male' # 区分雌雄 (FALSE-TRUE)
sexcode[i] = 1 # 定义雄性代码为1, 剩下的雌性依旧为0
as.data.frame(cbind(y, sex, sexcode, dose)) # 显示数据

# -----arguments for R2WinBUGS-----
data = list(n=length(y), N=N, y=y, x=ldose, z=sexcode)
params = list('alpha', 'beta', 'gamma', 'w')
inits = function() {
  list(alpha=rnorm(1), beta=rnorm(1),
        gamma=rnorm(1), w=rbinom(1,1,0.5))
}
```

```
# -----native WinBUGS code-----
modelFilename = 'd:/code/bugs/model.txt'
cat('
  model {
    alpha ~ dnorm(0, 0.01)
    beta ~ dnorm(0, 0.01)
    w ~ dbin(0.5,1) #确定雌雄
    gamma ~ dnorm(0, 0.01)
    for (i in 1:n) {
      y[i] ~ dbin(p[i], N)
      logit(p[i]) <- alpha + beta*x[i] + w*gamma*z[i]
    }
  }
', fill = TRUE, file = modelFilename)
```

## Codes

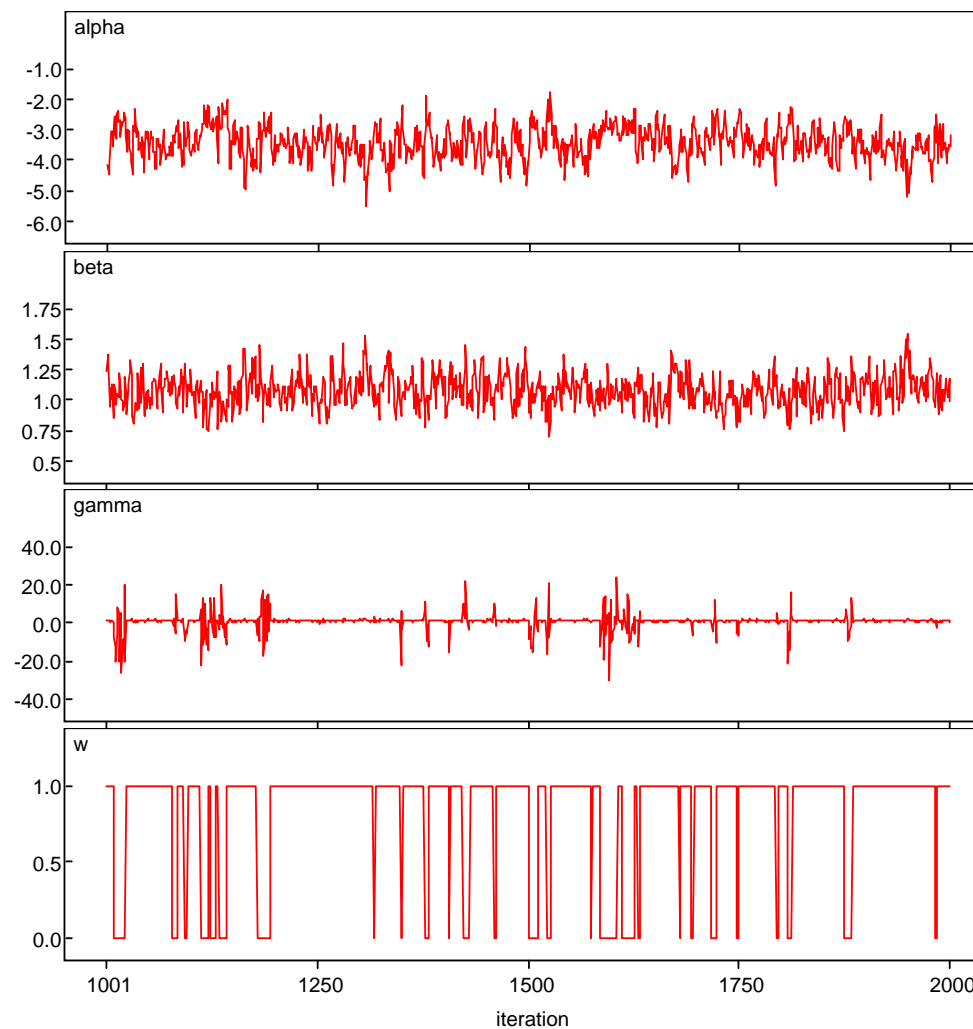
```
# -----call bugs() to fit model-----
library(R2WinBUGS)
modelFilename = 'd:/code/bugs/model.txt'
fit = bugs(data, inits, params, model.file = modelFilename,
  n.chains = 1, n.iter = 10000, n.burnin = 5000, n.thin = 5,
  bugs.seed = sample(1:9999, size=1), debug = TRUE,
  DIC = FALSE,
  bugs.directory = "d:/softwares/WinBUGS14/")
fit
fit$sims.matrix
```

### # Key WinBUGS code

```
alpha ~ dnorm(0, 0.01)
beta ~ dnorm(0, 0.01)
w ~ dbin(0.5, 1) #确定雌雄
gamma ~ dnorm(0, 0.01)
```

