



Estimation des fréquences avalancheuses

(modélisation spatiale avec Winbugs)

N. ECKERT, R. KIES

Cemagref, unité Erosion Torrentielle, Neige et Avalanches

38 402 Saint Martin d'Hères, France

Nicolas.eckert@cemagref.fr

E. PARENT

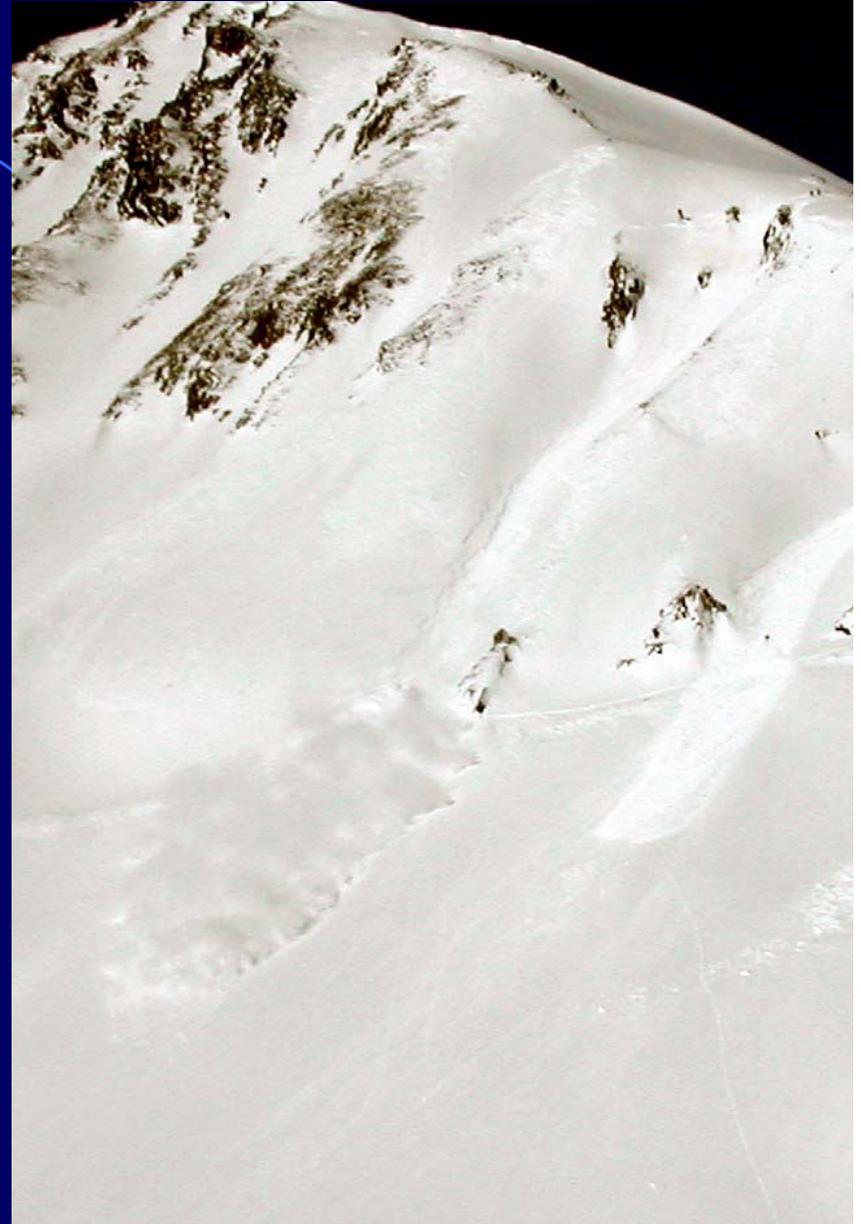
Equipe MORSE,

UMR 518 ENGREF/INRA/INAPG, 75732 Paris, France

Paris, 21 juin 2007

Plan de l'exposé

- Éléments de contexte
- Un modèle spatial de base
- Améliorations du modèle



Déclenchement sur site expérimental
Janvier 2004

Avalanches et statistique, un mariage nécessaire

- **Apprendre sur le phénomène** :
interprétation des données, lien météo-déclenchement...
- **Pallier l'insuffisance des connaissances physiques** :
loi de frottement, mécanismes de déclenchement...
- **Prévision et prédétermination** sont
liés à un cadre de travail probabiliste :

Prévision (Météo France) = réponse en temps réel (proba conditionnelle) : par massif

Prédétermination (Cemagref, RTM) = gestion à long terme (vision moyennée) : local



Prédétermination et aléa de référence

➔ En France longtemps +/- empirique (expertise)



Catastrophe de Montroc, 9 février 1999

➔ Définition plus rigoureuse de l'aléa de référence pour **zonage** et **dimensionnement**

➔ Guide PPR-A :
Période de retour

$$T(x) = \frac{1}{1 - P(X \leq x)}$$

Pourquoi s'intéresser aux fréquences avalancheuses?

Pour la période de retour, la fréquence avalancheuse est nécessaire, mais l'information locale est parfois insuffisante



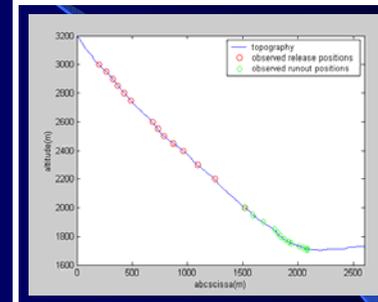
Partager l'information entre couloirs « proches »

Nombre y_k
d'avalanches
de l'année k

Fréquence
Avalancheuse $E[Y]$

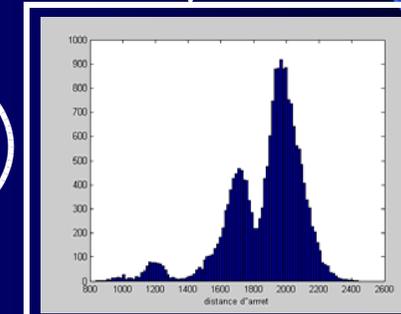
Période de retour

$$T(x) = \frac{1}{E[Y] \times (1 - P(X \leq x))}$$



Observations

X_1, X_2, \dots, X_N



Pdf $P(X=x)$
Ex: distance d'arrêt

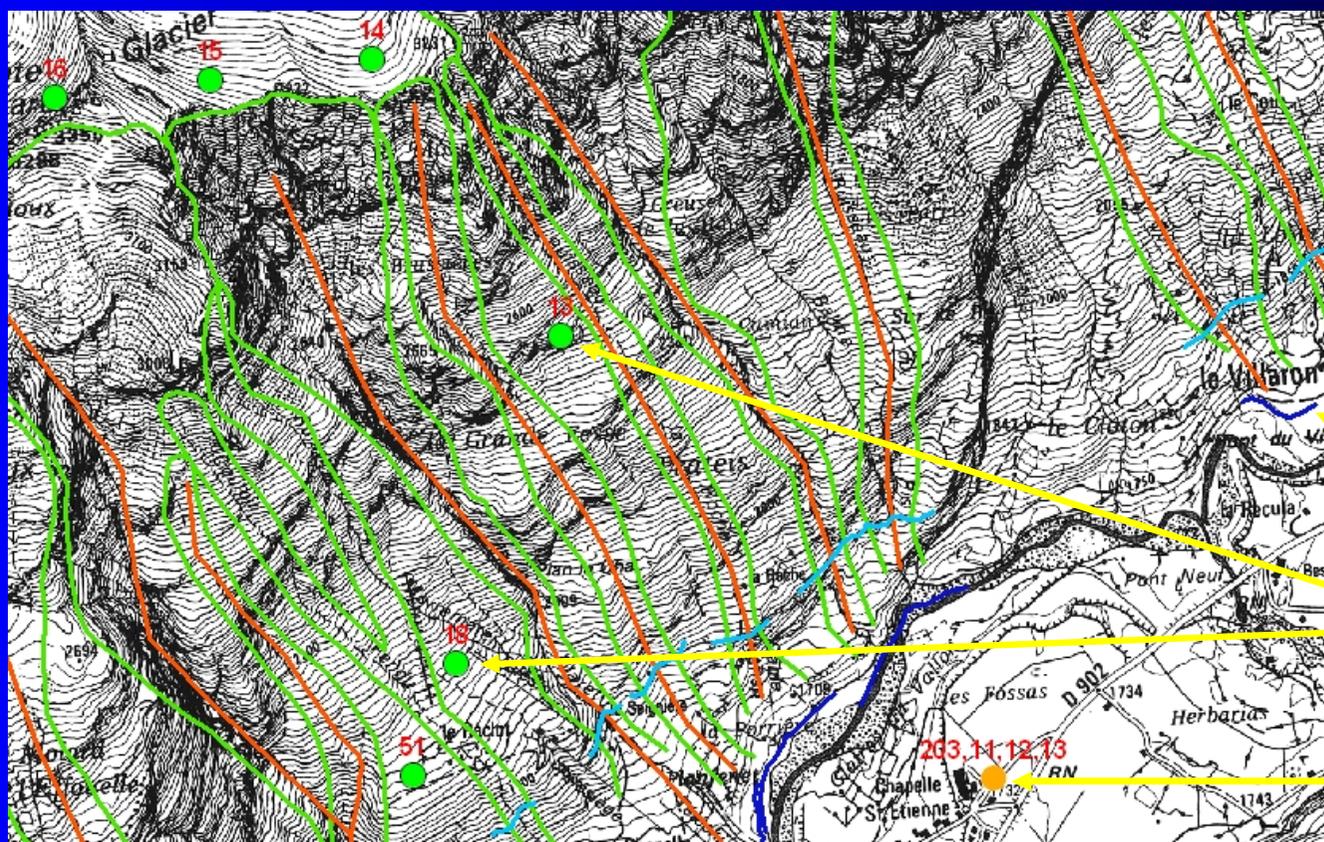
Etudier les motifs spatiaux est intéressant en soi :



compréhension du phénomène

L' Enquête Permanente sur les Avalanches (EPA)

- Près de 6000 couloirs dans les Alpes et les Pyrénées
- Suivi aussi exhaustif que possible, parfois depuis 1900, mais généralement après 1945
- Date des avalanches, informations quantitatives et qualitatives
- Qualité sujette à caution



