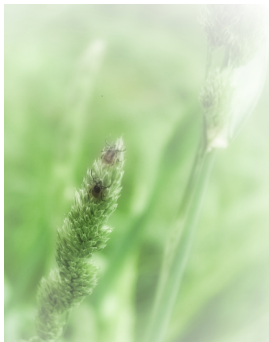


Modélisation spatiale de l'abondance de tiques *Ixodes ricinus*

Séverine Bord



Eric Parent, Isabelle Albert
UMR MIA - Equipe MORSE
AgroParisTech, Paris 5

Gwenaël Vourc'h
UMR EPIA
INRA VetAgroSup

AgroParisTech, Paris - 14/12/2017

Évaluer le **risque épidémiologique** lié à la présence de tiques dans un milieu (parcelle, forêt, région, pays) à partir de **l'estimation de l'abondance des populations** de tiques.

Évaluer le **risque épidémiologique** lié à la présence de tiques dans un milieu (parcelle, forêt, région, pays) à partir de **l'estimation de l'abondance des populations** de tiques.

Pourquoi chercher à estimer l'abondance de tiques ?

Évaluer le **risque épidémiologique** lié à la présence de tiques dans un milieu (parcelle, forêt, région, pays) à partir de **l'estimation de l'abondance des populations** de tiques.

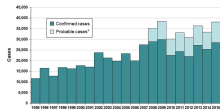
Pourquoi chercher à estimer l'abondance de tiques ?

- **transmission de nombreux agents pathogènes** comme *Borrelia* (Maladie de Lyme),

Évaluer le **risque épidémiologique** lié à la présence de tiques dans un milieu (parcelle, forêt, région, pays) à partir de **l'estimation de l'abondance des populations** de tiques.

Pourquoi chercher à estimer l'abondance de tiques ?

- **transmission de nombreux agents pathogènes** comme *Borrelia* (Maladie de Lyme),
- **Maladie de Lyme :**
 - une des maladies **les plus prévalentes** dans les zones tempérées de l'hémisphère nord (> 26000 nouveaux cas par an en France)
 - symptômes : érythème migrant, douleurs articulaires, fièvre, troubles neurologiques, fatigue, ...



Question :

Comment répartir l'échantillonnage dans l'espace de façon **optimale** pour estimer le risque de maladie lié aux tiques, **l'abondance de tiques et sa variabilité**

Question :

Comment répartir l'échantillonnage dans l'espace de façon **optimale** pour estimer le risque de maladie lié aux tiques, **l'abondance de tiques et sa variabilité**

Contraintes :

- **capacités d'échantillonnage limitées** (main d'oeuvre, temps, ...)
- la population de tique est **spatialement agrégée**
- covariables majeures associées à l'abondance des tiques sont connues

Question :

Comment répartir l'échantillonnage dans l'espace de façon **optimale** pour estimer le risque de maladie lié aux tiques, **l'abondance de tiques et sa variabilité**

Contraintes :

- **capacités d'échantillonnage limitées** (main d'oeuvre, temps, ...)
- la population de tique est **spatialement agrégée**
- covariables majeures associées à l'abondance des tiques sont connues

⇒ **modélisation de la structure spatiale**

Les données

Descriptif des
données

GLM Binomial
Négatif

Analyse des résidus
du GLM

Les données

Modélisation spatiale de l'abondance

Séverine Bord

La problématique biologique

Données et GLM BN

Les données

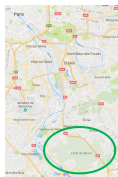
Descriptif des données

GLM Binomial Négatif

Analyse des résidus du GLM

Modélisation de la dépendance spatiale

Conclusion et perspectives

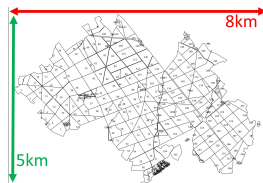


⇒ Données collectées en 2011 dans la forêt de Sénart.



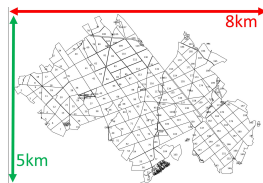
⇒ Données collectées en 2011 dans la forêt de Sénart.

- Unité d'observation : tranche de $10m^2$ ($1m \times 10m$)



⇒ Données collectées en 2011 dans la forêt de Sénart.

- Unité d'observation : tranche de 10m^2 ($1\text{m} \times 10\text{m}$)
- Mesure d'intérêt : nombre de tiques collectées sur une tranche (10m^2)
⇒ considéré comme un bon indicateur d'abondance.