```
for (i in 1:n) {
  y[i] ~ dbin(p[i], N)
  logit(p[i]) <- alpha + beta*x[i] + w*gamma*z[i]
}
```

## Results



node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	-3.423	0.5463	0.03888	-4.494	-3.406	-2.397	1001	1000
beta	1.077	0.1354	0.007451	0.827	1.07	1.364	1001	1000
gamma	1.143	1.885	0.06088	0.3462	1.187	2.05	1001	1000
w	0.822	0.3825	0.0342	0	1	1	1001	1000

## 分级模型 Hierarchical modeling

### 一个案例(an example)

物种在地点  $i$  的数量

(Abundance of a species at site  $i$ ):

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

每个物种个体在地点  $i$  被发现的概率

(Detection rate of an individual at site  $i$ ):

$$r_i$$

需要估计的参数

(Parameters to be estimated):

$$\lambda_i \quad r_i$$

## 分级模型的历史 History of hierarchical modeling

The idea of hierarchical modeling started in the mid 20th century (Gelman et al. 1995). The hierarchical spatio-temporal dynamic model methodology was illustrated with a case study concerned with predicting the abundance of the house finch (*Carpodacus mexicanus*) over the eastern half of the U.S. from 1966 through 2001, with data collected during the North American Breeding Bird Survey (BBS; Robbins et al. 1986).

Hierarchical models in environmental sciences were developed in recent studies (Berliner 1996; Wikle et al. 1998, 2001).

Hierarchical modeling is based on the simple fact from probability that the joint distribution of a collection of random variables can be decomposed into a series of conditional models (Wikle 2003a, 2003b).

Berliner, L. M. 1996. Hierarchical Bayesian time series models.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 1995. Bayesian Data Analysis (1st ed.), London: Chapman and Hall.

Wikle, C. K. 2003a. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* **84**:1382-1394.

Wikle, C. K. 2003b. Hierarchical models in environmental science. *International Statistical Review* **71**:181-199.

Wikle, C. K., L. M. Berliner, and N. Cressie. 1998. Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics* **5**:117-154.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner. 2001. Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds.


*Journal of the American Statistical Association* **96**:382-397.

## 代表性著作 Key literatures

All Publications > Ecology > June 2003 > HIERARCHICAL BAYESIAN MODELS FOR PREDICTING THE SPREAD OF ECOLOGICAL P... [Advanced Search](#)

Volume 84, Issue 6 (June)

< Previous Next >



[Current Issue](#)  
[Available Issues](#)  
[Preprints](#)

Journal Information

ISSN: 0012-9658  
Frequency: Monthly

[Mission and Scope](#)  
[Types of contributions](#)  
[Editorial Board](#)  
[Staff](#)  
[Instructions for Authors](#)  
[Reviewer Guidelines](#)

< Previous Article Next Article >

[Add to Favorites](#) | [Email](#) | [Download to Citation Manager](#) | [Track Citations](#) | [Permissions](#)

[Full-text](#) [PDF](#)

Wikle, Christopher K. 2003. HIERARCHICAL BAYESIAN MODELS FOR PREDICTING THE SPREAD OF ECOLOGICAL PROCESSES. *Ecology* 84:1382–1394. [doi: 10.1890/0012-9658(2003)084[1382:HBMFPTJ]2.0.CO;2]

Special Feature

**HIERARCHICAL BAYESIAN MODELS FOR PREDICTING THE SPREAD OF ECOLOGICAL PROCESSES**

Christopher K. Wikle<sup>1</sup>

Department of Statistics, 222 Math Science Building, University of Missouri, Columbia, Missouri 65211 USA

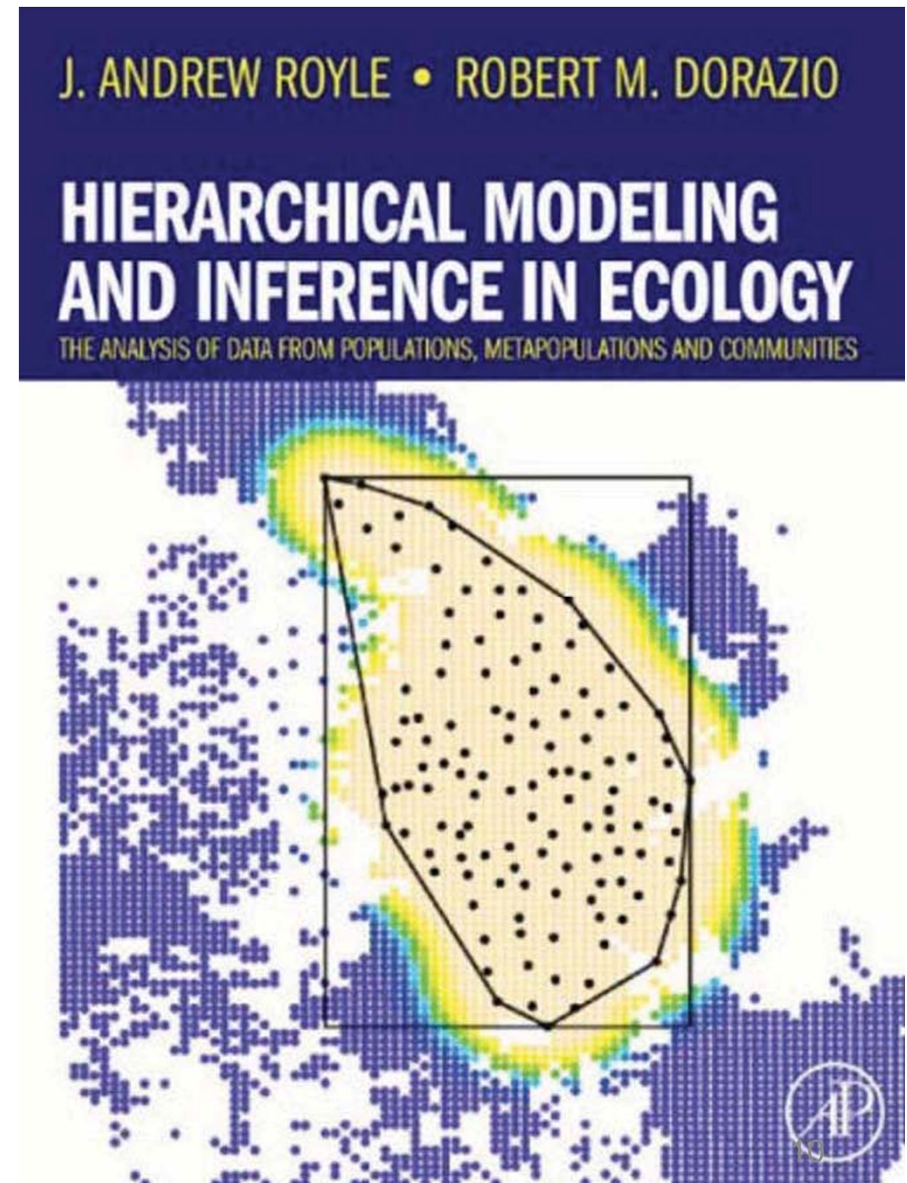
There is increasing interest in predicting ecological processes. Methods to accomplish such predictions must account for uncertainties in observation, sampling, models, and parameters. Statistical methods for spatiotemporal processes are powerful, yet difficult to implement in complicated high-dimensional settings. However, recent advances in hierarchical formulations for such processes can be utilized for ecological prediction. These formulations are able to account for the various sources of uncertainty and can incorporate scientific judgment in a probabilistically consistent manner. In particular, analytical diffusion models can serve as motivation for the hierarchical model for invasive species. We demonstrate by example that such a framework can be utilized to predict, spatially and temporally, the relative population abundance of House Finches over the eastern United States.

Corresponding Editor (ad hoc): J. S. Clark

Keywords: [Bayesian models](#), [Carpodacus mexicanus](#), [diffusion](#), [forecast](#), [hierarchical models](#), [House Finch](#), [invasive species](#), [state space](#), [uncertainty](#)

Received: December 3, 2001; Revised: May 7, 2002; Accepted: May 23, 2002; Final version received: June 28, 2002

<sup>1</sup> E-mail: [wikle@stat.missouri.edu](mailto:wikle@stat.missouri.edu)

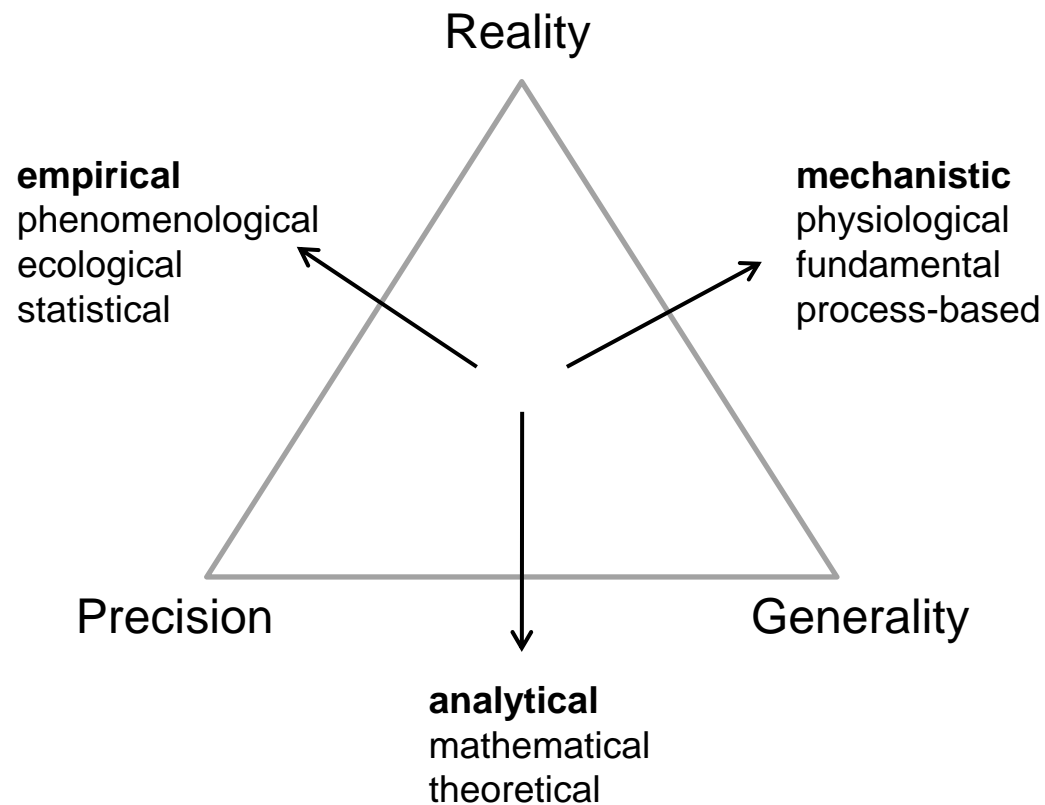


Wikle, C. K. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* **84**:1382-1394.

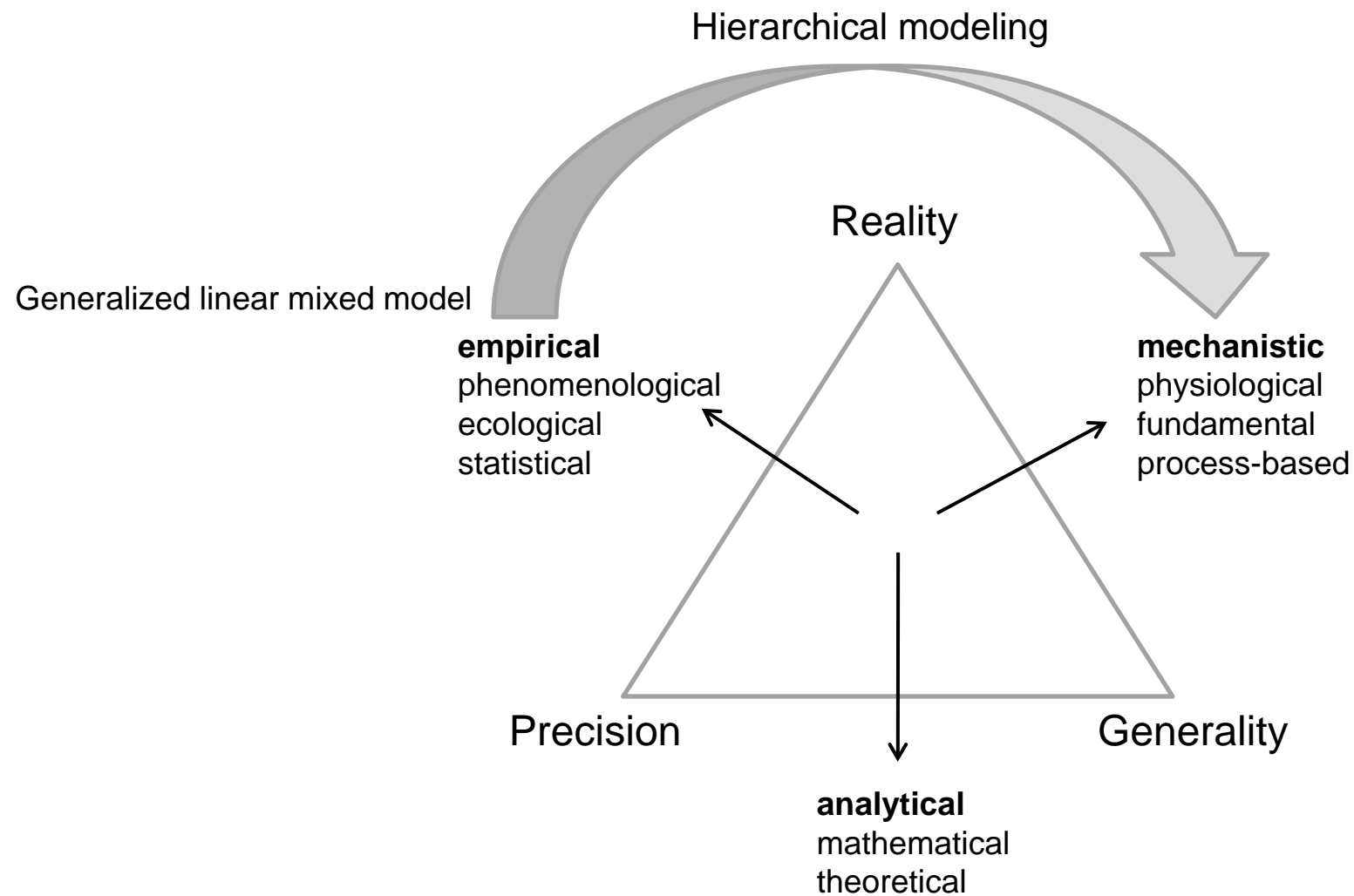
Royle, J. and R. Dorazio. 2008. Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities. Academic Press.

# A classification of models based on their intrinsic properties

After Levins (1966), Sharpe (1990), and Guisan & Zimmermann (2000)



## Improvement from hierarchical modeling



## 分级模型案例 An example of hierarchical modeling

(Royle and Dorazio 2008 Page 88)

Swiss bird survey data consisting of 3 replicate quadrat counts of the willow tit (*Parus montanus*, 褐头山雀) during the breeding season and covariates elevation (m) and forest cover (percent). Only a subset of quadrat counts are shown here. The symbol 'NA' indicates a missing value in an R data set.

repl	rep2	rep3	elevation	forest
0	0	0	910	21
0	0	0	540	11
0	1	NA	1880	32
1	0	1	1400	32
0	1	1	1210	75
1	0	0	380	23
0	0	NA	1840	0
1	0	NA	1910	18
1	2	2	1630	33
0	0	0	1540	21
2	4	3	1340	39
2	3	1	1410	52
14	13	9	2030	36
6	6	8	1880	66

## Models in WinBUGS format