Extrait de la
carte EPA

Seuil d'alerte

Couloirs suivis

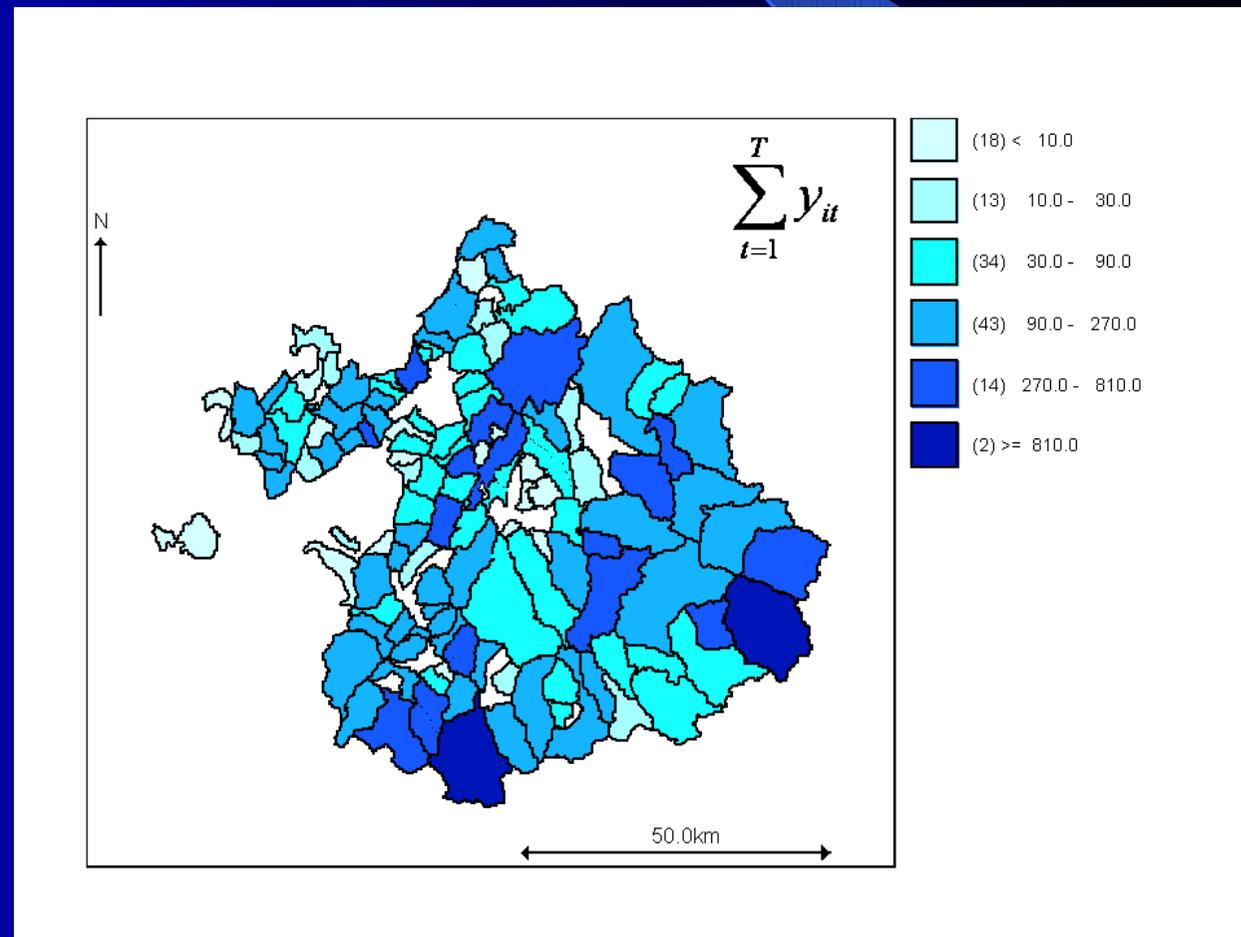
Point observation

Occurrences avalancheuses en Savoie

Savoie : - 124 communes avec des avalanches enregistrées
- 1347 couloirs avalancheux (1 a 49 par couloir)

Période: 1946-2005 divisée en 12 «tranches» de 5 ans

18 755 avalanches
- 0 à 352 par commune
et période
- Apparente stationnarité

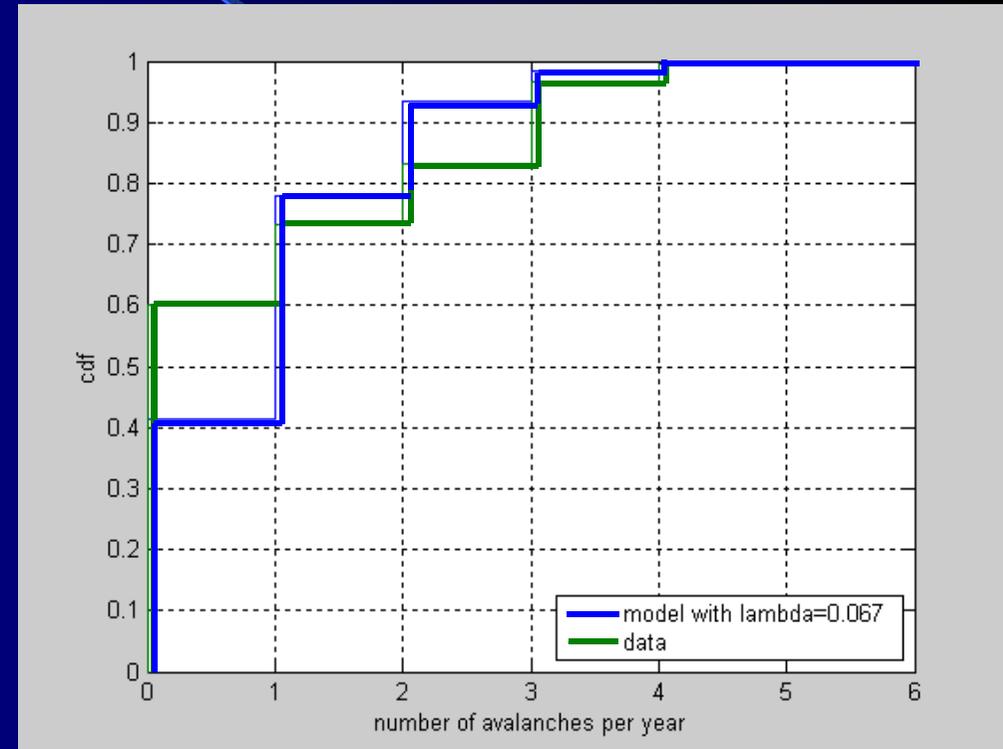


Modéliser les fréquences avalancheuses à l'échelle communale

Modèle poissonnien local homogène

$$y_{it} \sim P(\lambda_i) \quad \text{for } \begin{array}{l} i \in [1, N] \text{ communes} \\ t \in [1, T] \text{ périodes temporelles} \end{array}$$

répétitions indépendantes sur le temps mais structure spatiale



Modèle et cdf empirique pour un couloir

« Risque Relatif » multiplicatif

$$\lambda_i = e_i \times RR_i \quad \text{avec } e_i = \underbrace{\left(\frac{\sum_{i=1}^N \sum_{t=1}^T y_{it}}{T \times \sum_{i=1}^N c_i} \right)}_{\text{Normalisation avec le nombre de couloirs}} \times \frac{c_i}{\sum_{i=1}^N c_i}$$

Normalisation avec le nombre de couloirs

Deux couches hiérarchiques

Régression sur les covariables

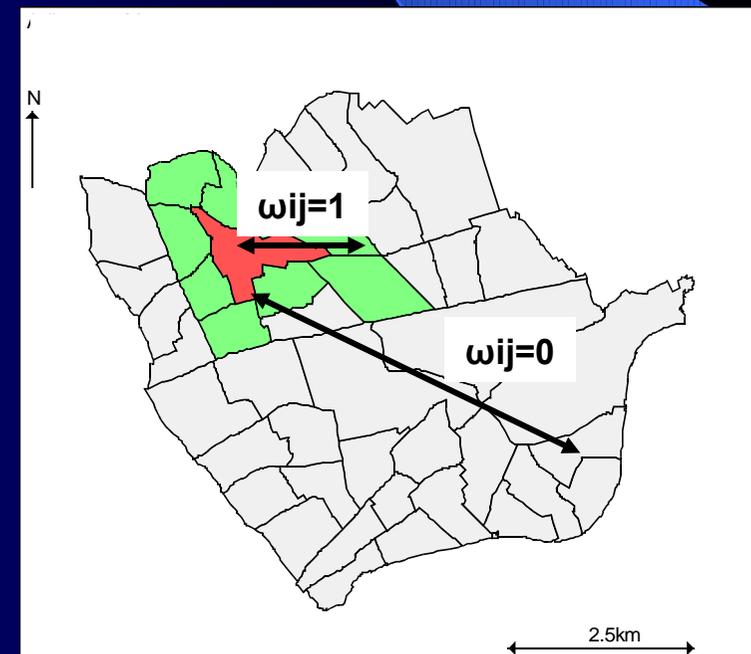
$$\ln(\lambda_i) = \ln(e_i) + \ln(RR_i) = \ln(e_i) + \sum_{j=1}^p \alpha_j \times X_{ij} + \varepsilon_i$$