⇒ Données collectées en 2011 dans la forêt de Sénart.

- Unité d'observation : tranche de 10m^2 ($1\text{m} \times 10\text{m}$)
- Mesure d'intérêt : nombre de tiques collectées sur une tranche (10m^2)
⇒ considéré comme un bon indicateur d'abondance.

Méthode d'échantillonnage :



Méthode du drapeau



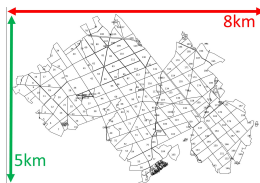
Tiques à l'affût : 2 adultes femelles (*Ixodes ricinus*)

Les données

Descriptif des
données

GLM Binomial
Négatif

Analyse des résidus
du GLM



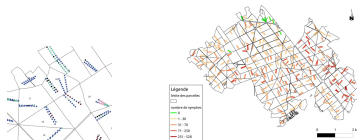
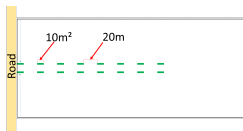
- 219 parcelles (environ 500m × 500m),

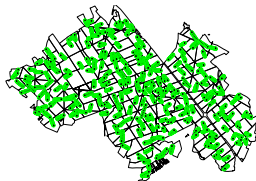


- **219 parcelles** (environ 500m × 500m),
- chaque parcelle : **16 tranches** (10m²) échantillonnées

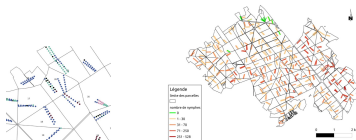
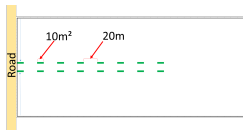


- 219 parcelles (environ 500m × 500m),
- chaque parcelle : 16 tranches (10m²) échantillonnées
- 16 tranches réparties en 2 lignes de 8 tranches,
- 2 lignes géo-localisées, placées perpendiculairement à la lisière,





- **219 parcelles** (environ 500m × 500m),
- chaque parcelle : **16 tranches** (10m²) échantillonnées
- **16 tranches** réparties en 2 lignes de 8 tranches,
- **2 lignes** géo-localisées, placées perpendiculairement à la lisière,



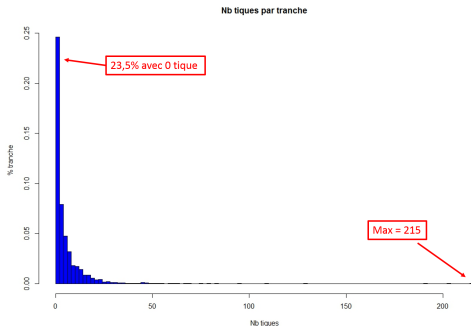
⇒ Au total **3504 tranches** ont été échantillonnées.

Variable d'intérêt :

Y_{iP} : nb de tiques collectées sur la tranche i dans parcelle P

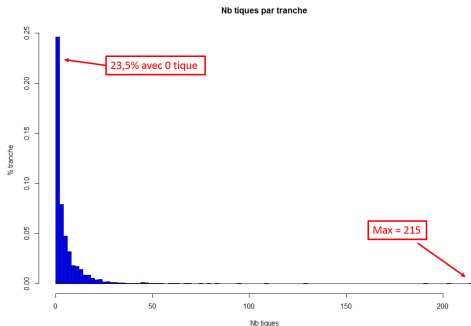
Variable d'intérêt :

Y_{iP} : nb de tiques collectées sur la tranche i dans parcelle P



Variable d'intérêt :

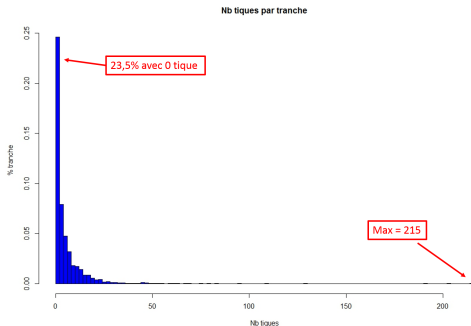
Y_{iP} : nb de tiques collectées sur la tranche i dans parcelle P



Données sur-dispersées.

Variable d'intérêt :

Y_{iP} : nb de tiques collectées sur la tranche i dans parcelle P



Données sur-dispersées.

⇒ distribution Binomiale Négative qui intègre un paramètre de dispersion.