```
for(i in 1:M){ # M sites
  z[i] ~ dbin(psi[i],1) # STATE MODEL, psi[i] is the probability of occurrence at site i
  logit(psi[i]) <- b0 + b1*elev[i] + b2*elev2[i] + b3*forest[i] # probability of occ.

  for(t in 1:J){ # J is the number of surveys at site i, which is 3
    mu[i, t] <- a0 + a1*date[i, t] + a2*date2[i, t] + a3*intensity[i, t]
    p[i, t] <- exp(mu[i, t])/(1+exp(mu[i, t])) # probability of detection
    mui[i, t] <- z[i]*p[i, t] # realized detection
    y[i, t] ~ dbin(mui[i, t],1) # OBSERVATION MODEL
  }
}
```

## R code

```

setwd('d:/models/hierarchical_modeling')
library("R2WinBUGS")
source("utilfns.Rd")

data      <- read.table("wtmatrix.csv", header=TRUE, sep="," , na.strings=c("NA")) # see previous slide
y         <- as.matrix(data[, c("y.1", "y.2", "y.3")])
M         <- nrow(y)
J         <- ncol(y)
length    <- as.vector(data[, "length"])
forest    <- as.vector(scale(data[, "forest"], center=TRUE))
elev      <- as.vector(scale(data[, "elev"], center=TRUE))
elev2     <- elev*elev
date       <- as.matrix(data[,c("day.1", "day.2", "day.3")])
mdate     <- mean(date, na.rm=TRUE)
sddate    <- sqrt(var(date[1:length(date)], na.rm=TRUE))
date       <- (date - mdate) / sddate
date2     <- date*date
dur        <- as.matrix(data[, c("dur.1", "dur.2", "dur.3")])
intensity <- dur/length
mint       <- mean(intensity, na.rm=TRUE)
sdint     <- sqrt(var(intensity[1:length(intensity)], na.rm=TRUE))
intensity <- (intensity - mint)/sdint
length     <- (length - mean(length))/sqrt(var(length))
date[is.na(y)] <- date2[is.na(y)] <- intensity[is.na(y)] <- 0

data      <- list ( "y", "M", "J", "forest", "elev", "elev2", "date", "date2", "intensity")
inits     <- function()
            list ( z=rbinom(M,1,.4),psi0=runif(1),b1=rnorm(1),b2=rnorm(1),
                  b3=rnorm(1),p0=runif(1),a1=rnorm(1),a2=rnorm(1),a3=rnorm(1))
parameters <- c("b0", "b1", "b2", "b3", "a0", "a1", "a2", "a3", "p0", "psi0")

```

### "utilfns.Rd"