X_{ij} : Covariables (Tendance spatiale)
Négligées dans un premier temps

Décomposition des résidus : $\varepsilon_i = U_i + V_i$

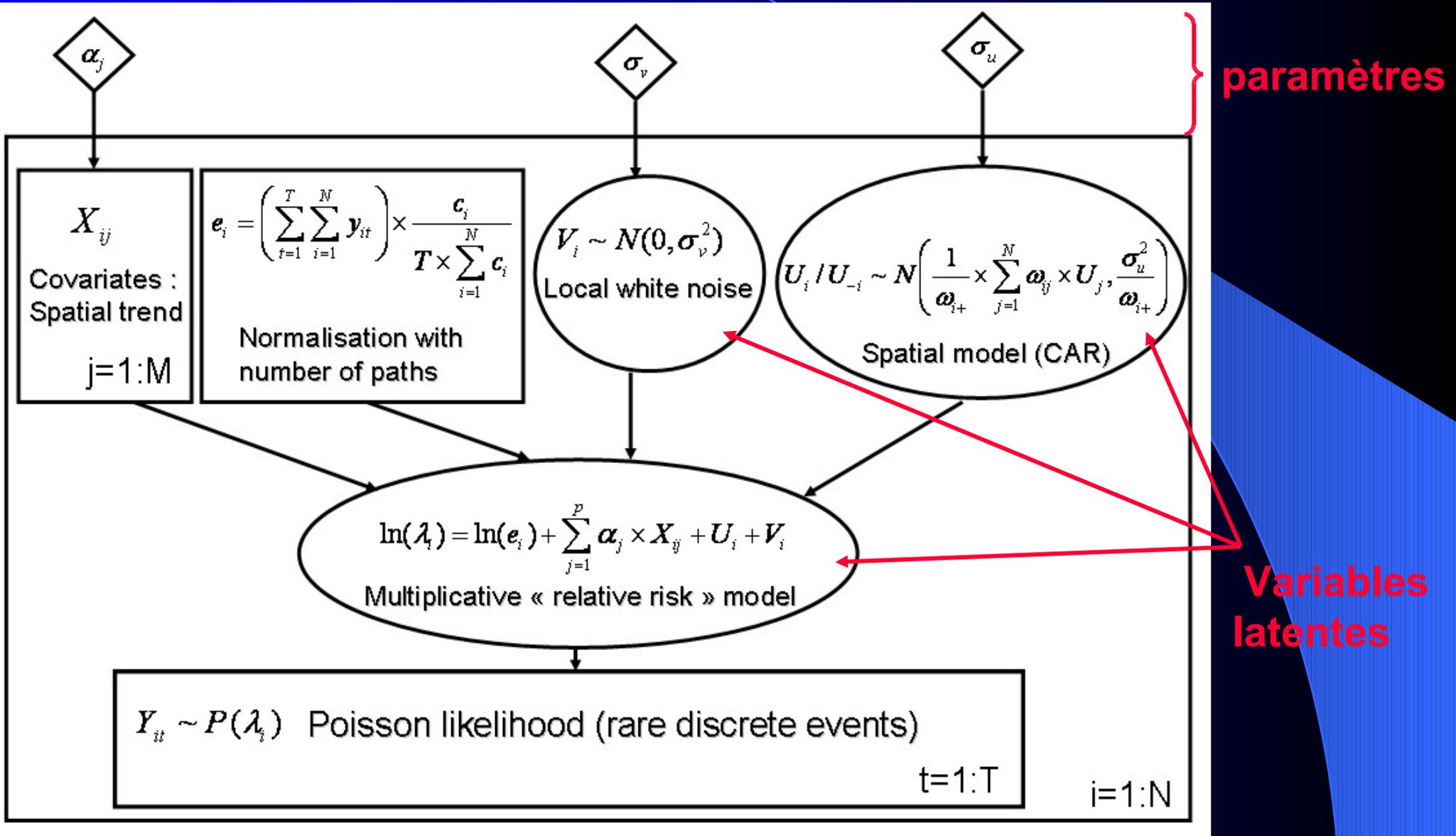
$$V_i / \sigma_V \sim N(0, \sigma_V) \quad \text{Bruit blanc local}$$

$$U_i / U_{-i} \sim N\left(\frac{1}{\omega_{i+}} \times \sum_{k=1}^N \omega_{ik} \times U_k, \frac{\sigma_u^2}{\omega_{i+}}\right) \quad \text{CAR Gaussien dit « intrinsèque »}$$



Voisinage de type « 0-1 »

Représentation graphique



Inférence bayésienne

→ Théorème de Bayes

Posterior joint

$$p(\theta / data) = \frac{l(data / \theta) \times p(\theta)}{\int l(data / \theta) \times p(\theta) \times d\theta}$$

Vraisemblance x prior
Normalisation

→ Paramètres $\sigma_U, \sigma_V, \alpha_j$
Variables latentes U_i, V_i, λ_i

$$[\lambda_i, u_i, v_i, \sigma_u, \sigma_v, \alpha_j / y_{it}, c_i, x_{ij}] \propto [\sigma_u, \sigma_v, \alpha_j] \times [u_i, v_i / \sigma_u, \sigma_v] \times [\lambda_i / c_i, \alpha_j, x_{ij}, u_i, v_i] \times \prod_{i=1}^I [y_{it} / \lambda_i]$$

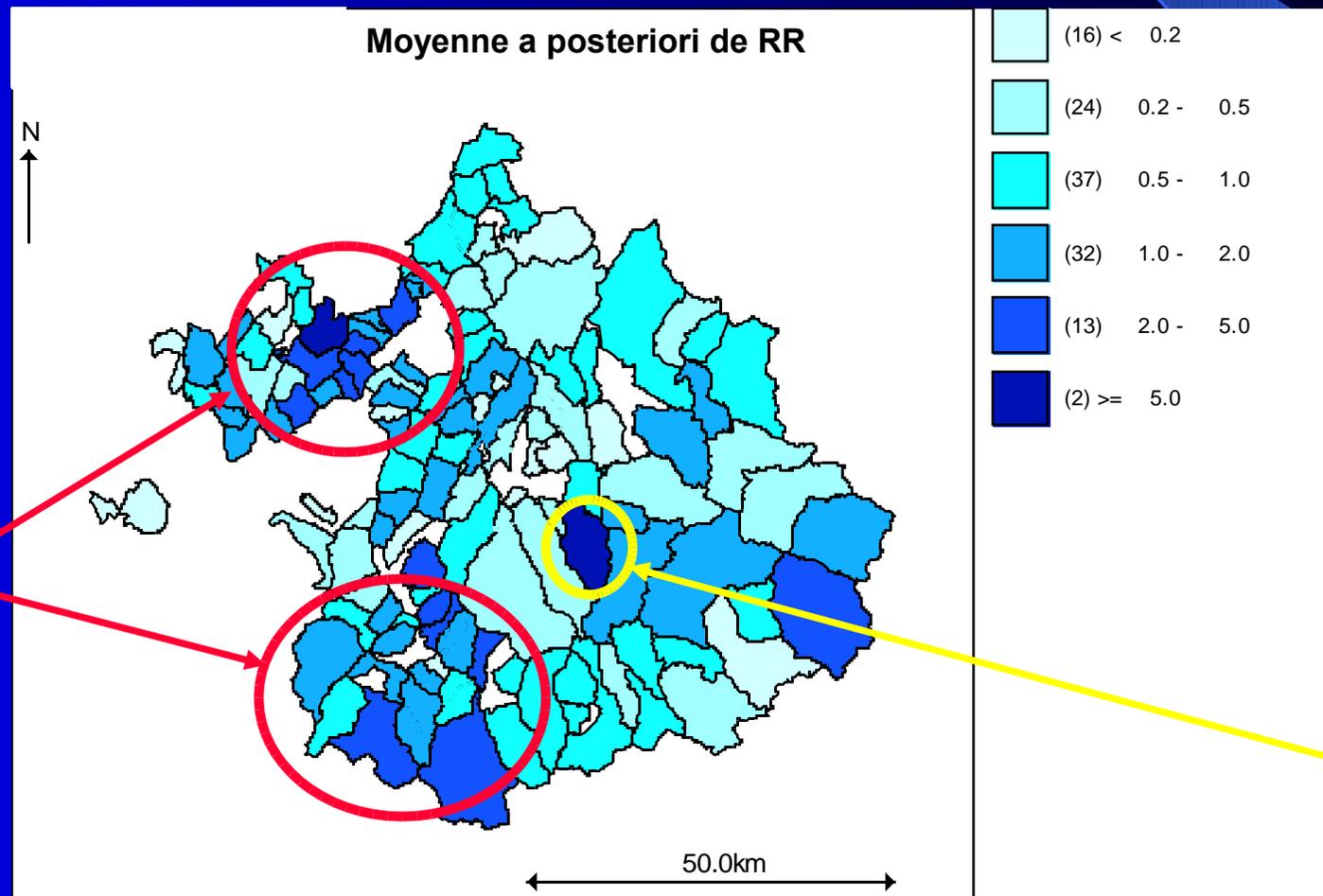
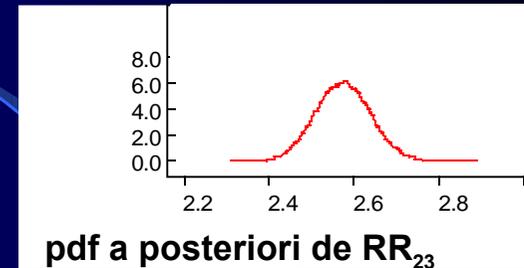
Distributions des variables latentes Vraisemblance