Poisson :

$$y \sim P(\lambda)$$

$$E(y) = V(y) = \lambda$$

Poisson :

$$y \sim P(\lambda)$$

$$E(y) = V(y) = \lambda$$

Binomial Négatif :

$$y \sim P(\lambda) \text{ avec } \lambda \sim \Gamma(k, \mu)$$

$$E(y) = \mu; V(y) = \mu + \mu^2/k \text{ avec } k^{-1} \text{ paramètre de dispersion}$$

Poisson :

$$y \sim P(\lambda)$$

$$E(y) = V(y) = \lambda$$

Binomial Négatif :

$$y \sim P(\lambda) \text{ avec } \lambda \sim \Gamma(k, \mu)$$

$$E(y) = \mu; V(y) = \mu + \mu^2/k \text{ avec } k^{-1} \text{ paramètre de dispersion}$$

Modèle GLM Binomial Négatif

- y_i , le nombre de tiques collectées sur la tranche i , $i \in [1; 3504]$

Poisson :

$$y \sim P(\lambda)$$

$$E(y) = V(y) = \lambda$$

Binomial Négatif :

$$y \sim P(\lambda) \text{ avec } \lambda \sim \Gamma(k, \mu)$$

$$E(y) = \mu; V(y) = \mu + \mu^2/k \text{ avec } k^{-1} \text{ paramètre de dispersion}$$

Modèle GLM Binomial Négatif

- y_i , le nombre de tiques collectées sur la tranche i , $i \in [1; 3504]$

$$\log(E(y_i)) = \beta_0 + \sum_{j=1}^8 \beta_j \cdot X_j,$$

- 8 covariables, 2 niveaux :

Poisson :

$$y \sim P(\lambda)$$

$$E(y) = V(y) = \lambda$$

Binomial Négatif :

$$y \sim P(\lambda) \text{ avec } \lambda \sim \Gamma(k, \mu)$$

$$E(y) = \mu; V(y) = \mu + \mu^2/k \text{ avec } k^{-1} \text{ paramètre de dispersion}$$

Modèle GLM Binomial Négatif

- y_i , le nombre de tiques collectées sur la tranche i , $i \in [1; 3504]$

$$\log(E(y_i)) = \beta_0 + \sum_{j=1}^8 \beta_j \cdot X_j$$

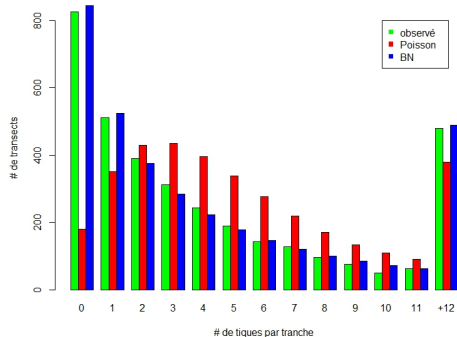
- 8 covariables, 2 niveaux :
 - parcelle (saturation déficit, ab. chevreuil, ab. sanglier, couverture forestière, nature du sol)
 - tranche (végétation, rang d'échantillonnage)

Estimations :

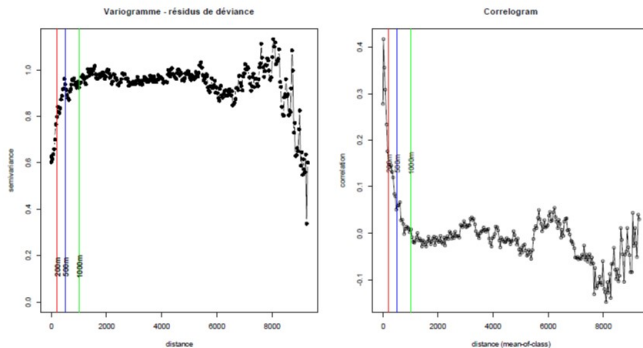
Facteurs	Modalités	Référence	OR	ic95i	ic95s	signif
Déficit humidité	[9,1,13,4] [13,5,22,2]	[4,4,9,0]	0,96 0,84	0,87 0,75	1,06 0,93	ns ***
chevreuil	[4,7] [8,14]	[0,3]	1,03 1,21	0,92 1,04	1,16 1,42	ns ***
sangliers	[5,10] [11,30]	[0,4]	0,92 0,46	0,83 0,41	1,03 0,53	ns ***
Dist Tani	[3,2,5,3] [5,3,8,5]	[0,3,2]	0,84 1,57	0,75 1,35	0,93 1,82	*** ***
Végétation	Fougères		1,07	0,75	1,52	ns
	Bruyère		0,50	0,32	0,76	***
	Herbe	Terre	1,39	1,05	1,86	***
	Arbustes		3,26	1,93	5,51	***
	Feuilles mortes		1,21	0,91	1,61	ns
Couvert forestier	chêne	Conifères	1,22	1,01	1,47	***
	Autres feuillus		0,71	0,55	0,91	***
Sol	Argileux	Sableux	1,02	0,82	1,27	ns
	limoneux		1,75	1,54	1,98	***
Ordre éch.	t3-t8	t1-t2	1,44	1,31	1,59	***

Prédictions :

Ajustement modèle Poisson vs BN



Variogramme et correlogramme des résidus de déviance du GLM BN



⇒ Dépendance spatiale jusqu'à 700 m.