```

`expit` <- function(x){
  exp(x)/(1+exp(x))
}

`logit` <- function(x){
  log(x/(1-x))
}

```

## R code

```

sink("model.txt")
cat("
model {
# prior distributions
p0      ~ dunif(0,1)
a0      <- log(p0/(1-p0))
a1      ~ dnorm(0,.001)
a2      ~ dnorm(0,.001)
a3      ~ dnorm(0,.001)
psi0    ~ dunif(0,1)
b0      <- log(psi0/(1-psi0))
b1      ~ dnorm(0,.001)
b2      ~ dnorm(0,.001)
b3      ~ dnorm(0,.001)
for(i in 1:M){
  z[i] ~ dbin(psi[i],1) # STATE MODEL
  logit(psi[i]) <- b0 + b1*elev[i] + b2*elev2[i] + b3*forest[i]
  for(t in 1:J){
    mu[i, t] <- a0 + a1*date[i, t] + a2*date2[i, t] + a3*intensity[i, t]
    p[i, t]   <- exp(mu[i, t])/(1+exp(mu[i, t]))
    muy[i, t] <- z[i]*p[i, t]
    y[i, t]   ~ dbin(muy[i, t],1) # OBSERVATION MODEL
  }
}
}
",fill=TRUE)
sink()

```

```

fit = bugs(data, inits, parameters, "model.txt",
          n.chains=3, n.iter=2000, n.burnin=1000, n.thin=1,
          bugs.seed=sample(1:9999, size=1),
          debug=TRUE, DIC=FALSE,
          bugs.directory = "d:/softwares/WinBUGS14/")
fit

```

## Results

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
a0	1.288	0.236	0.01057	0.814	1.293	1.752	1001	3000
a1	0.1816	0.1775	0.008539	-0.1515	0.1811	0.5331	1001	3000
a2	-0.04908	0.1576	0.008315	-0.3486	-0.04592	0.2685	1001	3000
a3	0.1571	0.2308	0.01171	-0.2552	0.1324	0.6192	1001	3000
b0	-0.1762	0.2751	0.01053	-0.7205	-0.1768	0.3636	1001	3000
b1	2.043	0.3083	0.01126	1.468	2.035	2.691	1001	3000
b2	-1.156	0.2654	0.01152	-1.716	-1.143	-0.6635	1001	3000
b3	0.8612	0.2362	0.007432	0.4179	0.8551	1.357	1001	3000
p0	0.7812	0.04028	0.001818	0.693	0.7846	0.8522	1001	3000
psi0	0.4568	0.06704	0.002553	0.3273	0.4559	0.5899	1001	3000



<http://www.flickr.com/photos/micr0tubulus/2358719527/>

## Fitting model to the willow tit data using the likelihood specification (non-Bayesian method)

```
# Data
source("D:/models/hierarchical_modeling/utlfn.Rd")
data      <- read.table("wtmatrix.csv", header=TRUE, sep=",", na.strings=c("NA"))
forest    <- as.vector(scale(data[, "forest"], center=TRUE))
elev      <- as.vector(scale(data[, "elev"], center=TRUE))
ymat      <- as.matrix(data[, c("y.1", "y.2", "y.3")])
M         <- nrow(ymat)
length    <- as.vector(data[, "length"])
date      <- as.matrix(data[, c("day.1", "day.2", "day.3")])
mdate     <- mean(date, na.rm=TRUE)
sddate    <- sqrt(var(date[1:length(date)], na.rm=TRUE))
date      <- (date - mdate) / sddate
dur       <- as.matrix(data[, c("dur.1", "dur.2", "dur.3")])
intensity <- dur/length
mint      <- mean(intensity, na.rm=TRUE)
sdint     <- sqrt(var(intensity[1:length(intensity)], na.rm=TRUE))
intensity <- (intensity - mint)/sdint
length    <- (length - mean(length))/sqrt(var(length))
```

(Royle and Dorazio 2008 Page 117)

```
# Function for parameter estimation
lik <- function(parms,vars){
  tmp <- rep(0,9)
  names(tmp) <- c("pconst","psiconst","length","elev1","elev2",
                 "forest","intensity","date1","date2")
  tmp[vars] <- parms
  ones <- rep(1,M) # number of row
  pmat <- expit (tmp[1]*ones + tmp[7]*intensity + tmp[8]*date
                + tmp[9]*(date^2)) # detection rate
  psi  <- expit (tmp[2]*ones + tmp[3]*length + tmp[4]*elev
                + tmp[5]*(elev^2) + tmp[6]*forest) # occurrences rate
  loglik <- rep(NA, M)
  for(i in 1:M){
    yvec    <- ymat[i,]
    navec   <- is.na(yvec)
    nd      <- sum(yvec[!navec])
    pvec    <- pmat[i,]
    cp      <- (pvec^yvec)*((1-pvec)^(1-yvec))
    cp[navec] <- 1
    loglik[i] <- log(prod(cp)*psi[i] + ifelse(nd==0,1,0)*(1-psi[i]))
  }
  sum(-1*loglik)
}
```

## Stepwise fitting for model selection

```
nam <- c("pconst", "psiconst", "elev1", "elev2", "forest", "length",
        "intensity", "date1", "date2", "aic")
out <- matrix(NA, nrow=10, ncol=length(nam))
dimnames(out) <- list(1:10, nam)
```

```
v <- c("pconst", "psiconst", "elev1", "forest")
x <- nlm(lik, c(0, 0, 0, 0), vars=v, hessian=TRUE)
out[1, v] <- x$estimate
out[1, "aic"] <- 2*x$minimum + 2*length(v)
```

```
v <- c("pconst", "psiconst", "elev1", "elev2", "forest")
x <- nlm(lik, c(0, 0, 0, 0, 0), vars=v, hessian=TRUE)
out[2, v] <- x$estimate
out[2, "aic"] <- 2*x$minimum + 2*length(v)
```

```
v <- c("pconst", "psiconst", "elev1", "elev2", "forest", "length")
x <- nlm(lik, c(0, 0, 0, 0, 0, 0), vars=v, hessian=TRUE)
out[3, v] <- x$estimate
out[3, "aic"] <- 2*x$minimum + 2*length(v)
```

```
v <- c("pconst", "psiconst", "elev1", "elev2", "forest", "length",
        "intensity")
x <- nlm(lik, c(0, 0, 0, 0, 0, 0, 0), vars=v, hessian=TRUE)
out[4, v] <- x$estimate
out[4, "aic"] <- 2*x$minimum + 2*length(v)
```

```
v <- c("pconst", "psiconst", "elev1", "elev2", "forest", "length",
        "intensity", "date1")
x <- nlm(lik, c(0, 0, 0, 0, 0, 0, 0, 0), vars=v, hessian=TRUE)
out[5, v] <- x$estimate
out[5, "aic"] <- 2*x$minimum + 2*length(v)
```

```
v <- c("pconst", "psiconst", "elev1", "elev2", "forest", "length",
        "intensity", "date1", "date2")
x <- nlm(lik, c(0, 0, 0, 0, 0, 0, 0, 0, 0), vars=v, hessian=TRUE)
out[6, v] <- x$estimate
out[6, "aic"] <- 2*x$minimum + 2*length(v)
```

```
v <- c("pconst", "psiconst", "intensity")
x <- nlm(lik, c(0, 0, 0), vars=v, hessian=TRUE)
out[7, v] <- x$estimate
out[7, "aic"] <- 2*x$minimum + 2*length(v)
```

```
v <- c("pconst", "psiconst", "intensity", "date1")
x <- nlm(lik, c(0, 0, 0, 0), vars=v, hessian=TRUE)
out[8, v] <- x$estimate
out[8, "aic"] <- 2*x$minimum + 2*length(v)
```