Risques relatifs d'avalanche

$$RR_i = \frac{\lambda_i}{e_i}$$



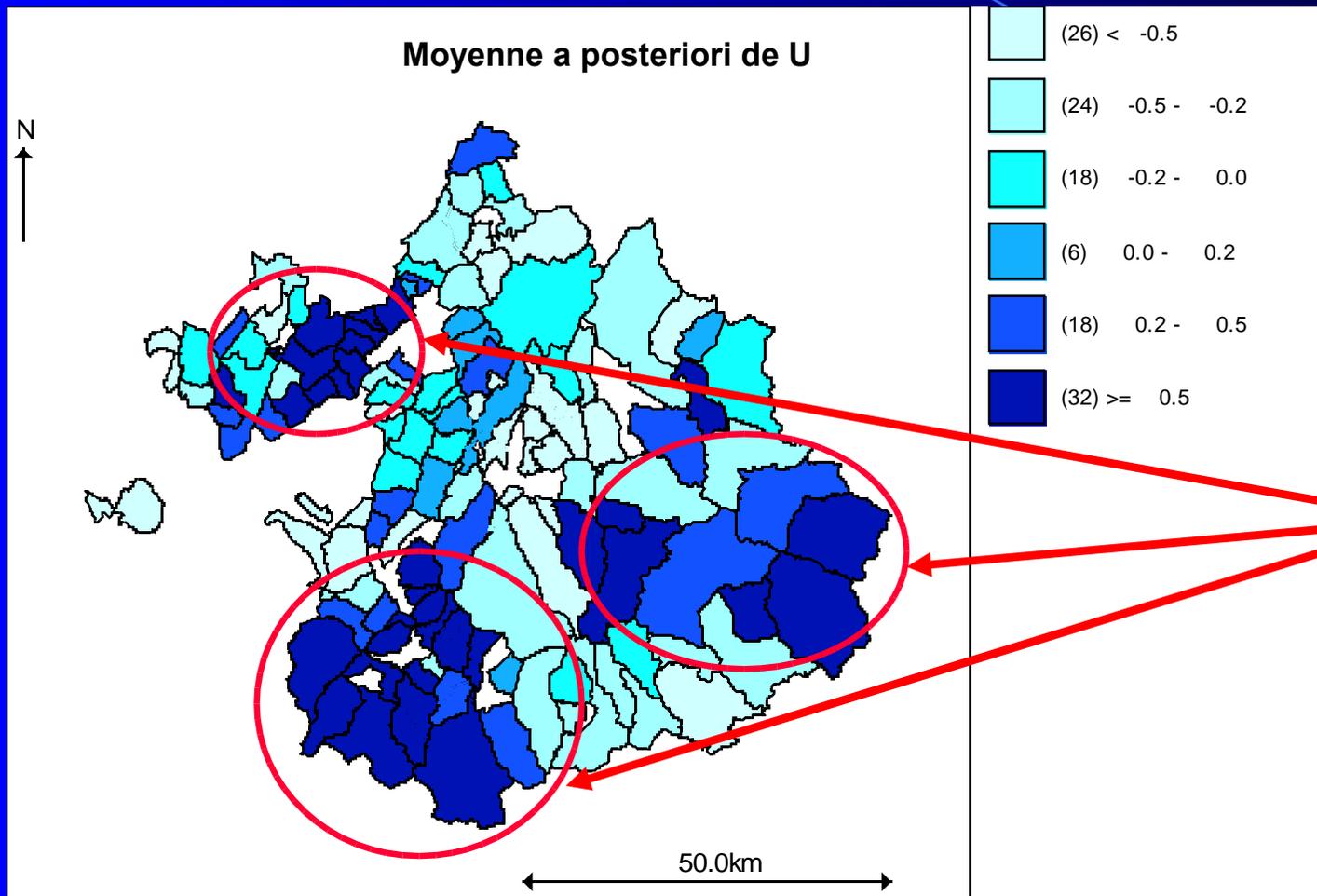
Si homogénéité spatiale
 $RR_i = 1$ partout



Excès spatialement structuré

Excès local

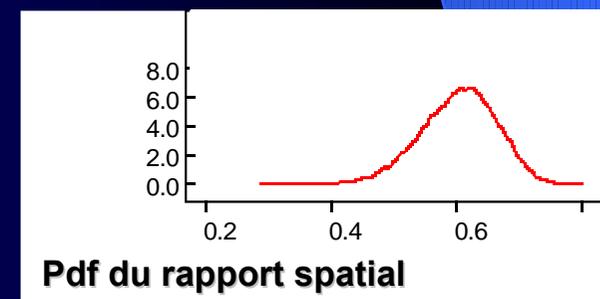
Motifs spatiaux



Structure spatiale
de l'activité
avalancheuse

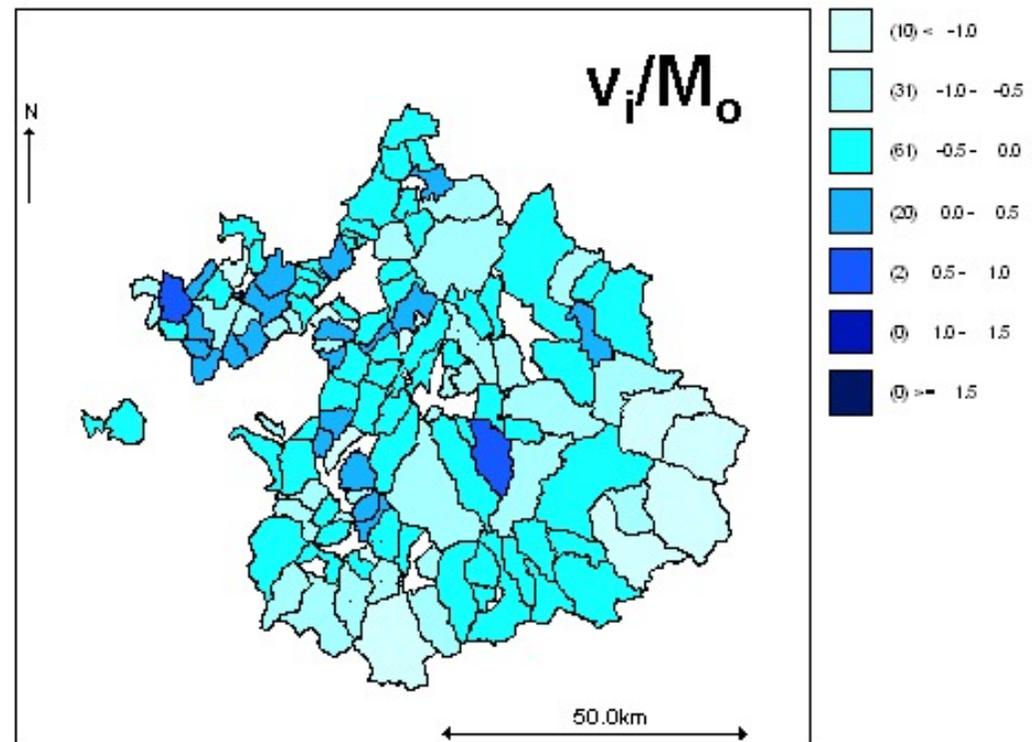
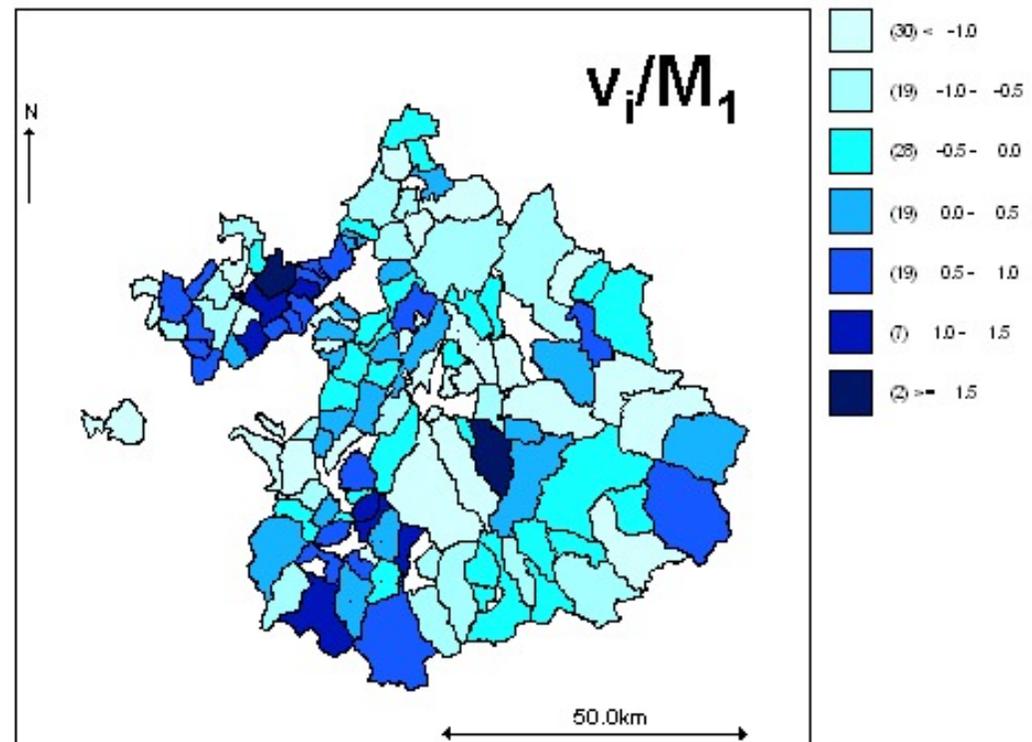
$$s_r = \frac{\text{Var}(U)}{\text{Var}(U) + \sigma_v^2} = 0.60$$

Variabilité spatiale
versus variabilité totale



Intérêt du terme spatial

	M_0 Local +spatial	M_1 local	M_2 spatial
DIC	17 039	17 045	19,830

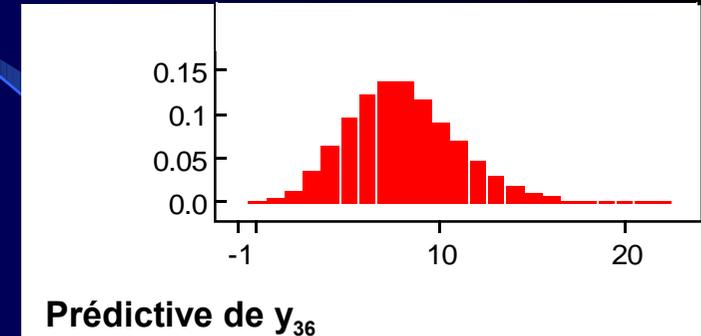


Loi prédictive

$$\left[y_i^{pred} / y_{ii}, c_i, x_{ij} \right] = \int l\left(y_i^{pred} / \lambda_i \right) \times \left[\lambda_i, u_i, v_i, \sigma_u, \sigma_v, \alpha_j / y_{ii}, c_i, x_{ij} \right] \times d\sigma_u \times d\sigma_v \times d\alpha_j$$