Approche naïve :

La
problématique
biologique

Données et
GLM BN

Modélisation
de la
dépendance
spatiale

Approche naïve

Décomposition de la
dépendance spatiale

Démarche

Modèle CAR propre,
niveau parcelle

Inference Bayésienne
avec INLA

Conclusion et
perspectives

Approche naïve :

⇒ ajouter un effet aléatoire spatial s au modèle GLM,

Approche naïve :

⇒ ajouter un effet aléatoire spatial s au modèle GLM,
avec :

$$\log(E(y_i)) = \beta_0 + \sum_{j=1}^g \beta_j \cdot X_{ij} + s_i$$

Approche naïve :

⇒ ajouter un effet aléatoire spatial s au modèle GLM,

avec :

$$\log(E(y_i)) = \beta_0 + \sum_{j=1}^8 \beta_j \cdot X_{ij} + s_i$$

$$\text{où } s \sim \mathcal{N}(0, \Sigma), \text{ avec } \Sigma = \begin{pmatrix} \sigma^2 & \dots & \sigma^2 \exp^{-\rho d_{ii'}} \\ \vdots & \sigma^2 & \vdots \\ \sigma^2 \exp^{-\rho d_{ii'}} & \dots & \sigma^2 \end{pmatrix}_{3504 \times 3504}$$

Approche naïve :

⇒ ajouter un effet aléatoire spatial s au modèle GLM,

avec :

$$\log(E(y_i)) = \beta_0 + \sum_{j=1}^8 \beta_j \cdot X_{ij} + s_i$$

$$\text{où } s \sim \mathcal{N}(0, \Sigma), \text{ avec } \Sigma = \begin{pmatrix} \sigma^2 & \dots & \sigma^2 \exp^{-\rho d_{ii'}} \\ \vdots & \sigma^2 & \vdots \\ \sigma^2 \exp^{-\rho d_{ii'}} & \dots & \sigma^2 \end{pmatrix}_{3504 \times 3504}$$

⇒ modélisation de la **dépendance spatiale** par une fonction de **covariance de Matern** (exponentiel) :

$$\text{Cov}(Y(s_i), Y(s_{i'})) \equiv C(d_{ii'}) = \sigma^2 \exp^{-\rho d_{ii'}}$$

pour $i \neq i'$ et $d_{ii'}$ distance entre s_i et $s_{i'}$

avec

- $d_{ii'}$ la distance entre i et i' ,
- σ^2 , l'erreur de mesure,
- et ρ le paramètre d'échelle.

Approche naïve :

⇒ ajouter un effet aléatoire spatial s au modèle GLM,

avec :

$$\log(E(y_i)) = \beta_0 + \sum_{j=1}^8 \beta_j \cdot X_{ij} + s_i$$

$$\text{où } s \sim \mathcal{N}(0, \Sigma), \text{ avec } \Sigma = \begin{pmatrix} \sigma^2 & \dots & \sigma^2 \exp^{-\rho d_{ii'}} \\ \vdots & \sigma^2 & \vdots \\ \sigma^2 \exp^{-\rho d_{ii'}} & \dots & \sigma^2 \end{pmatrix}_{3504 \times 3504}$$

⇒ modélisation de la **dépendance spatiale** par une fonction de **covariance de Matern** (exponentiel) :

$$\text{Cov}(Y(s_i), Y(s_{i'})) \equiv C(d_{ii'}) = \sigma^2 \exp^{-\rho d_{ii'}}$$

pour $i \neq i'$ et $d_{ii'}$ distance entre s_i et $s_{i'}$

avec

- $d_{ii'}$ la distance entre i et i' ,
- σ^2 , l'erreur de mesure,
- et ρ le paramètre d'échelle.