```
v <- c("pconst", "psiconst", "intensity", "date1", "date2")
x <- nlm(lik, c(0, 0, 0, 0, 0), vars=v, hessian=TRUE)
out[9, v] <- x$estimate
out[9, "aic"] <- 2*x$minimum + 2*length(v)
```

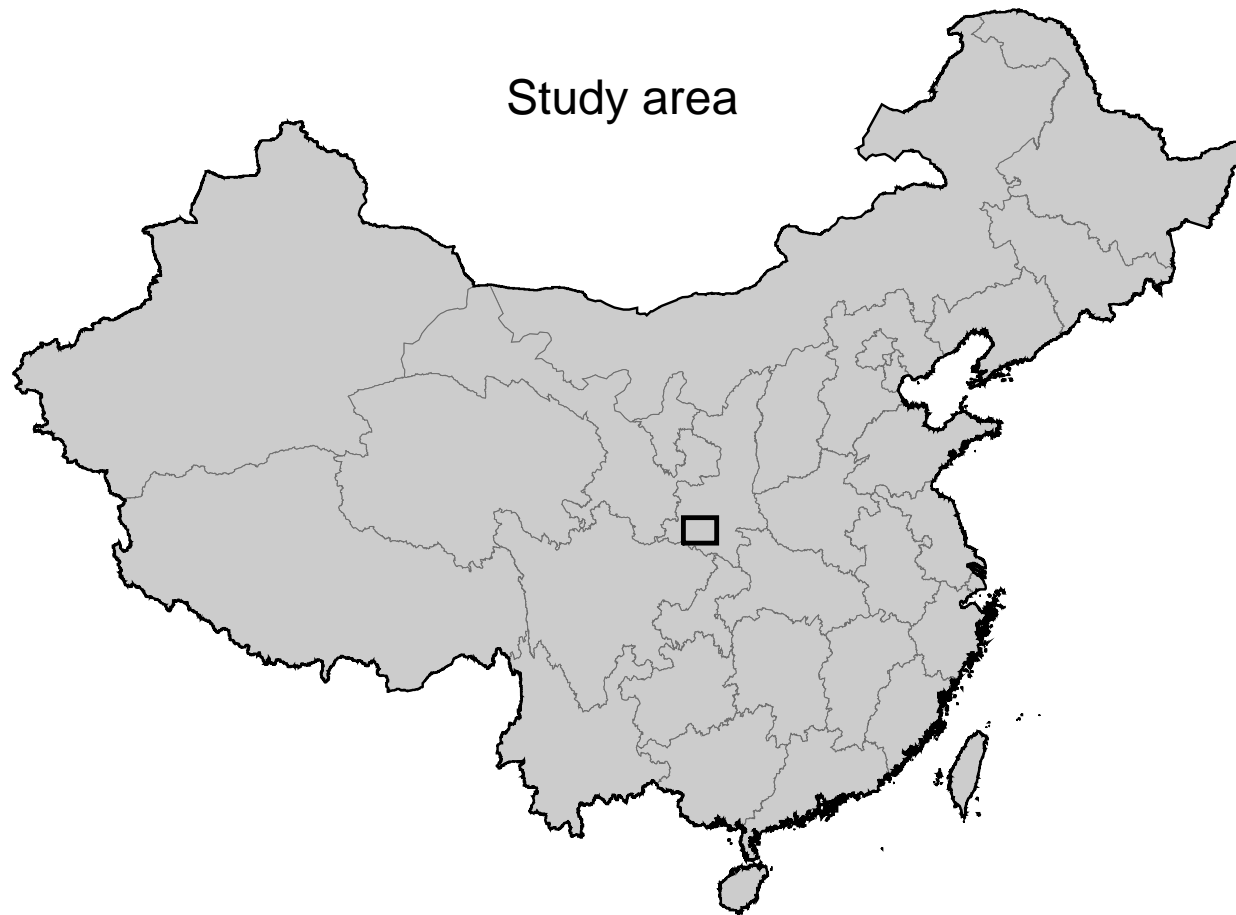
```
v <- c("pconst", "psiconst", "length", "intensity", "date1", "date2")
x <- nlm(lik, c(0, 0, 0, 0, 0, 0), vars=v, hessian=TRUE)
out[10, v] <- x$estimate
out[10, "aic"] <- 2*x$minimum + 2*length(v)
```

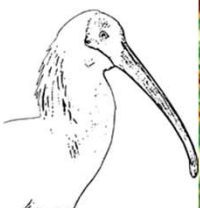
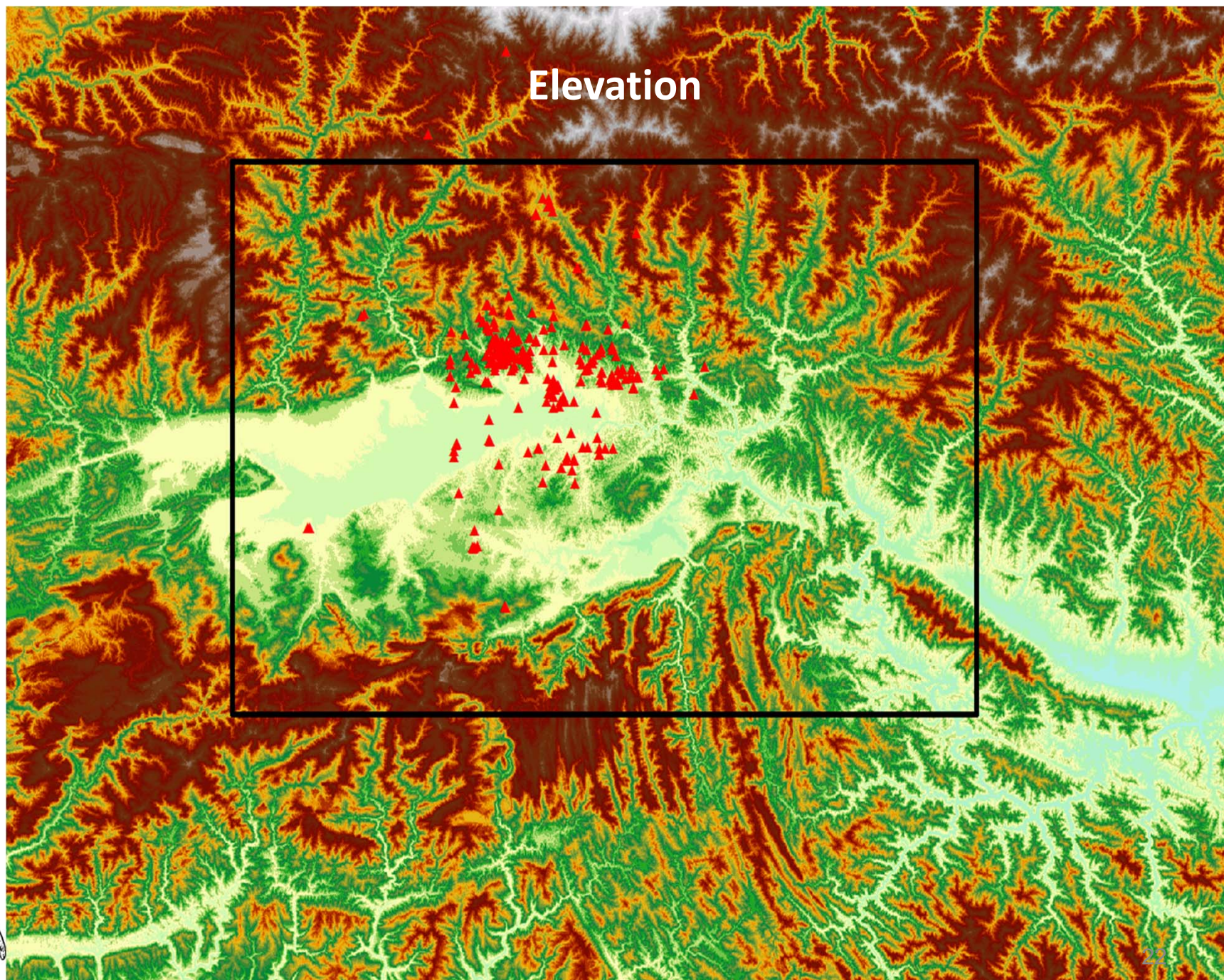
## Results of model selection

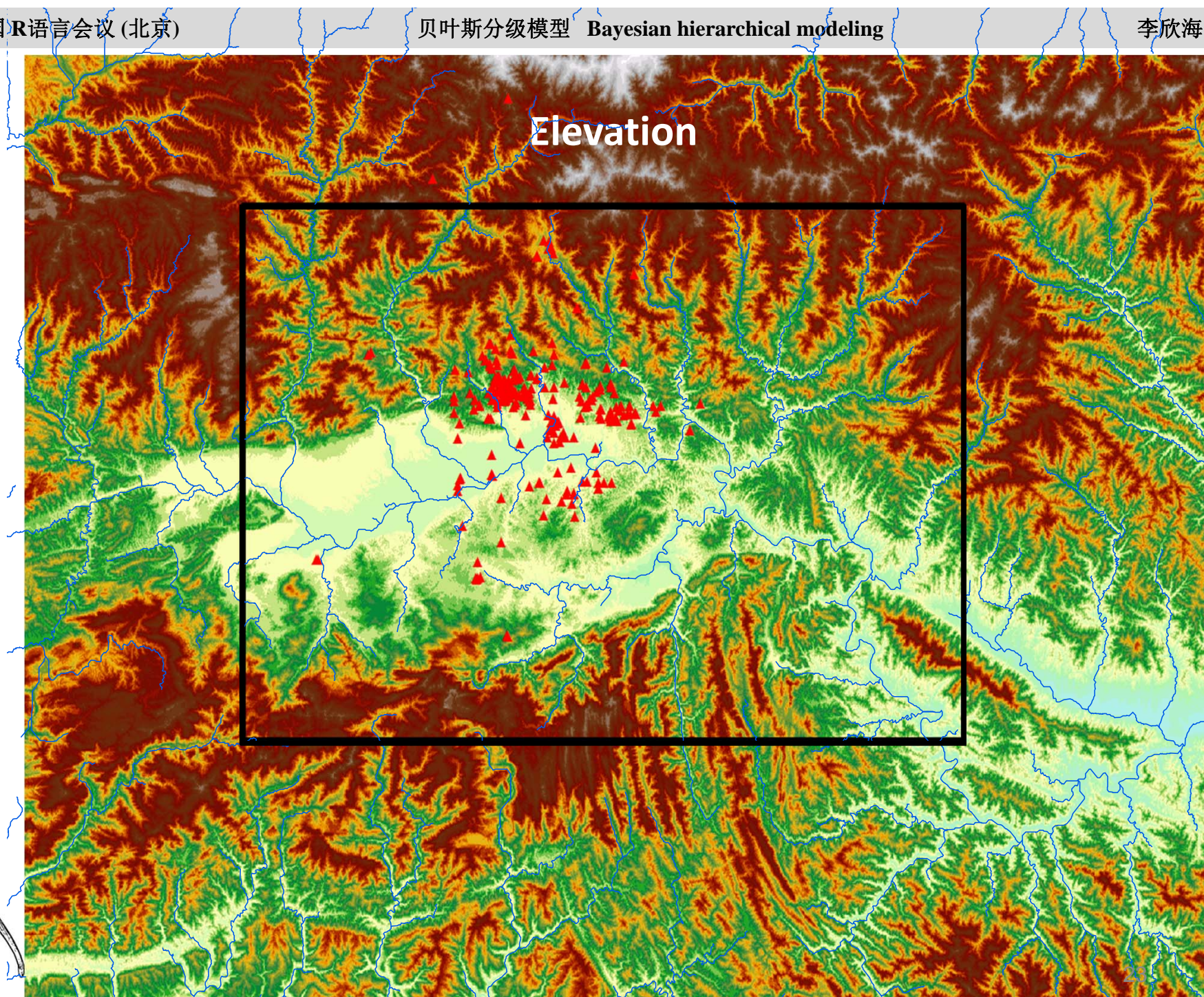
out[1:10, ]

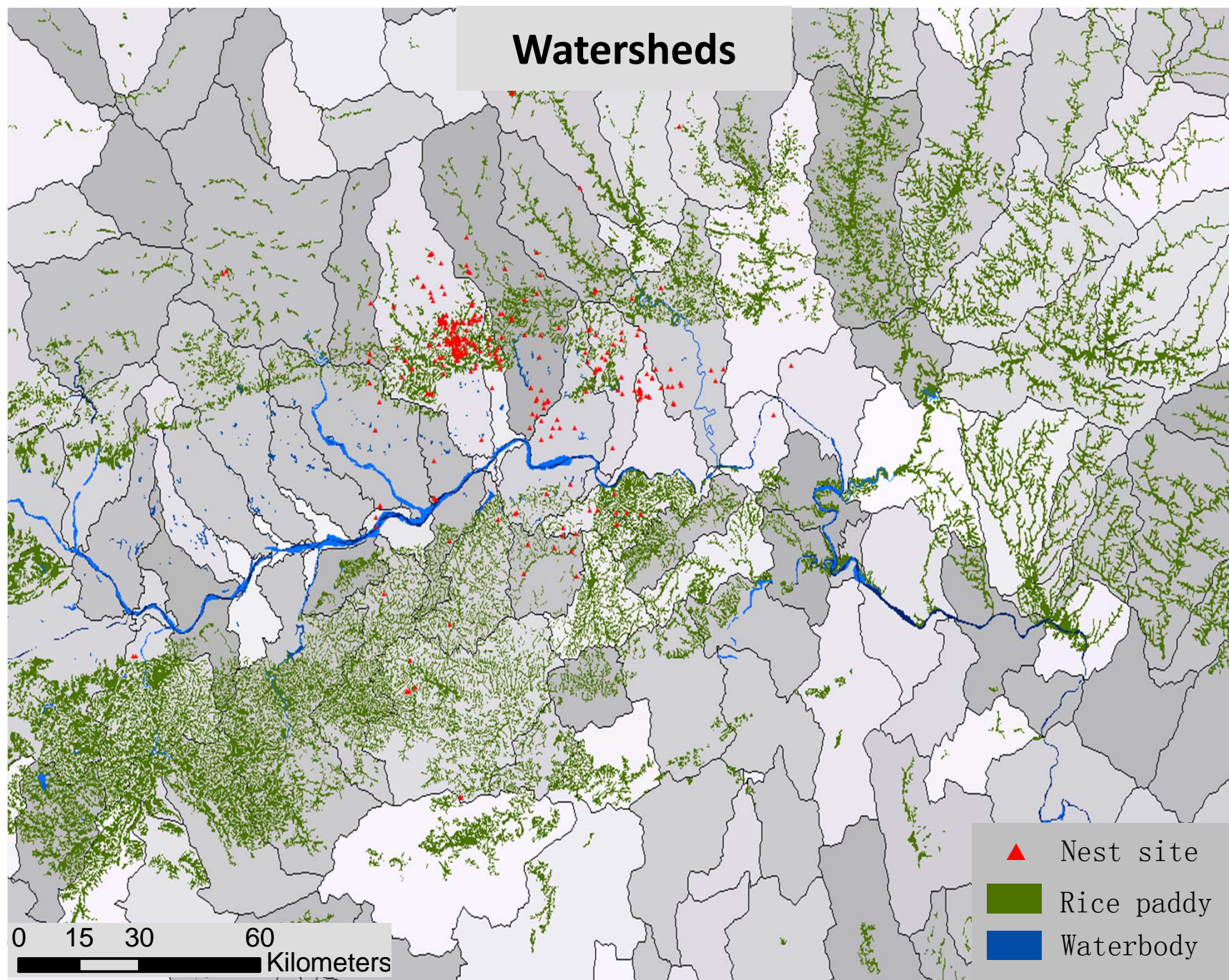
	pconst	psiconst	elev1	elev2	forest	length	intensity	date1	date2	aic
1	1.316	-1.050	1.348	NA	1.219	NA	NA	NA	NA	452.311
2	1.329	-0.207	1.985	-1.091	0.846	NA	NA	NA	NA	433.468