vraisemblance

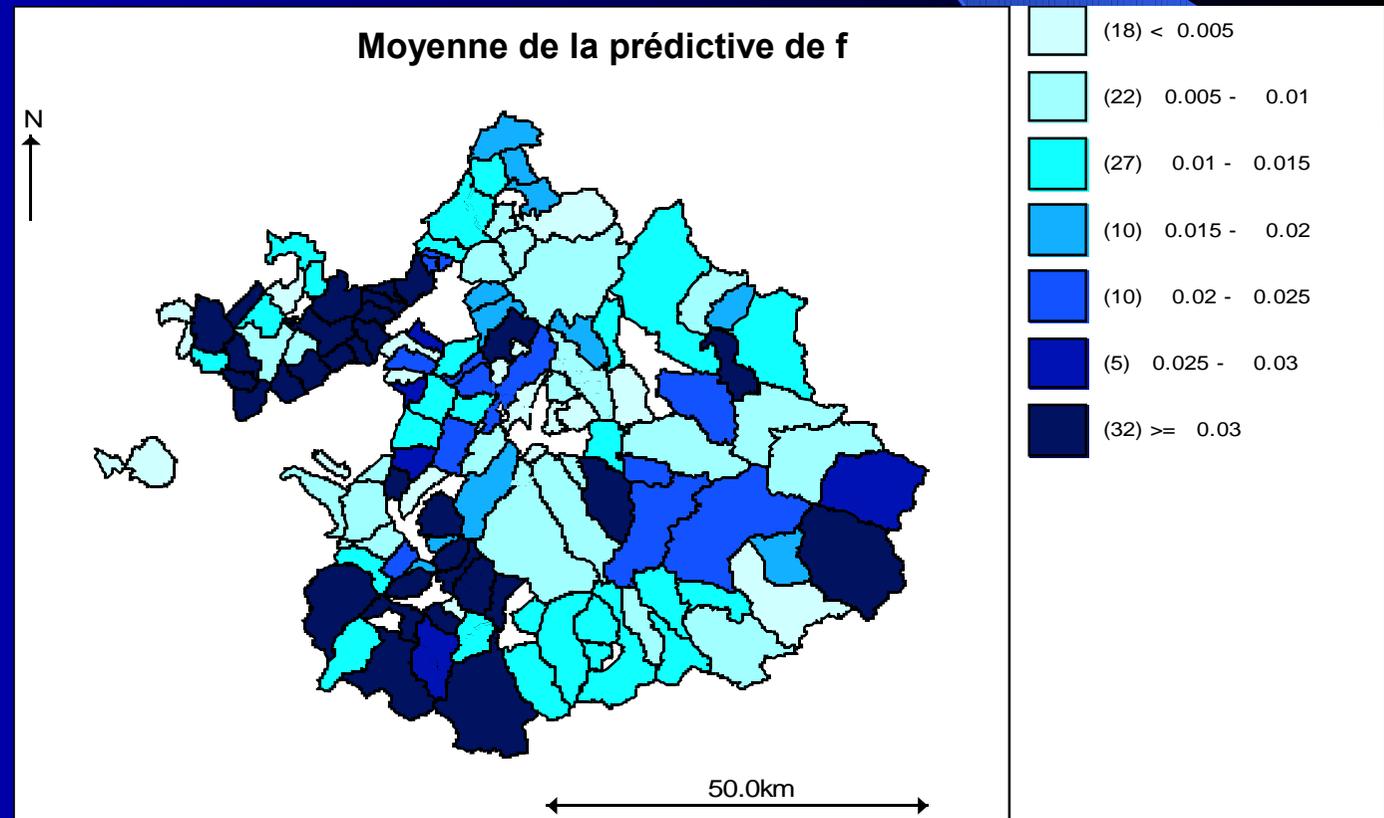
posterior



Prédictive
des fréquences
avalancheuses



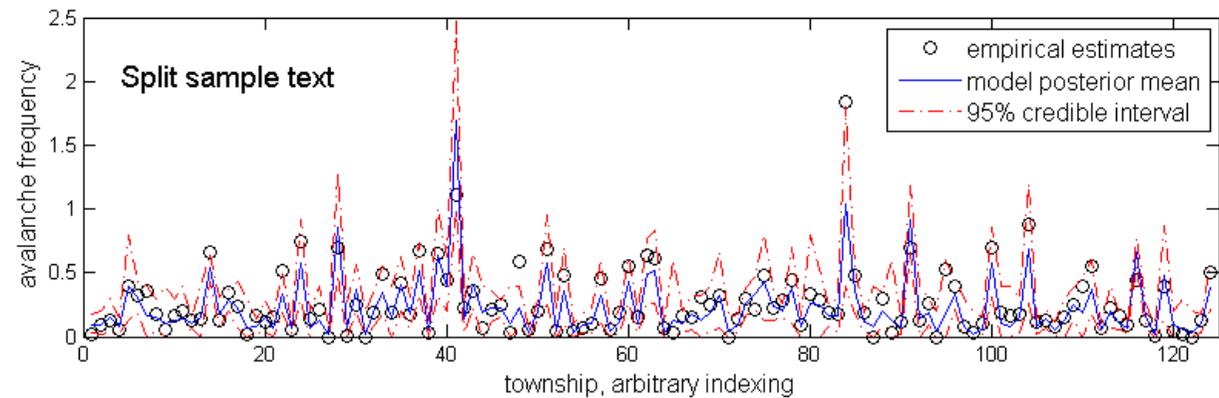
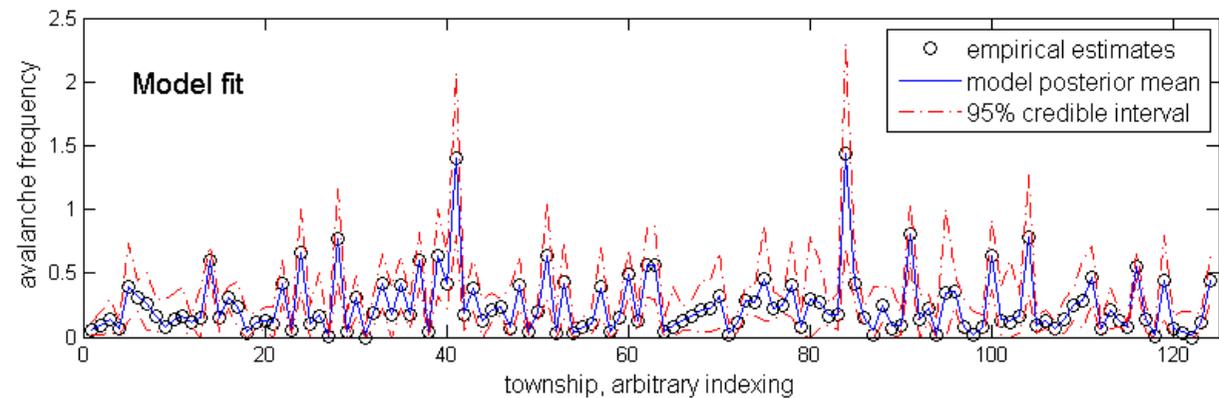
$$f_i^{pred} = \frac{y_i^{pred}}{c_i \times \left(\frac{t_{obs}}{T} \right)}$$



Bilan et limites

➔ **60% de variabilité expliquée par la spatialisation :**
Modélisation spatiale « pas idiote »

➔ **Adéquation raisonnable entre données et modèle :**



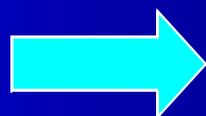
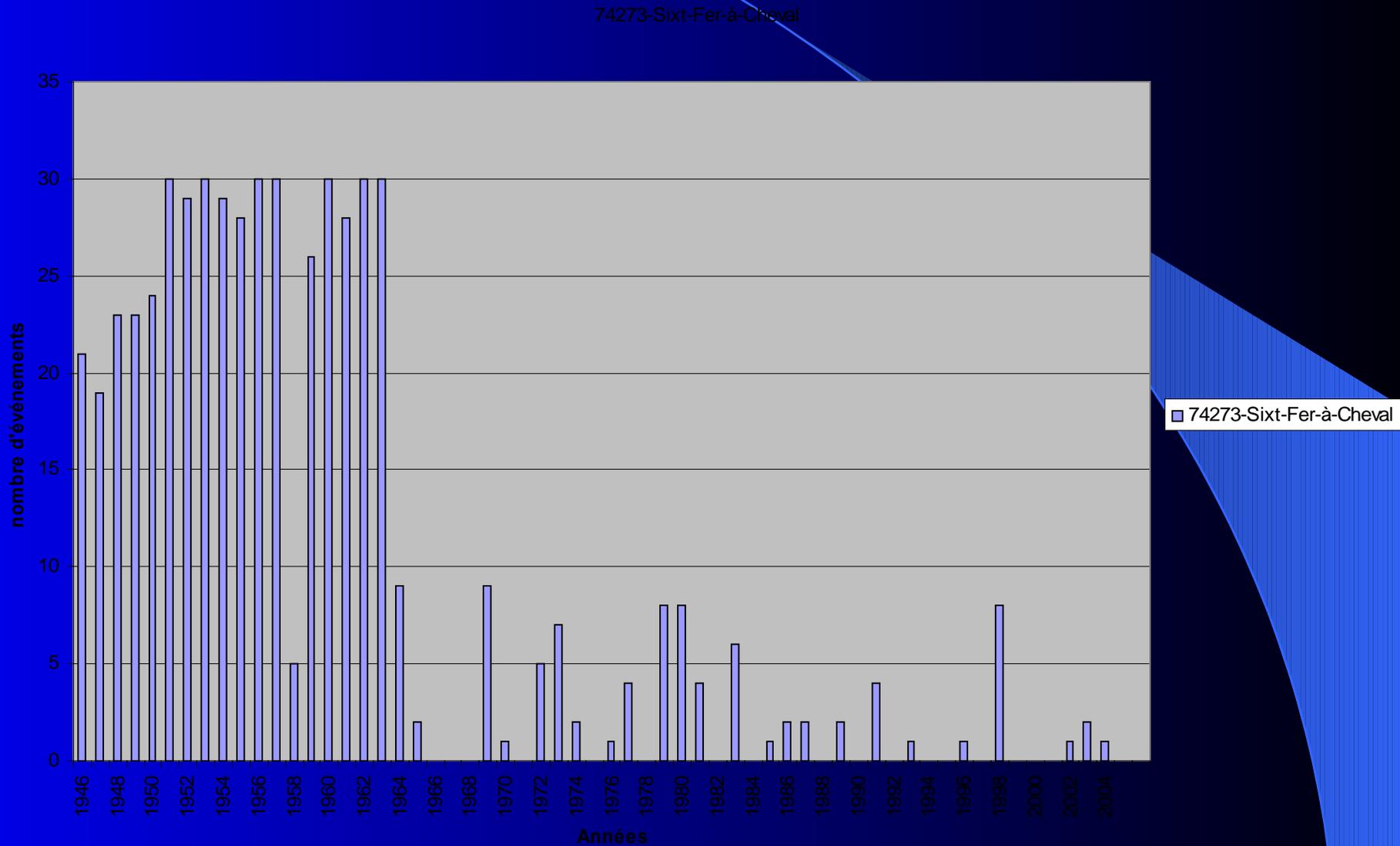
➔ **Mais nombreuses limites :**
Effort d'observation, pouvoir prédictif, hypothèses de modélisation.