⇒ **méthodes MCMC** pour estimer les paramètres de la structure spatiale

⇒ pb de **grande dimension** (matrice variance-covariance (3504*3504))

⇒ Alternatives envisagées :

- ① découpage de la structure spatiale en 2 niveaux :
 - (a) INTER parcelle : structure de dépendance entre parcelle
 - (b) INTRA parcelle : structure de dépendance au sein de chaque parcelle (réduction de la dimension)
- ② INLA : Integrated Nested Laplace Approximation,
 - technique qui permet d'estimer des champs aléatoires Gaussiens par inférence Bayésienne et approximation de Laplace.
 - disponible dans un package R.

Notations :

- y_{iP} le nombre de tiques collectées sur tranche i dans la parcelle P ,
 - i numéro de la tranche, $i \in [1;16]$,
 - P numéro de la parcelle, $P \in [1;219]$,
- X_{iP} , covariables associées à la tranche iP ,
- X'_P , covariables associées à la parcelle P .

Notations :

- y_{iP} le nombre de tiques collectées sur tranche i dans la parcelle P ,
 - i numéro de la tranche, $i \in [1;16]$,
 - P numéro de la parcelle, $P \in [1;219]$,
- X_{iP} , covariables associées à la tranche iP ,
- X'_P , covariables associées à la parcelle P .

$$(y_{iP} | \lambda_{iP}) \sim \mathcal{P}(\lambda_{iP})$$

Notations :

- y_{iP} le nombre de tiques collectées sur tranche i dans la parcelle P ,
 - i numéro de la tranche, $i \in [1;16]$,
 - P numéro de la parcelle, $P \in [1;219]$,
- X_{iP} , covariables associées à la tranche iP ,
- X'_P , covariables associées à la parcelle P .

$$(y_{iP} | \lambda_{iP}) \sim \mathcal{P}(\lambda_{iP})$$

$$\log(\lambda_{iP}) = \mu + X_{iP}\beta + X'_P\alpha + S_{iP}$$

Notations :

- y_{iP} le nombre de tiques collectées sur tranche i dans la parcelle P ,
 - i numéro de la tranche, $i \in [1;16]$,
 - P numéro de la parcelle, $P \in [1;219]$,
- X_{iP} , covariables associées à la tranche iP ,
- X'_P , covariables associées à la parcelle P .

$$(y_{iP} | \lambda_{iP}) \sim \mathcal{P}(\lambda_{iP})$$

$$\log(\lambda_{iP}) = \mu + X_{iP}\beta + X'_P\alpha + S_{iP}$$

Effet Spatial : $S_{iP} = T_{iP} + P_P$

Effet Spatial : $S_{iP} = T_{iP} + P_P$

Effet Spatial : $S_{iP} = T_{iP} + P_P$

INTRA parcelle : modèle de covariance Matérn
(exponentiel) : $T_{iP} \sim \mathcal{N}(0, \Sigma_P)$

Effet Spatial : $S_{iP} = T_{iP} + P_P$

INTRA parcelle : modèle de covariance Matérn
(exponentiel) : $T_{iP} \sim \mathcal{N}(0, \Sigma_P)$

$$\Sigma_P = \begin{pmatrix} \sigma^2 & \dots & \sigma^2 \exp^{-\rho d_{ii'}} \\ \vdots & \sigma_P^2 & \vdots \\ \sigma^2 \exp^{-\rho d_{ii'}} & \dots & \sigma^2 \end{pmatrix}$$

Hyper paramètres :

- ρ , paramètre d'échelle
- σ^2 , erreur de mesure (nugget)

(communs à toutes les parcelles)
16x16

Effet Spatial : $S_{iP} = T_{iP} + P_P$

INTRA parcelle : modèle de covariance Matérn
(exponentiel) : $T_{iP} \sim \mathcal{N}(0, \Sigma_P)$

$$\Sigma_P = \begin{pmatrix} \sigma^2 & \dots & \sigma^2 \exp^{-\rho d_{ii'}} \\ \vdots & \sigma_P^2 & \vdots \\ \sigma^2 \exp^{-\rho d_{ii'}} & \dots & \sigma^2 \end{pmatrix}$$

Hyper paramètres :

- ρ , paramètre d'échelle
- σ^2 , erreur de mesure (nugget)

(communs à toutes les parcelles)
16x16

Avantages :

- estimation parcelle par parcelle \rightarrow réduction de la matrice de dépendance (16*16)
- amélioration de la puissance statistique pour estimer ρ et σ .

