Prediction of count data with spatial dependency and zero-inflation

A hierarchical bayesian approach

O. Flores & F. Mortier

Cirad

January 31, 2007

Contents

Context

When **count data** are sampled in the field
(number of trees, flowers, seeds, tornadoes, accidents,. . .),

1. *spatial autocorrelation* (biology is contagious. . . !),
2. *zero-inflation* (low abondance, clumped pattern, sampling design)

. . . are likely ! !

$\bigoplus$ *multiple* descriptors of the environment

Modelling issues

1. how to model taking those features into account ?
2. how to select relevant explicative variables and fit the models ?

Classical models for count data

- **Poisson model**

Example :
beans dropped over a chess game and co
within the cells $\rightarrow Z \sim \mathcal{P}(\lambda)$

$$\mathbb{P}(Z = z|\lambda) = \frac{\lambda^{\mathbf{z}}}{\mathbf{z}!} e^{-\lambda}$$

$$\mathbb{E}(Z) = \lambda \quad \text{and} \quad \mathbb{V}(Z) = \lambda$$

- **Negative Binomial Model**
  Continuous mixture of Poisson distributions with Gamma-distributed intensity $\rightarrow Z \sim \mathcal{NB}(\lambda, \tau)$

$$\mathbb{P}(Z = \mathbf{z}|\lambda, \tau) = \frac{\Gamma(\mathbf{z} + \tau)}{\mathbf{z}!\Gamma(\tau)} \left(\frac{\tau}{\lambda + \tau}\right)^{\tau} \left(\frac{\lambda}{\lambda + \tau}\right)^{\mathbf{z}}, \quad (\lambda, \tau) > 0$$

$$\mathbb{E}(Z) = \lambda \quad \text{and} \quad \mathbb{V}(Z) = \lambda + \frac{\lambda}{}$$

Models for count data with zero-inflation I

Zero Inflated Poisson (ZIP) models

Two processes acting simultaneously :
- Is the distribution a $\mathcal{P}$oisson or certainly nul ?
- If Poisson, how many ?

ZIP as a Mixture Poisson model :

$$Z \sim \omega\delta(0) + (1-\omega)\mathcal{P}(\lambda)$$

$$\mathbb{P}(Z = z|\omega,\theta) = \begin{cases} \omega + (1-\omega)\mathbb{P}(Z = 0|\theta), & \text{if } z = 0 \\ (1-\omega)\mathbb{P}(Z \neq 0|\theta), & \text{if } z > 0 \end{cases}$$

$$\mathbb{E}(Z) = (1-\omega)\lambda \quad \text{and} \quad \mathbb{V}(Z) = \left(1 + \frac{\lambda}{\omega}\right)\lambda$$

Models for count data with zero-inflation II

ZI models as missing data models

Let $C = (C_1, \ldots, C_n)$ be a latent random variable so that C_i equals
- $c_i = 1$ if $Z_i = 0$ and drawn from (0)
- $c_i = 0$ if $Z_i > 0$ or if Z_i is null and drawn from $\mathcal{P}(\lambda)$

Marginal distribution : $C \sim Bernoulli(\omega)$

The new joint distribution is

$$
\begin{aligned}
f(Z, C|\omega, \lambda) &= \prod_{i=1}^{n} f(z_i|C_i = c_i, \omega, \lambda)\pi(C_i|\omega) \\
&= \prod_{i=1}^{n} p^{c_i} \left[(1-\omega)\,\mathbb{P}(Z_i = z_i|\lambda)\right]^{1-c_i}
\end{aligned}
$$

Taking explicative variables into account

Mixture proportion (ω) and Poisson intensity (λ) dependent on co-variables ($\mathbf{B}, \mathbf{X}$) :

- The mixture proportion is expressed as a function of **B** :

$$\text{logit}(\omega_i) = \mathbf{B}_i \beta$$

- The Poisson intensity depends on the environment via **X** :

$$\log(\lambda_i) = \mathbf{X}_i \gamma + \alpha_i$$

- α : spatial random effect allowing for autocorrelation between observations,

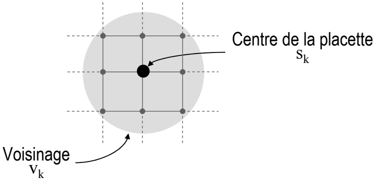
- **B** and **X** may have columns in common or not