3	1.330	-0.171	2.032	-1.127	0.823	0.141	NA	NA	NA	434.945
4	1.252	-0.161	2.036	-1.130	0.823	0.153	0.196	NA	NA	436.086
5	1.241	-0.163	2.029	-1.131	0.825	0.150	0.146	0.153	NA	437.376
6	1.307	-0.163	2.028	-1.130	0.824	0.150	0.155	0.170	-0.074	439.145
7	1.223	-0.656	NA	NA	NA	NA	0.237	NA	NA	529.876
8	1.200	-0.656	NA	NA	NA	NA	0.190	0.177	NA	530.929
9	1.280	-0.654	NA	NA	NA	NA	0.206	0.195	-0.094	532.553
10	1.292	-0.661	NA	NA	NA	-0.164	0.188	0.192	-0.095	533.244

## Case study – nest site selection of crested ibis







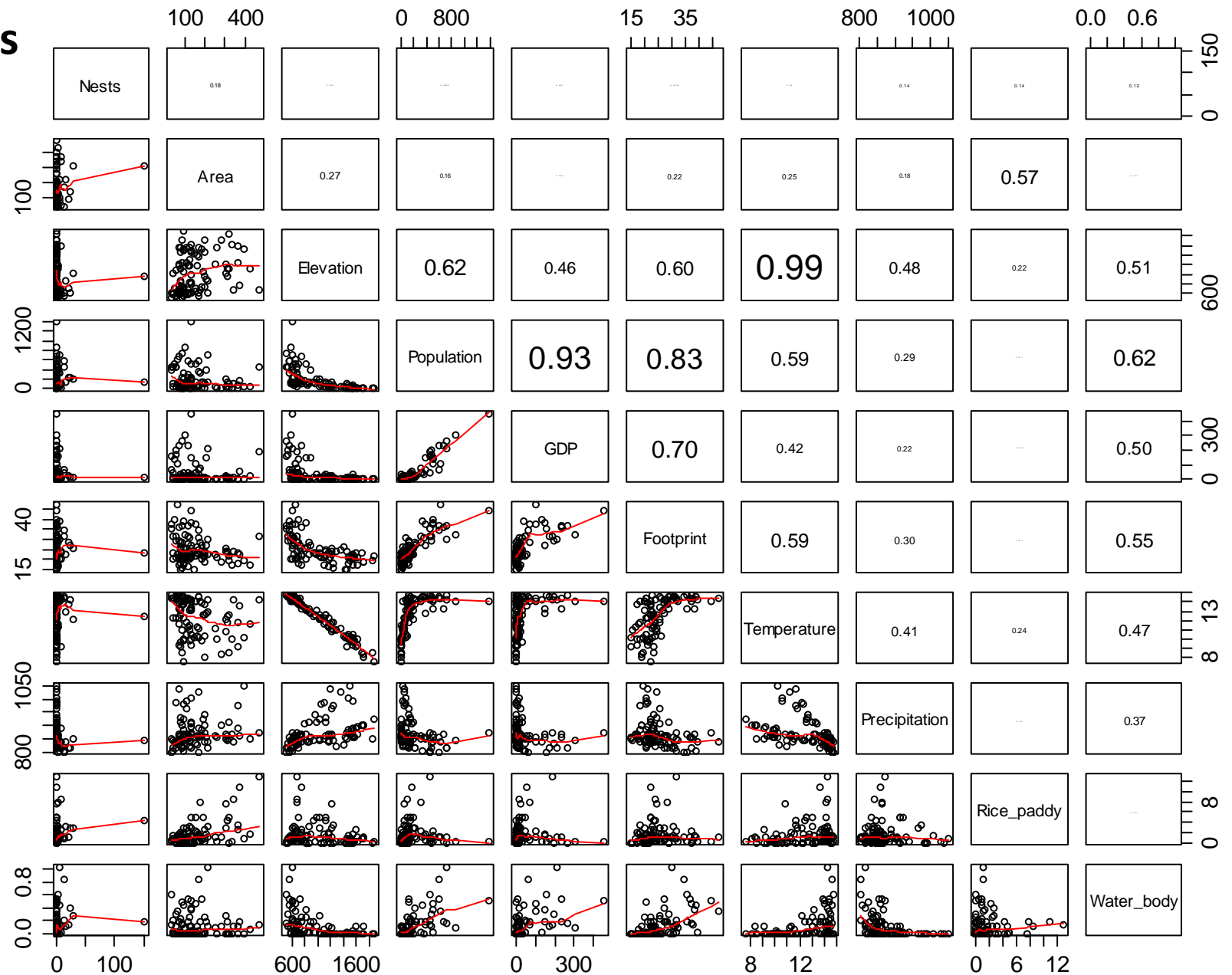


# Variables

## In every watersheds

- Average elevation
- SD of elevation
- Area of rice paddy
- Area of water body
- Human footprint index
- Population density
- GDP
- Temperature
- Precipitation
- Area of the watershed

# Correlation of variables



## Variable standardization