Effet Spatial : $S_{iP} = T_{iP} + P_P$

Effet Spatial : $S_{iP} = T_{iP} + P_p$

INTER parcelle : l'effet spatial P_p d'une parcelle p dépend de ses voisins (CAR propre)

$$P_p | P_{-p} \sim \mathcal{N}\left(\frac{\phi \sum_{j \in \delta_p} a_{pj} P_j}{a_{p+}}, \frac{\tau^2}{a_{p+}}\right), j \neq p$$

- ϕ , pondération de la dépendance spatiale
- P_{-p} , effet spatial autres parcelles $\neq p$
- $\sum_{j \in \delta_p} a_{pj} P_j$ somme des effets spatiaux des voisins de p
- a_{p+} , nombre de voisins de p
- a_{pj} , 1 si j voisin de p , 0 sinon
- τ^2 , variance.

Modèle complet, DAG

Modélisation spatiale de l'abondance

Séverine Bord

La problématique biologique

Données et GLM BN

Modélisation de la dépendance spatiale

Approche naïve

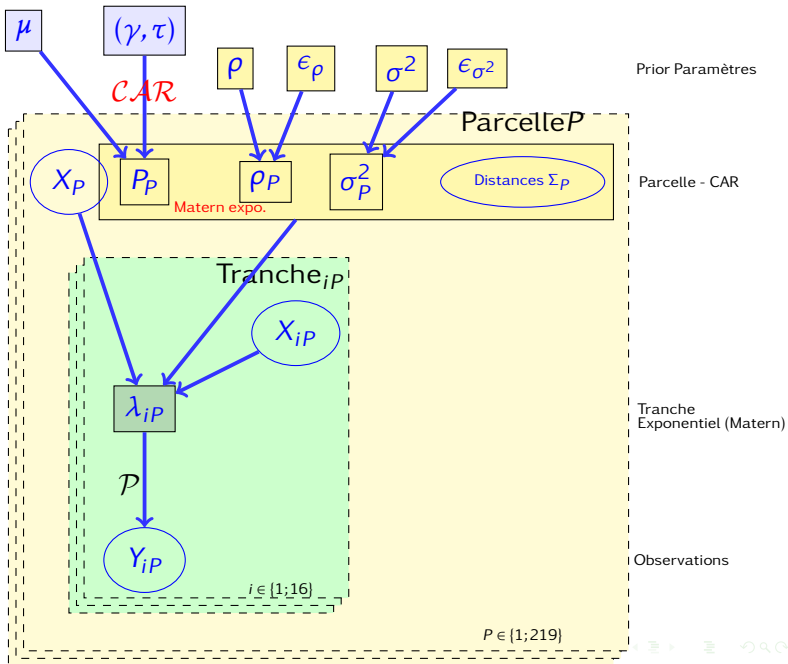
Décomposition de la dépendance spatiale

Démarche

Modèle CAR propre, niveau parcelle

Inference Bayésienne avec INLA

Conclusion et perspectives



- 1 Estimation du **Modèle CAR : inter parcelle** : agrégation des données au niveau parcelle,
- 2 Estimation du **Modèle Matérn exponentiel : intra parcelle**
- 3 Association des niveaux **Parcelle** et **tranche** (1 et 2).

Modèle CAR : inter parcelle

⇒ **1. Agrégation des données** à l'échelle de la parcelle :

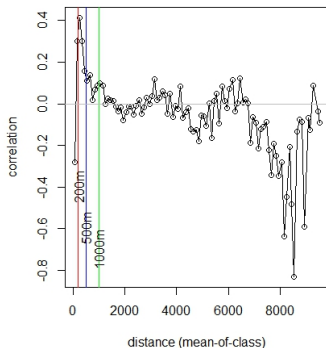
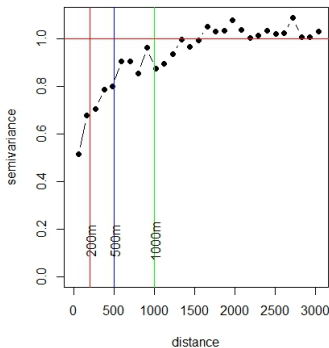
- observations : somme des captures par parcelle

$$y_P = \sum_{i=1}^{16} y_{iP}, P \in [1, 219]$$

- Covariables parcelle : ok
- Covariables tranche : modalité majoritaire

⇒ 2. Modèle GLM Binomial Négatif - Parcelle

⇒ 3. Analyse des résidus du modèle GLM BN - Parcelle



⇒ Dépendance spatiale jusqu'à 700m.

⇒ 4. Implémentation du **Modele CAR** sous Winbugs

⇒ 4. Implémentation du **Modele CAR** sous Winbugs

Problème : modèle complet (8 covariables) ne tourne pas.