Random spatial effect

Conditional auto-regressive process (CAR) on discret domaine (lattice)

$$\alpha_i | \alpha_j, j \in V_i \sim \mathcal{N}\left(\sum_{j \in V_i} \rho M_{ij} \alpha_j, \sigma^2\right)$$

- V_i neighborhood of individual i
- $E(\alpha) = 0$
- σ^2 : conditional variance
- ρ : spatial correlation
- $M = (M_{ij})$: known weights



$\theta = (\rho, \sigma^2)$
Hyper-prior : $\rho \sim U]a, b[$, $\sigma^2 \sim IG$

Variable selection

Let a unknown latent binary variable (to be estimated) indicate which explicative variables are included in the model :

$$\eta = \{\eta_j\}_1^p$$

where p is the total number of explicative variables.
The linear predictors are modified

$$\xi_i = \sum_{j=1}^{p} \mathbf{Y}_{ij}\delta_j\eta_j, \;\; i = 1, \ldots, n,$$

with $\xi = (\text{logit}(\omega), \log(\lambda)), \;\; \mathbf{Y} = (\mathbf{B}, \mathbf{X}), \delta = (\beta, \gamma)$

Hierarchical Bayesian models I

Three basic levels of hypotheses

1. *Data level* : conditional distribution of data

$$Z_i|\theta_1, \xi \sim \mathcal{F}(\theta_1, \xi_i)$$

$$\text{and } (Z_i|\theta_1, \xi_i) \perp (Z_j|\theta_1, \xi_j)$$

2. *Process Level* : distributions of parameters controling data level

$$\xi|\theta_2 \sim \Upsilon(\theta_2)$$

3. *Parameter level* : prior distributions of unknown parameters

$$\Theta = (\theta_1, \theta_2) \sim \Phi(\theta_3)$$

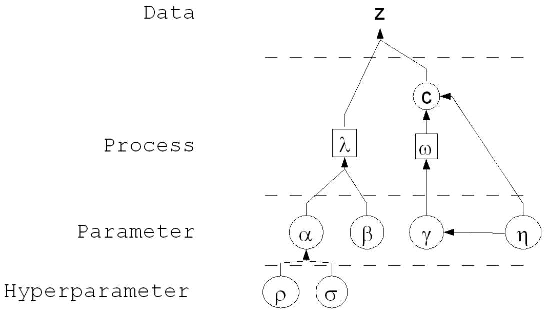
with θ_3 set *a priori*

Hierarchical Bayesian models II

x

Cyclic graph for spatial ZIP with variable selection : stochastic nodes (circles) or deterministic (squares)

Hierarchical Bayesian models III



Estimation : Bayesian principle

Aim : estimate (posterior) distribution of Θ given data z

- Given prior distribution on Θ : π_0,
- Posterior distribution (Bayes' theorem) :

$$\pi(\Theta|z) = \frac{f(z|\Theta)\pi_0(\Theta)}{\int f(z|\Theta)\pi_0(\Theta)d\Theta}$$

- In general, we do not know how to calculate $\pi(\Theta|z)$

Method : Approximate $\pi(\Theta|z)$ using a Monte Carlo Markov Chain algorithm

The ZIP case
Simulate the posterior distribution

In the spatial ZIP case with variable selection :

$$\Theta = (\eta, \beta, \gamma, \mathbf{c}, \alpha, \rho, \sigma)$$

The posterior distribution is :

$$\begin{aligned}\pi(\eta, \mathbf{c}, \gamma, \beta, \alpha, \rho, \sigma|\mathbf{z}) &= f(\mathbf{z}|\eta, \beta, \gamma, \mathbf{c}, \alpha)\pi(\mathbf{c}|\gamma)\pi(\alpha|\rho, \sigma^2) \\ &\quad \pi(\beta|\eta)\pi(\gamma)\pi(\rho)\pi(\sigma^2)\pi(\eta),\end{aligned}$$

where $f(\mathbf{z}|\eta, \beta, \gamma, \mathbf{c}, \alpha) = \ell(\eta, \beta, \gamma, \mathbf{c}, \alpha|\mathbf{z})$ is the likelihood of the parameter set given data.