Améliorations du modèle

- ➔ **Qualité des données** : pré-filtrage
- ➔ **Pouvoir prédictif** : Introduction de covariables et recalcul des nombres attendus
- ➔ **Stationnarité** : introduction d'un effet temporel
- ➔ **Hypothèses de modélisation** :
Test d'autres structures

Application à Savoie-Haute Savoie

Filtrage des données manquantes



Passage à des décomptes annuels

Introduction des covariables

Variables utilisées:

- altitude minimale
- altitude maximale
- altitude moyenne
- superficie
- précipitations sur 3 jours

Variables significatives:

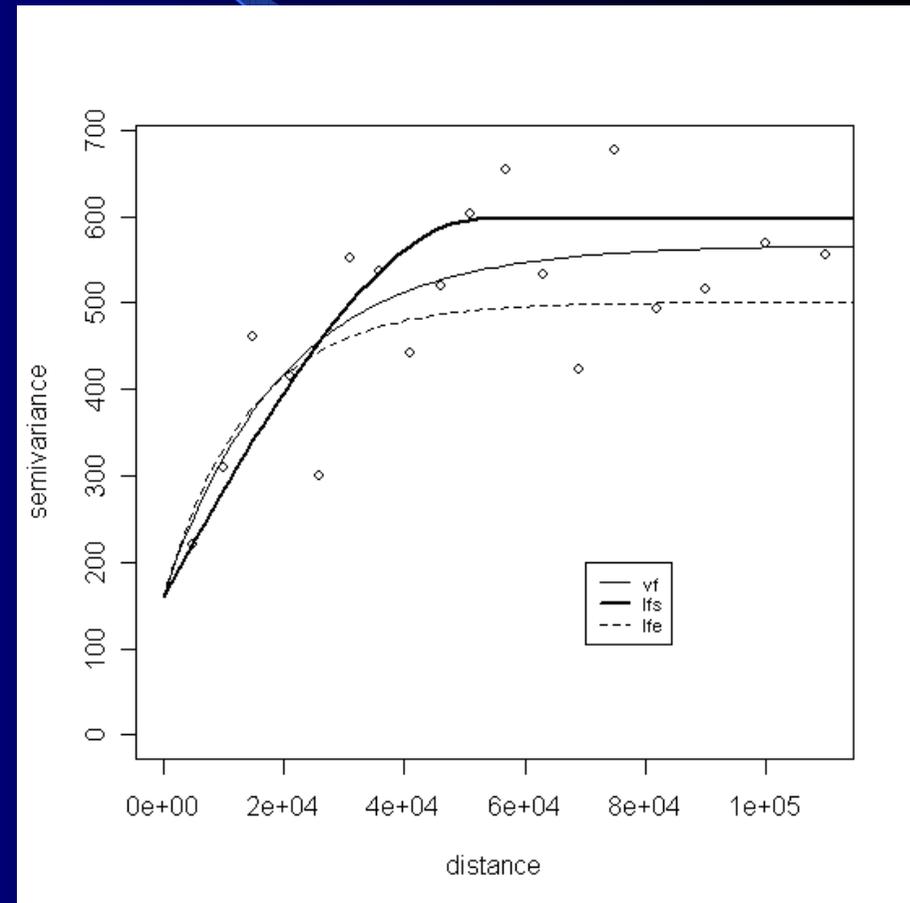
- altitude moyenne
- précipitations sur 3 jours

Intérêt: recalcul des nombres attendus

$$\ln(E'_i) = \ln(E_i) + \sum_{j=1}^M \alpha_j \times X_{ij}$$

Obtention du cumul de précipitation sur 3 jours à l'échelle communale

- Q_{10} de précipitation hivernale
- Altitude constante, gradient Nord/Sud
- Krigeage des résidus
- Prédiction au niveau des centroïdes des communes

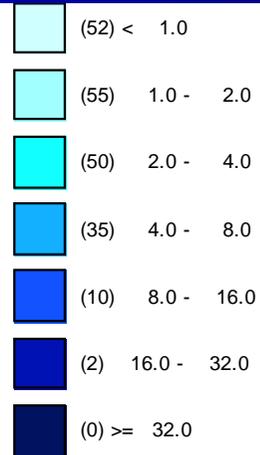
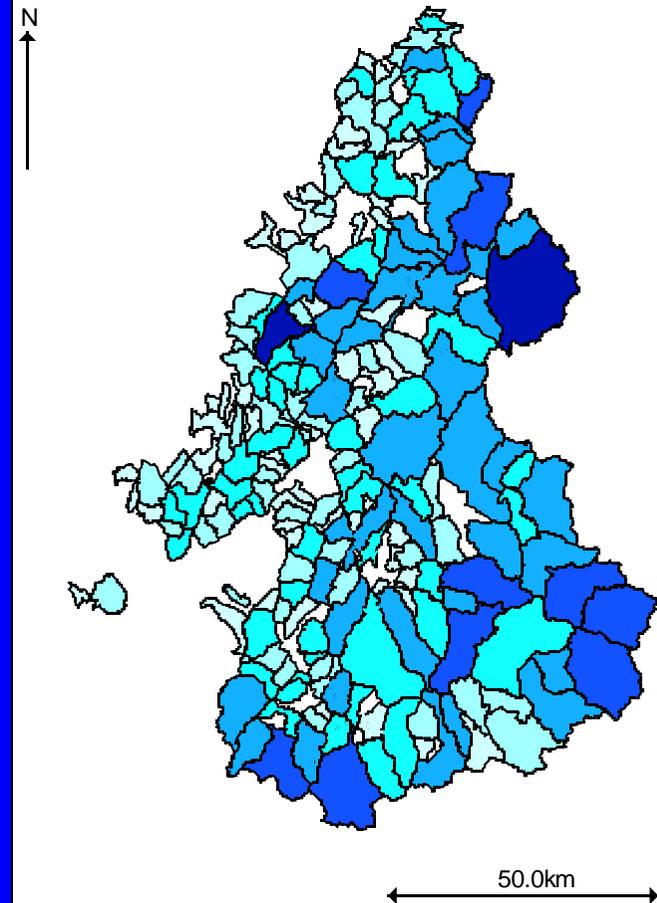


**Variogramme empirique
Ajustement avec 2 variogrammes
exponentiels et un variogramme sphérique**

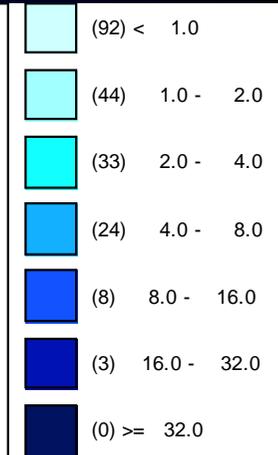
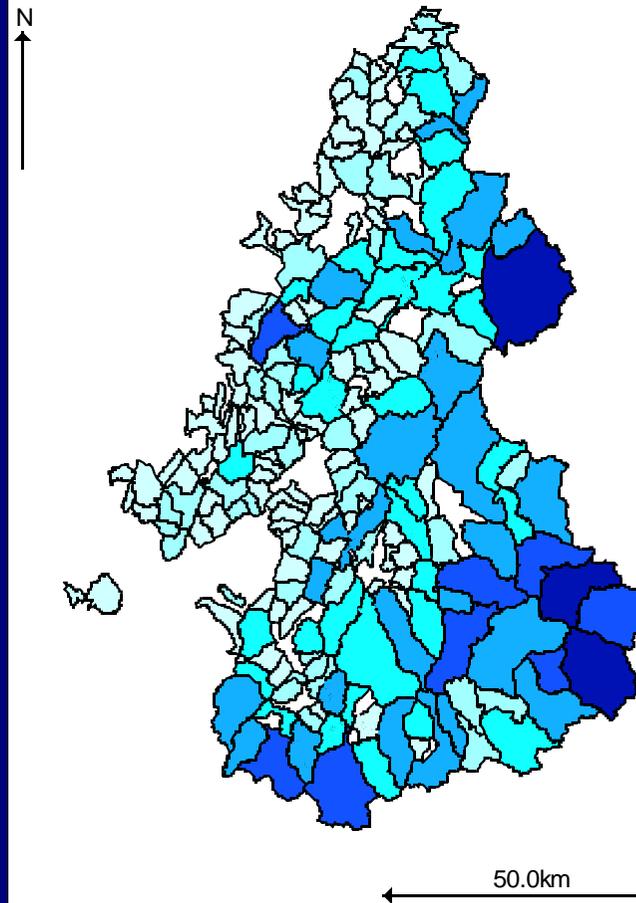
Recalcul des nombres attendus