⇒ 4. Implémentation du **Modèle CAR** sous Winbugs

Problème : modèle complet (8 covariables) ne tourne pas.

Solutions testées :

⇒ 4. Implémentation du **Modele CAR** sous Winbugs

Problème : modèle complet (8 covariables) ne tourne pas.

Solutions testées :

- modèle avec 2 covariables tourne mais **nécessite des priors très informatifs**.

⇒ 4. Implémentation du **Modèle CAR** sous Winbugs

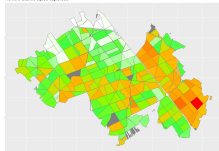
Problème : modèle complet (8 covariables) ne tourne pas.

Solutions testées :

- modèle avec 2 covariables tourne mais **nécessite des priors très informatifs**.
- initialiser le modèle complet avec les **estimations du modèle GLM Binomial négatif** estimé précédemment :
 - pour l'effet spatial parcelle P_p avec les résidus
 - pour les β avec estimations $\hat{\beta}$

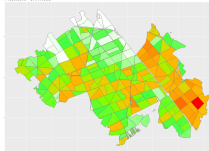
Résultats modèle CAR :

Nombre total de liques capturées



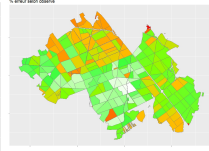
Les observations

Prédictions - CAR modèle



Les prédictions

% erreur selon observé



Les erreurs

Étape 1 : Agrégation au niveau parcelle

Modélisation spatiale de l'abondance

Séverine Bord

La problématique biologique

Données et GLM BN

Modélisation de la dépendance spatiale

Approche naïve

Décomposition de la dépendance spatiale

Démarche

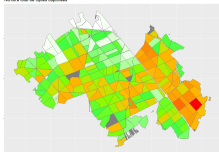
Modèle CAR propre, niveau parcelle

Inference Bayésienne avec INLA

Conclusion et perspectives

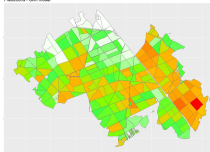
Résultats modèle CAR :

Nombre total de tiques capturées



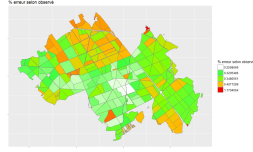
Les observations

Prédictions - CAR modèle



Les prédictions

% erreur selon observ

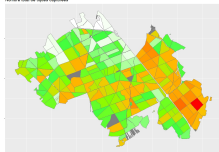


Les erreurs

Les étapes suivantes...

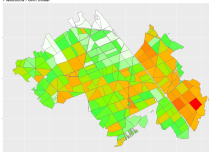
Résultats modèle CAR :

Nombre total de tiques capturées



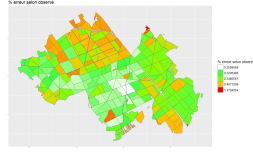
Les observations

Prédictions - CAR modèle



Les prédictions

% erreur selon observ



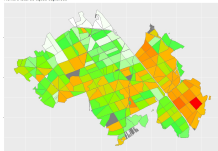
Les erreurs

Les étapes suivantes...

- modéliser la structure de dépendance "INTRA parcelle",

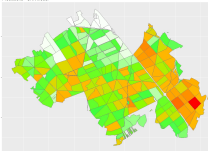
Résultats modèle CAR :

Nombre total de tiques capturées



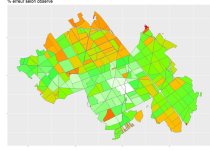
Les observations

Prédictions - CAR modèle



Les prédictions

% erreur selon observé



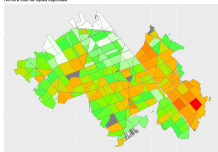
Les erreurs

Les étapes suivantes...

- modéliser la structure de dépendance "INTRA parcelle",
- agréger les modèles "INTER parcelle"(CAR) et "INTRA parcelle"(exponentiel)

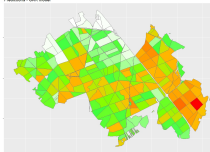
Résultats modèle CAR :

Nombre total de tiques capturées



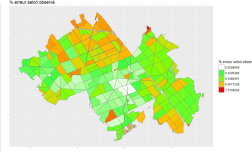
Les observations

Prédictions - CAR modèle



Les prédictions

% erreur selon observé



Les erreurs

Les étapes suivantes...

- modéliser la structure de dépendance "INTRA parcelle",
- agréger les modèles "INTER parcelle"(CAR) et "INTRA parcelle"(exponentiel)
- comparer avec INLA

INLA : Integrated Nested Laplace Approximation,

- technique qui permet d'estimer des **champs aléatoires Gaussiens par inférence Bayésienne et approximation de Laplace.**
- disponible dans un package R.