Monte Carlo Markov Chain Algorithm

Aim : sample values of $\Theta = (\Theta_1, \ldots, \Theta_N)$ from an unknown distribution π

- Construct a markov chain whose asymptotic distribution is π
- When distribution π is obtained (convergence), extract samples $\Theta^{(k)} = (\Theta_1^{(k)}, \ldots, \Theta_N^{(k)})$ to estimate posterior mode, median, mean...

MCMC algorithm principle

One of *mutation/selection* algorithms in two steps :

1. Propose a new value for parameters (mutation) : $\Theta \longrightarrow \Theta^*$
2. Accept or reject mutation (selection)

Different types of algorithm :

- Mutation rule ? $\rightsquigarrow$ flexible : *independent, random walk, gradient-orientated*. . .
- Selection rule ? $\rightsquigarrow$ imposed by theory (Metropolis-Hastings, 1970)

Metropolis-Hasting algorithm

Require: Θ^0, initial point

for $i = 0$ to N_{iter} **do**

Let $\Theta^\star \sim Q(\Theta|\Theta^i)$, with Q the proposal distribution (mutation)

Accept

$$\Theta^{i+1} = \begin{cases} \Theta^\star & \text{with probability } r(\Theta^i, \Theta^\star) \\ \Theta^i & \text{with probability } 1 - r(\Theta^i, \Theta^\star) \end{cases}$$

where

$$r(\Theta^i, \Theta^\star) = min(r^\star, 1) = \min\left\{ \frac{\pi(\Theta^\star)}{\pi(\Theta^i)} \frac{Q(\Theta^i|\Theta^\star)}{Q(\Theta^\star|\Theta^i)}, 1 \right\}$$

end for

Gibbs sampling algorithm

Principle : parameters sequentially updated knowing the full conditional distributions $\pi_i(\Theta_i|\Theta_{-i})$

$\Theta = \Theta_1, \ldots, \Theta_n$ with known conditional distributions $\pi_1, \ldots, \pi_n$.

In the mutation step, one can simulate

1. $\Theta_1^{i+1} \sim \pi_1(\Theta_1^i|\Theta_2^i, \ldots, \Theta_n^i)$
2. $\Theta_2^{i+1} \sim \pi_2(\Theta_2^i|\Theta_1^{i+1}, \Theta_3^i, \ldots, \Theta_n^i)$
3. $\ldots$
4. $\Theta_n^{i+1} \sim \pi_n(\Theta_n^i|\Theta_1^{i+1}, \ldots, \Theta_{n-1}^{i+1})$

In this case, one can verify $r^\star = 1$ $\Rightarrow$ proposals are optimal (following MH
$\Rightarrow$ all proposals are accepted

Metropolis within Gibbs sampling

Some of the full conditional conditions may be unknown.
In this case, implement a Metropolis step for the corresponding parameters.
Overview of the overall algorithm :

1. Initialization

$$\Theta_0 = (\eta_0, \beta_0, \gamma_0, \mathbf{c}_0, \alpha_0, \rho_0, \sigma_0)$$

2. Sequential updates :
   - $\eta_{t+1} \mid \mathbf{z}, \beta_t, \gamma_t, \mathbf{c}_t, \alpha_t$ the latent indicator variable : $\eta_t \rightsquigarrow \eta_{t+1}$,
   - $(\beta_{t+1}, \gamma_{t+1}) \mid \mathbf{z}, \eta_{t+1}, \mathbf{c}_t, \alpha_t$ the regression coefficients : $(\beta_t, \gamma_t) \rightsquigarrow (\beta_{t+1}, \gamma_{t+1})$
   - $\mathbf{c}_{t+1} \mid \mathbf{z}, \eta_{t+1}, \beta_{t+1}, \gamma_{t+1}, \alpha_t$ the latent class variable : $\mathbf{c}_t \rightsquigarrow \mathbf{c}_{t+1}$
   - $\alpha_{t+1} \mid \mathbf{z}, \eta_{t+1}, \beta_{t+1}, \gamma_{t+1}, \rho_t, \sigma_t$ the spatial random effect : $\alpha_t \rightsquigarrow \alpha_{t+1}$
   - $\rho_{t+1} \mid \alpha_{t+1}, \sigma_t$ the spatial parameter mesuring dependency : $\rho_t \rightsquigarrow \rho_{t+1}$
   - $\sigma_{t+1} \mid \alpha_{t+1}, \rho_{t+1}$ the conditional variance parameter : $\sigma_t \rightsquigarrow \sigma_{t+1}$