 E_i
 E'_i

values for E

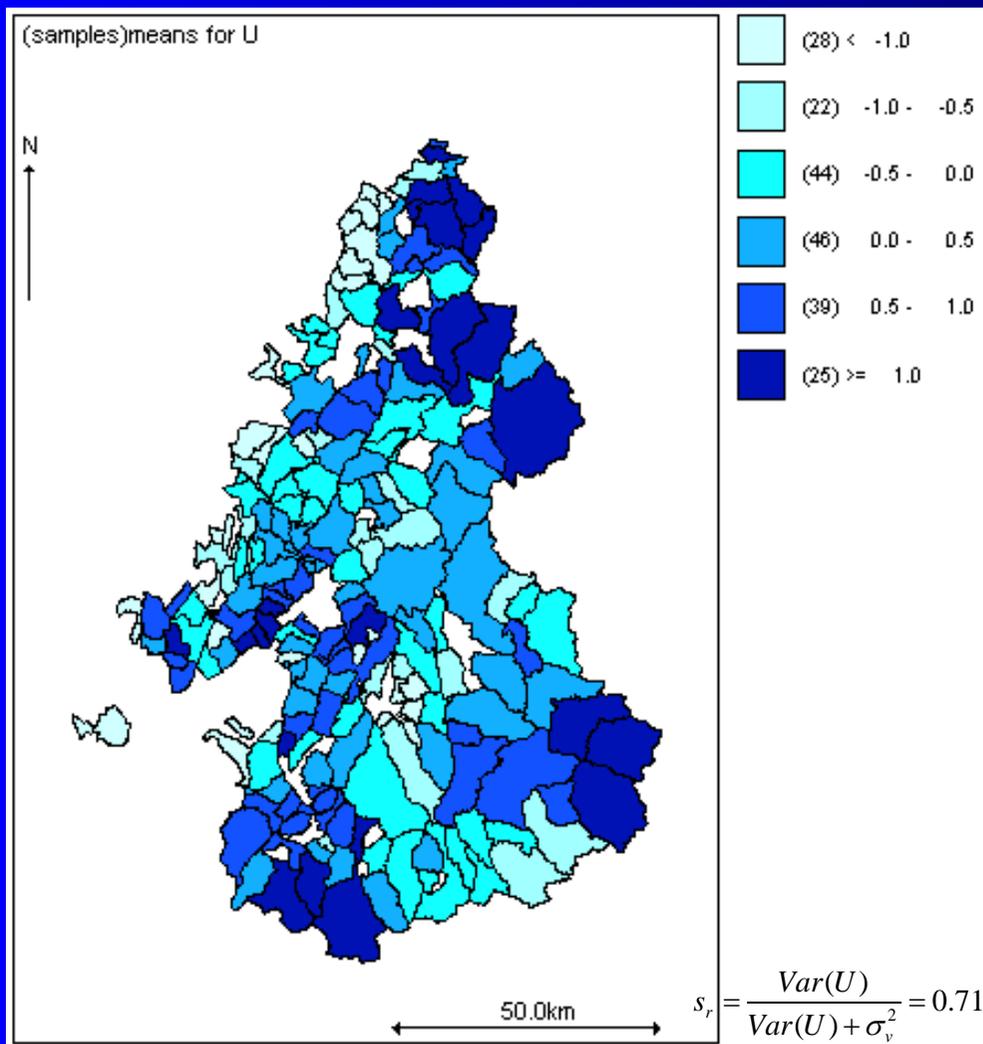


(samples)means for mref

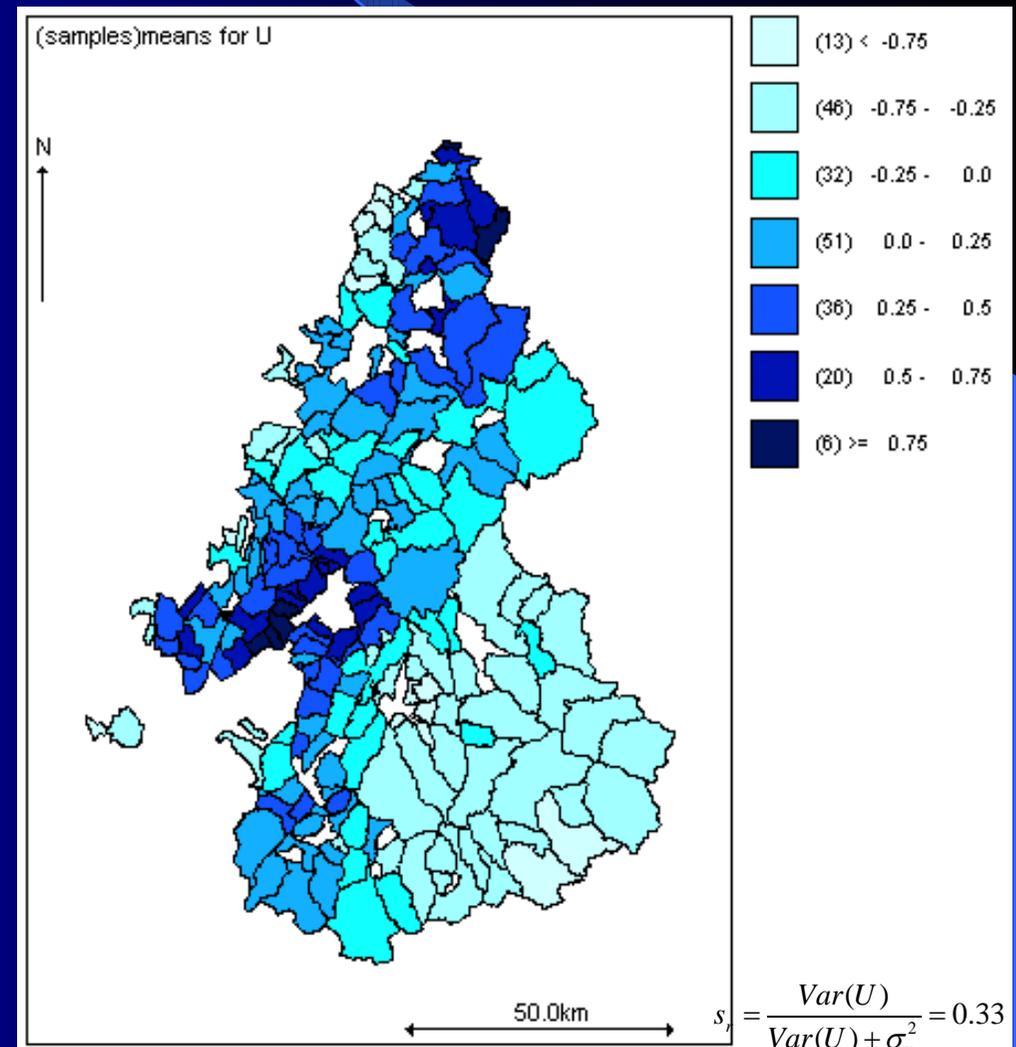


Recalcul des motifs spatiaux

U modèle de base (avec données manquantes)



U avec covariables (avec données manquantes)



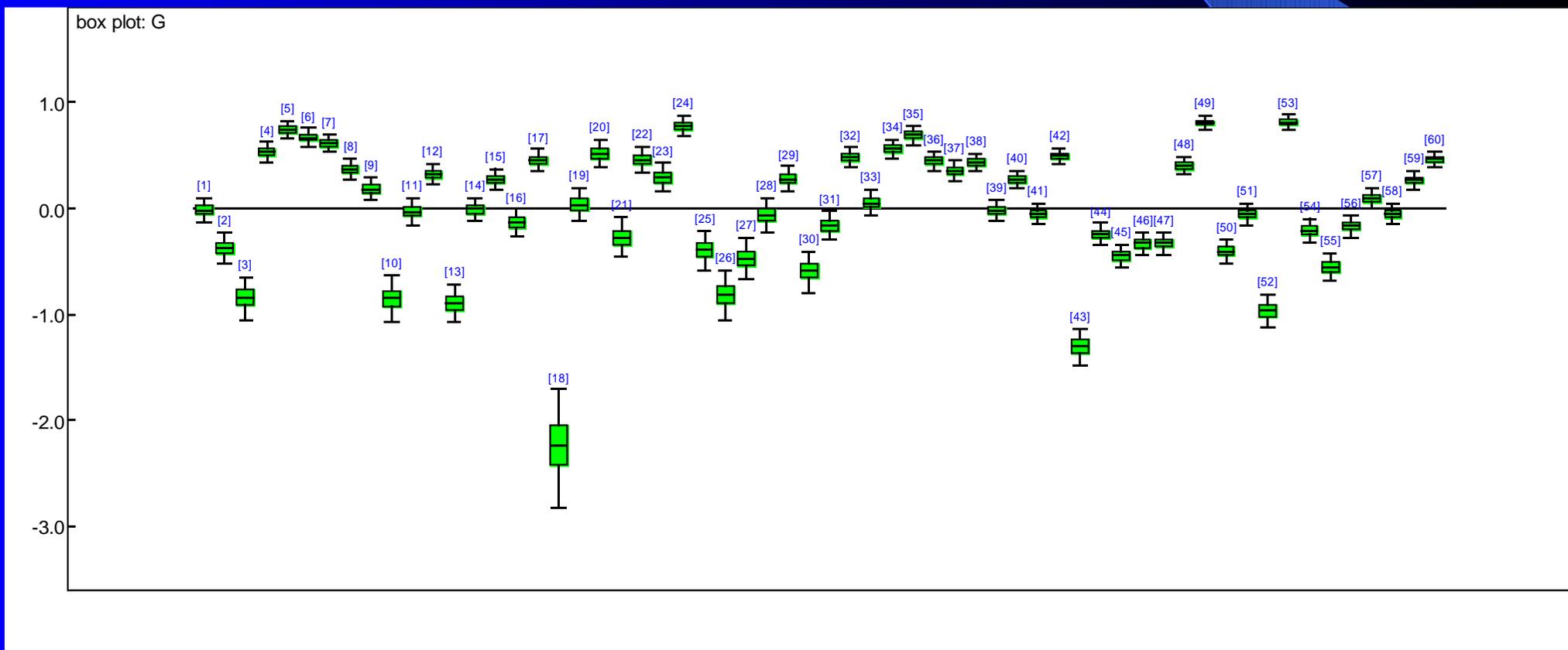
Introduction du temps : Extraction de la composante temporelle

G modèle de base

(avec données manquantes et covariables)