Approche SPDE (Stochastic Partial Differential Equation, Lindgren *et al.* (2011)) permet de faire des estimations à partir d' :

- une **triangulation de l'espace** avec G sommets
- une structure qui approche le champs latent Gaussien par un **GMRF** (Gaussian Markov Random Field) et une matrice de **précision sparse** (basée sur voisinage).

⇒ permet des temps de calculs réduits.

Triangulation de l'espace ⇒ construction du maillage :

Triangulation de l'espace

⇒ construction du maillage :

- **Plus le maillage est fin et meilleure est approximation mais plus le temps de calcul est long**

Triangulation de l'espace

⇒ construction du maillage :

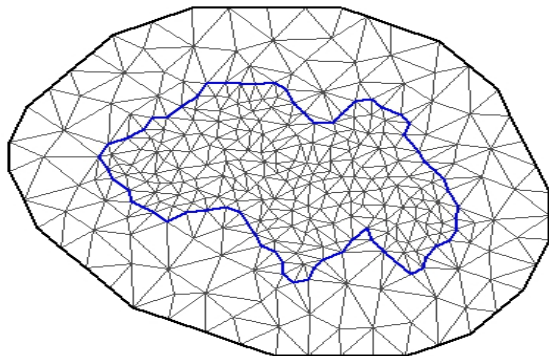
- **Plus le maillage est fin et meilleure est approximation** mais plus le temps de calcul est long
- **Il faut étendre le maillage en dehors du domaine spatial** pour éviter des effets de bords (variance des observations qui vont être augmentés aux frontières du domaine)

Triangulation de l'espace

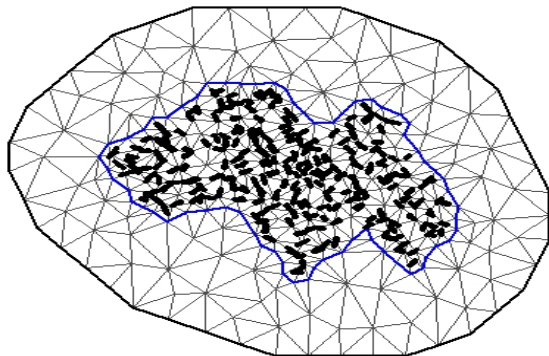
⇒ construction du maillage :

- **Plus le maillage est fin et meilleure est approximation** mais plus le temps de calcul est long
- **Il faut étendre le maillage en dehors du domaine spatial** pour éviter des effets de bords (variance des observations qui vont être augmentés aux frontières du domaine)

⇒ choix des paramètres du maillage.



226 triangles à l'intérieur, 333 en tout donc matrice A_{ijg} de taille (3504×333)



INLA permet d'implémenter des modèles GLM classiques,
tels que :

INLA permet d'implémenter des modèles GLM classiques, tels que :

- Poisson, binomial négatif

INLA permet d'implémenter des modèles GLM classiques, tels que :

- Poisson, binomial négatif

- distributions type 0 :

$$Prob(y|...) = p \times \mathbb{1}_{[y=0]} + (1 - p) \times \text{Poisson}(y|y > 0)$$

INLA permet d'implémenter des modèles GLM classiques, tels que :

- Poisson, binomial négatif

- distributions type 0 :

$$Prob(y|...) = p \times \mathbb{1}_{[y=0]} + (1 - p) \times \text{Poisson}(y|y > 0)$$

- distributions type 1 :

$$Prob(y|...) = p \times \mathbb{1}_{[y=0]} + (1 - p) \times \text{Poisson}(y)$$

⇒ Sur nos données :

⇒ Sur nos données :

- test de différents modèles GLM, on conserve le modèle binomial négatif (DIC).

⇒ Sur nos données :

- test de différents modèles GLM, on conserve le modèle binomial négatif (DIC).
- test d'un modèle CAR :
 - modèle qui tourne en 1h environ,
 - voir l'effet des choix des paramètres du maillage sur les estimations,
 - reste la validation du modèle,
 - construire des cartes de prévisions sur les zones non échantillonnées.

Notre modèle spatial n'est pas compliqué mais difficile à estimer.

Les approches bayésiennes classiques avec Winbugs peuvent être instables voire difficiles à mettre en oeuvre.

INLA,

- outil puissant qui gère les grandes dimensions,
- mais nécessite un paramétrage fin de l'algorithme.



Merci.