Subalgorithms I
Examples

Independent Metropolis step :
η update for variable selection

- Prior $\eta_i \sim \mathcal{B}(0.5)$
- Proposal
  - randomly chosen $i \in \{1, \ldots, n_{var}\}$;
  - $\eta_i^\star \sim \mathcal{B}(0.5)$ ($\eta^\star = 1$ or 0)
- Selection

$$r^\star = \frac{\ell(z|\alpha, \beta, \eta^\star, \gamma)}{\ell(z|\alpha, \beta, \gamma)}$$

is the likelihood ratio

Subalgorithms II
Examples

Random Walk Metropolis step : ρ update

- Prior

$$\pi_0(\rho) \sim \mathcal{N}(0,1)\mathbb{1}_{[a,b]}$$

- Proposal

$$\rho^\star|\rho \sim \mathcal{N}(\rho, \sigma_\rho^2)\mathbb{1}_{[a,b]}$$

- Selection

$$\begin{aligned}\log(r^*) &= \frac{\ell(\rho^\star|\alpha,\sigma^2)}{\ell(\rho|\alpha,\sigma^2)}\frac{\mathcal{N}(\rho^\star,\sigma_\rho^2)}{\mathcal{N}(\rho,\sigma_\rho^2)} \\ &= \frac{\ell(\alpha|\rho^\star,\sigma^2)\pi_0(\rho^\star)}{\ell(\alpha|\rho,\sigma^2)\pi_0(\rho)}\frac{\mathcal{N}(\rho^\star,\sigma_\rho^2)}{\mathcal{N}(\rho,\sigma_\rho^2)}\end{aligned}$$

numerically tractable thanks to CAR properties

Subalgorithms III
Examples

Langevin-Metropolis step (gradient-orientated) : α update

- Prior : CAR model
- Proposal $\alpha^*|\alpha \sim \mathcal{N}(\mu_\alpha, h\mathbf{I})$, $\mu_\alpha = \alpha + \frac{h}{2}\nabla(\alpha)$

$$\nabla(\alpha) = (1 - \mathbf{c})(\mathbf{z} - \lambda) - {}^{\circ}\,\alpha$$

- Selection

$$\log(r^*) \quad = \quad \log\left[\frac{\ell(\alpha^*|\mathbf{z})}{\ell(\alpha|\mathbf{z})}\right] \frac{\pi(\alpha^*|\rho,\sigma)}{\pi(\alpha^*|\rho,\sigma)} \frac{\mathcal{N}(\mu_\alpha, h\mathbf{I})}{\mathcal{N}(\mu^\star_\alpha, h\mathbf{I})}$$

Posterior simulation and estimation with R I
Without variable selection

Parameters

$\beta = (-1, 0.5)$,

$\gamma = (0.8, 1.2)$,

$\rho = 0.9, \ \sigma = 1$

Covariables

$\mathbf{B} \sim \mathcal{N}(0, 0.7\mathbf{I}_2)$

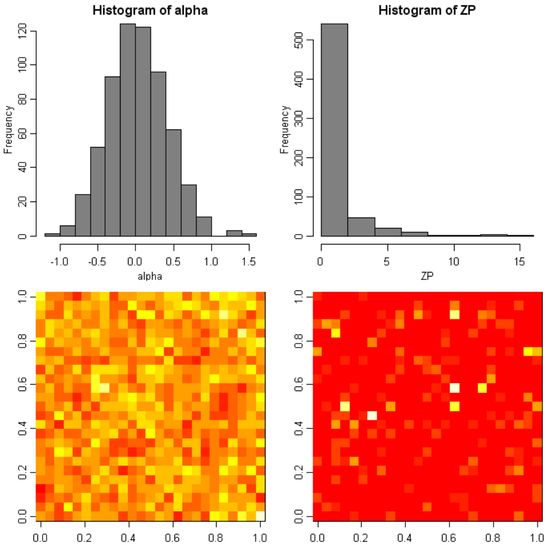
$\mathbf{X} \sim \mathcal{N}(0, 0.7\mathbf{I}_2)$