```
sheds <- read.csv('d:/text/ibis_Bayes/watersheds4.csv', header=TRUE) #read data

elev <- as.vector(scale(sheds[, "elev"], center=TRUE)) #elevation
elev.sd <- as.vector(scale(sheds[, "elev.SD"], center=TRUE)) #variance of elevation
foot <- as.vector(scale(sheds[, "footprint"], center=TRUE)) #Human Footprint Index
foot2 <- foot*foot #The square term
elev2 <- elev*elev #The square term
precip <- as.vector(scale(sheds[, "precip"], center=TRUE)) #Precipitation
T <- as.vector(scale(sheds[, "T"], center=TRUE)) #Temperature
pop <- as.vector(scale(sheds[, "pop"], center=TRUE)) #human population density
paddy <- as.vector(scale(sheds[, "paddy"], center=TRUE)) #area of rice paddy
water <- as.vector(scale(sheds[, "water"], center=TRUE)) #area of water body (pond and river)

wetland <- as.vector(scale(sheds[, "paddy"] * sheds[, "water"], center=TRUE)) #interaction term
```



Variable **wetland** is the product of **paddy** and **waterbody**.

## Variable selection using Generalized additive model (GAM)

```
library(mgcv)
model = gam(log(count + 1) ~ s(x1) + s(x2) + ... + s(x10), data = sheds)

Xs <- c("elev", "elev.SD", "footprint", "GDP", "population", "precip",
        "paddy", "waterbody", "wetland", "temperature")
```

From the full model, select models with 2-5 variables

Based on: deviance explained and GCV (Generalized Cross Validation)



## WinBUGS code

```
library(R2WinBUGS)
# Write WinBUGS code
sink("d:/text/ibis/ibis.txt")
cat("
model{
  for(i in 1:nsites){
    logit(p[i]) <- a0 + a1*foot[i] + a2*foot2[i] #detection rate
    logit(psi[i]) <- b0 + b1*wet[i] + b2*precip[i] #rate of occurrence
    z[i] ~ dbin(psi[i],1) #1=presence, 0=absence
    tmp[i] <- z[i]*p[i] #rate of recording
    y[i] ~ dbin(tmp[i],J[i]) #number of recording
  }
  a0 ~ dnorm(0,.001)
  a1 ~ dnorm(0,.001)
  a2 ~ dnorm(0,.001)
  b0 ~ dnorm(0,.001)
  b1 ~ dnorm(0,.001)
  b2 ~ dnorm(0,.001)
  logit(psi0) <- b0 #截距对应的 p值要指定,或者直接给初始值
  logit(p0) <- a0 #同上
}
",fill=TRUE)
sink()

data <- list ( "y","J","nsites","foot","foot2","wet","precip")
inits <- function()
  list(z=rbinom(nsites,1,.4),a0=rnorm(1),a1=rnorm(1),a2=rnorm(1),
       b0=rnorm(1),b1=rnorm(1),b2=rnorm(1))
parameters <- c("a0","a1", "a2","b0","b1","b2","psi0","p0")

out <- bugs(data, inits, parameters, "d:/text/ibis/ibis.txt",
            n.chain=3, n.burnin=1000, n.iter=10000, debug=T,
            bugs.directory = "d:/softwares/WinBUGS14/")
```

## WinBUGS code

朱鹮选择集水区*i* 作巢址的概率:

```
logit(psi[i]) <- b0 + b1*wetland[i] + b2*precipitation[i]
```

每个朱鹮巢址在集水区*i* 被发现的概率:

```
logit(p[i]) <- a0 + a1*footprint[i] + a2*footprint2[i]
```

朱鹮是否选择集水区*i* 作巢址:

```
z[i] ~ dbin(psi[i], 1)    #1=presence, 0=absence
```

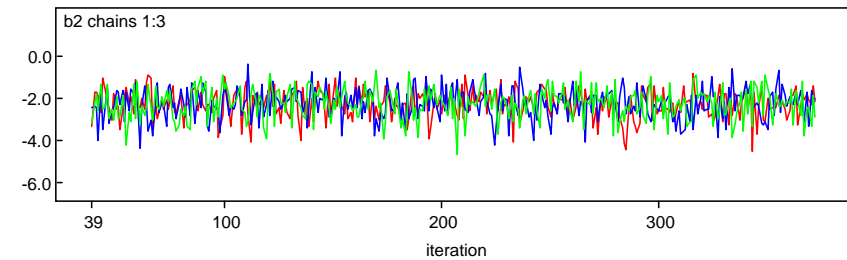
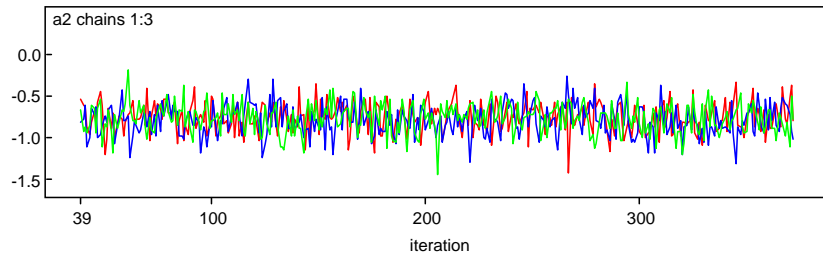
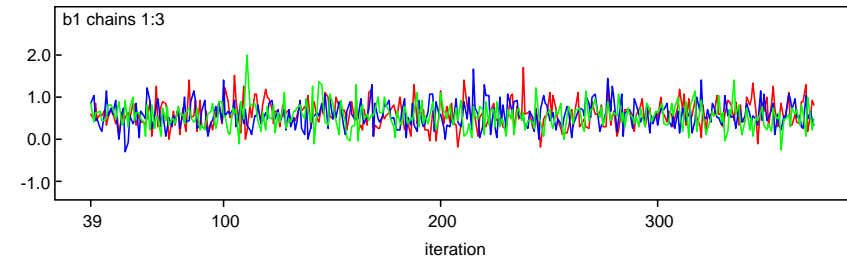
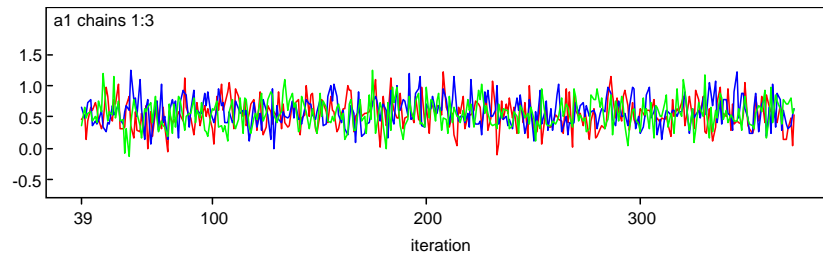
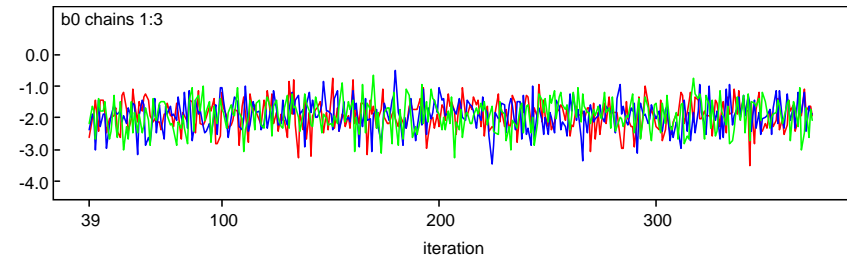
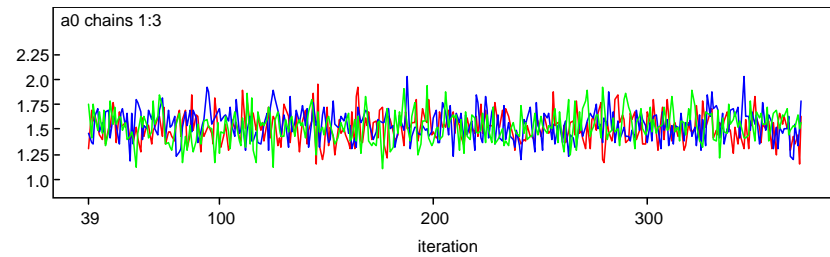
在集水区*i* 记录到朱鹮巢址的概率:

```
tmp[i] <- z[i]*p[i]
```

朱鹮巢址在集水区 *i* 的记录数(*J*[*i*]为在集水区 *i* 的调查次数):

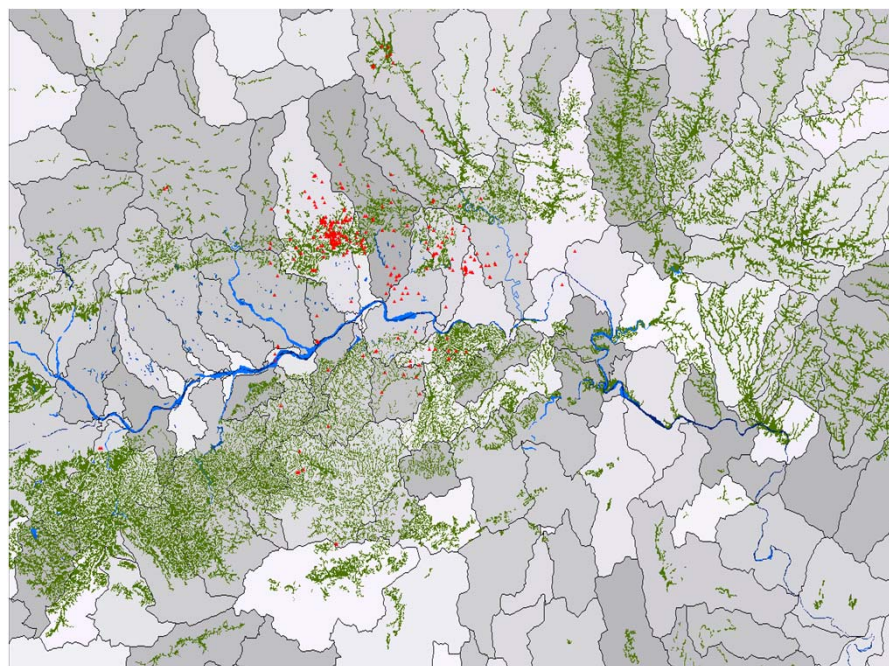
```
y[i] ~ dbin(tmp[i], J[i])
```

## Parameter estimation using MCMC



## Parameter estimation

node	mean	sd	MC error	2.50%	median	97.50%
a0	1.531	0.1484	0.005556	1.256	1.533	1.836
a1_footprint	0.5567	0.2303	0.00751	0.129	0.5492	1.029
a2_footprint <sup>2</sup>	-0.7765	0.18	0.006249	-1.129	-0.7768	-0.4362
b0	-1.924	0.4737	0.01423	-2.897	-1.893	-1.062
b1_wetland	0.5695	0.295	0.008542	0.02534	0.5557	1.207
b2_precipitation	-2.272	0.701	0.02063	-3.78	-2.222	-1.033



谢谢!

朱鹮育雏  
摄于2013年5月14日  
陕西洋县桂峰村