Data simulation

$\mathbf{C} \sim \mathcal{B}(\omega = \mathbf{B}\beta)$,

$\mathbf{P} \sim \mathcal{P}(\lambda = \mathbf{X}\gamma)$

$\mathbf{ZP} = (1 - \mathbf{C})\mathbf{P}$



Posterior simulation and estimation with R II

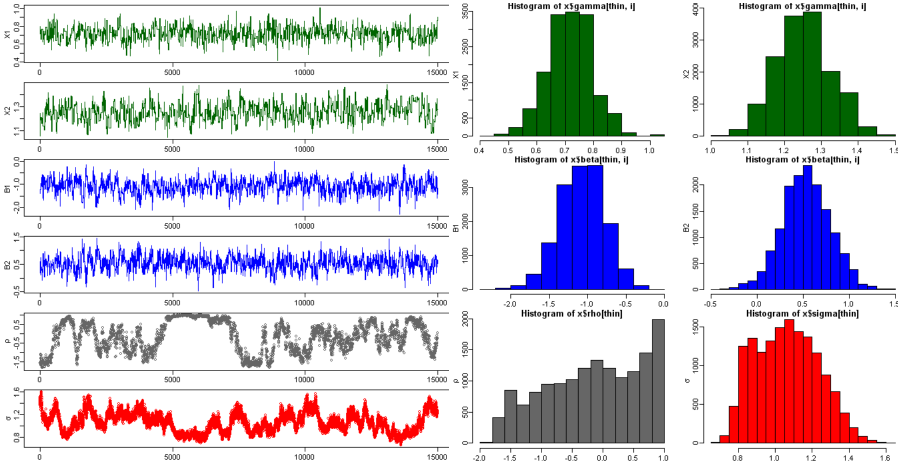
Without variable selection

Summary of MCMC samples (no variable selection)

```
Iterations: 20000, Burn-in phase: 5000,  Thinning number: 100

Coefficients in Binomial distribution
      Mean      Sd      2.5%  Median   97.5%
B1  -1.088  0.294  -1.6698  -1.090  -0.578
B2   0.546  0.238   0.0701   0.509   1.040

Coefficients in Poisson distribution
     Mean       Sd    2.5%  Median  97.5%
X1  0.714  0.0786  0.556   0.711  0.873
X2  1.250  0.0761  1.082   1.249  1.401

Spatial parameters in CAR model
         Mean      Sd     2.5%  Median  97.5%
rho    -0.178  0.800  -1.673  -0.136   0.98
sigma  1.063  0.171   0.795   1.066   1.41
```

Posterior simulation and estimation with R III

Without variable selection



Posterior simulation and estimation with R I

With variable selection

Parameters

$\beta = (-1, 0.5, 0, 0, 0)$,

$\gamma = (0.8, 1.2, 0, 0, 0)$,

$\rho = 0.9, \ \sigma = 1$

Covariables

$\mathbf{B}' = (\mathbf{B}, \mathcal{N}(0, 0.7\mathbf{I}_3))$

$\mathbf{X}' = (\mathbf{X}, mathcalN(0, 0.7\mathbf{I}_2))$

Data simulation

$\mathbf{C} \sim \mathcal{B}(\omega = \mathbf{B}\beta)$,

$\mathbf{P} \sim \mathcal{P}(\lambda = \mathbf{X}\gamma)$

$\mathbf{ZP} = (1 - \mathbf{C})\mathbf{P}$

Posterior simulation and estimation with R II
With variable selection

Summary of MCMC samples for parameter η in variable selection

Variable selection in Binomial distribution

	Mean	Sd	2.5%	Median	97.5%
B1	0.947	0.225	0	1	1
B2	0.680	0.468	0	1	1
B3	0.533	0.501	0	1	1
B4	0.573	0.496	0	1	1
B5	0.467	0.501	0	0	1

Variable selection in Poisson distribution

	Mean	Sd	2.5%	Median	97.5%
X1	1.000	0.000	1	1	1
X2	1.000	0.000	1	1	1
X3	0.313	0.465	0	0	1
X4	0.640	0.482	0	1	1
X5	0.400	0.492	0	0	1

Conclusions

- Hierarchical Bayseian : flexible framework for modelling,
- Mutation/selection algorithms are robust and tunable,
- Computing realized in C language can be easily interfaced with R,
- All routines and more will be included in a free R package