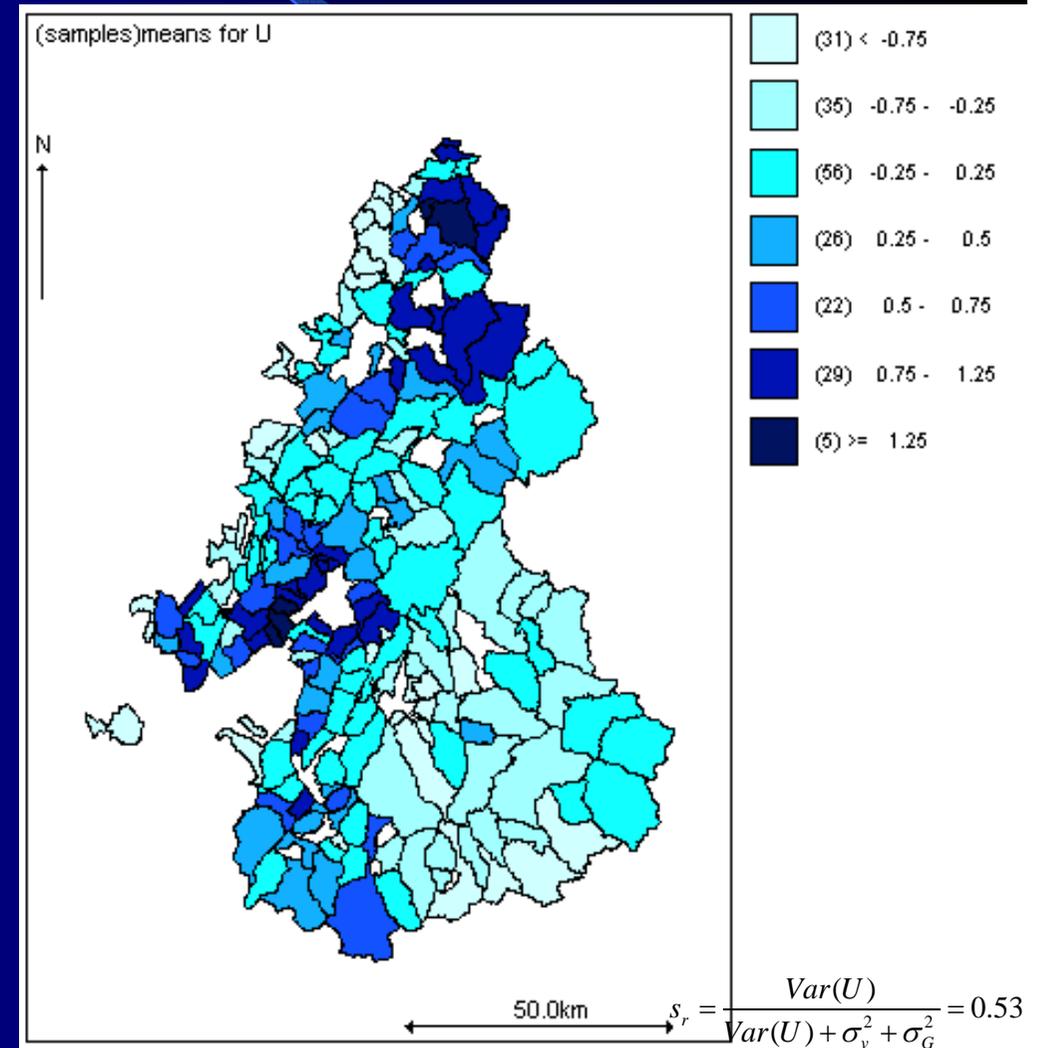
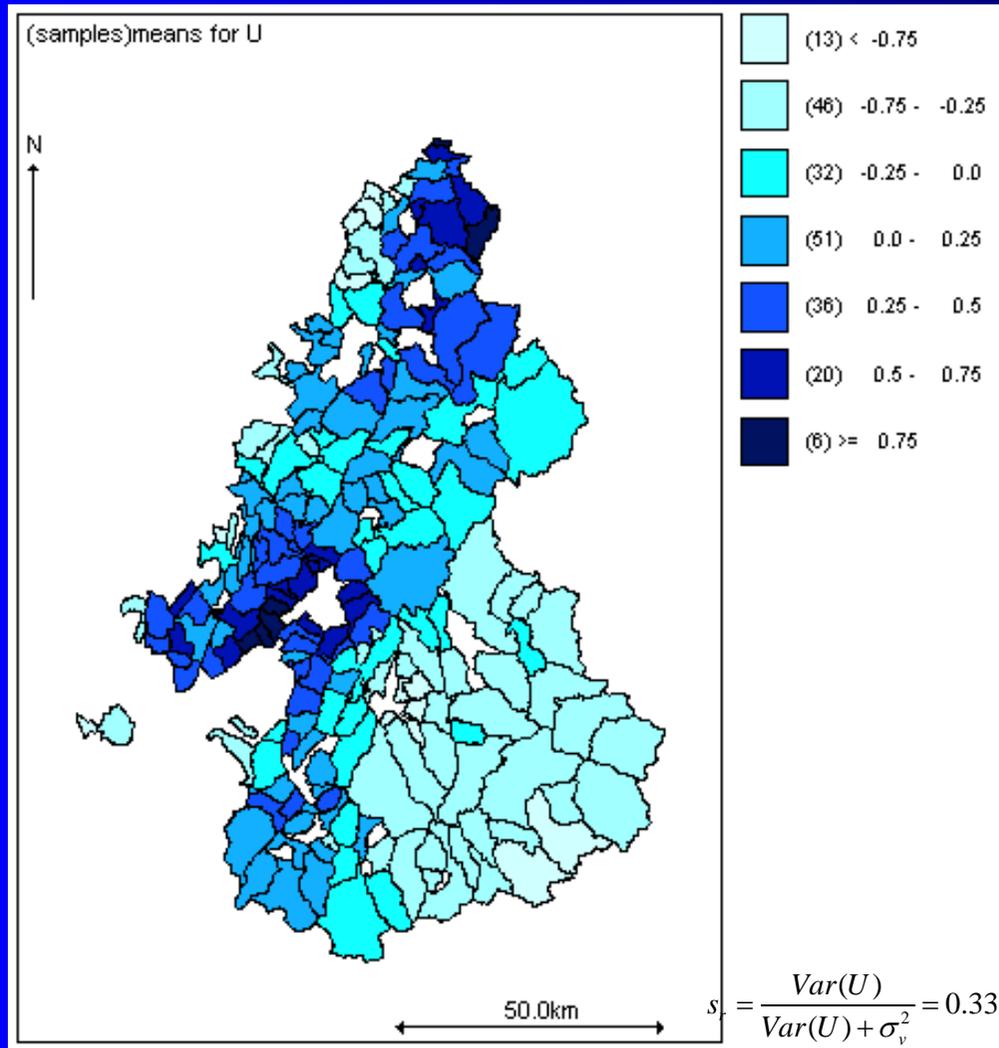
$$\ln(\lambda_{i,t}) = \ln(E_i) + v_i + u_i + G_t$$

$$G_t \sim N(0, \sigma_G)$$



Recalcul du terme spatial

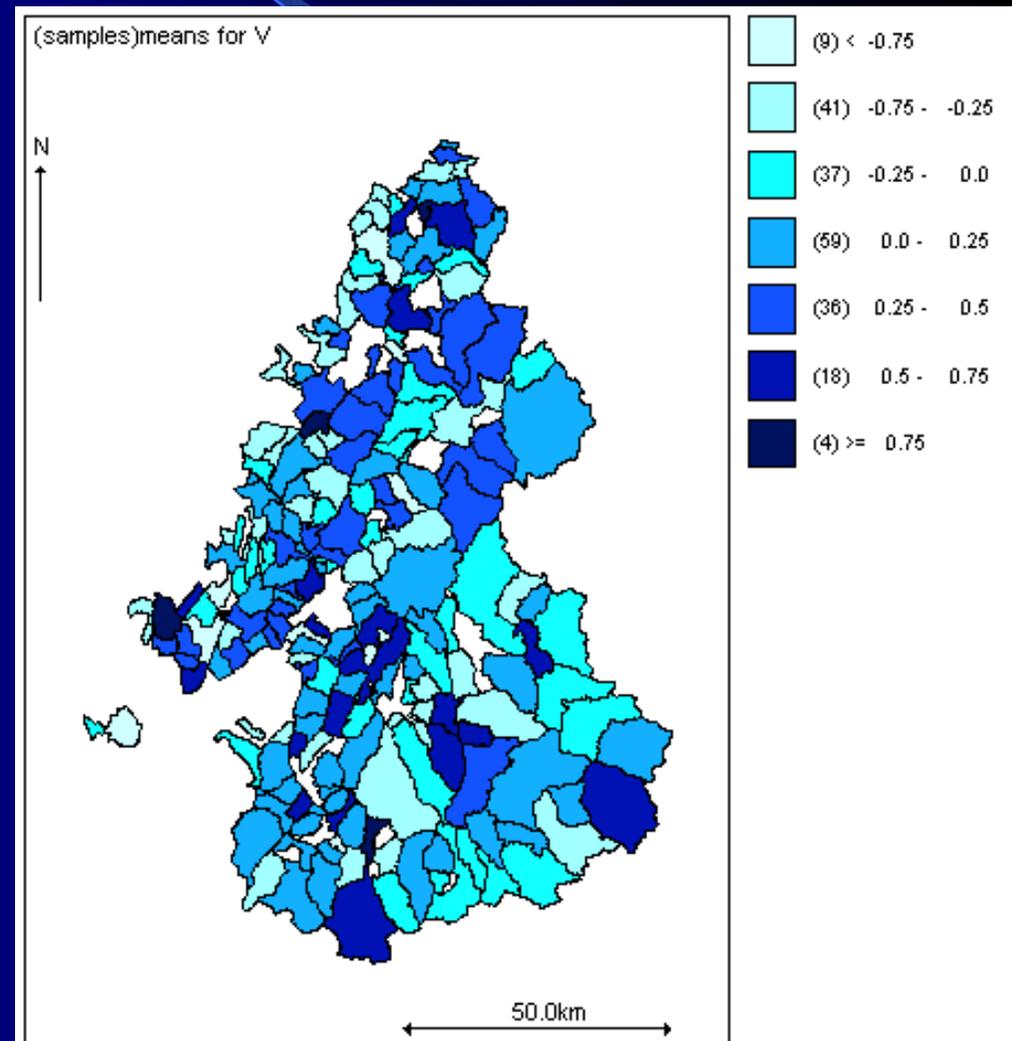
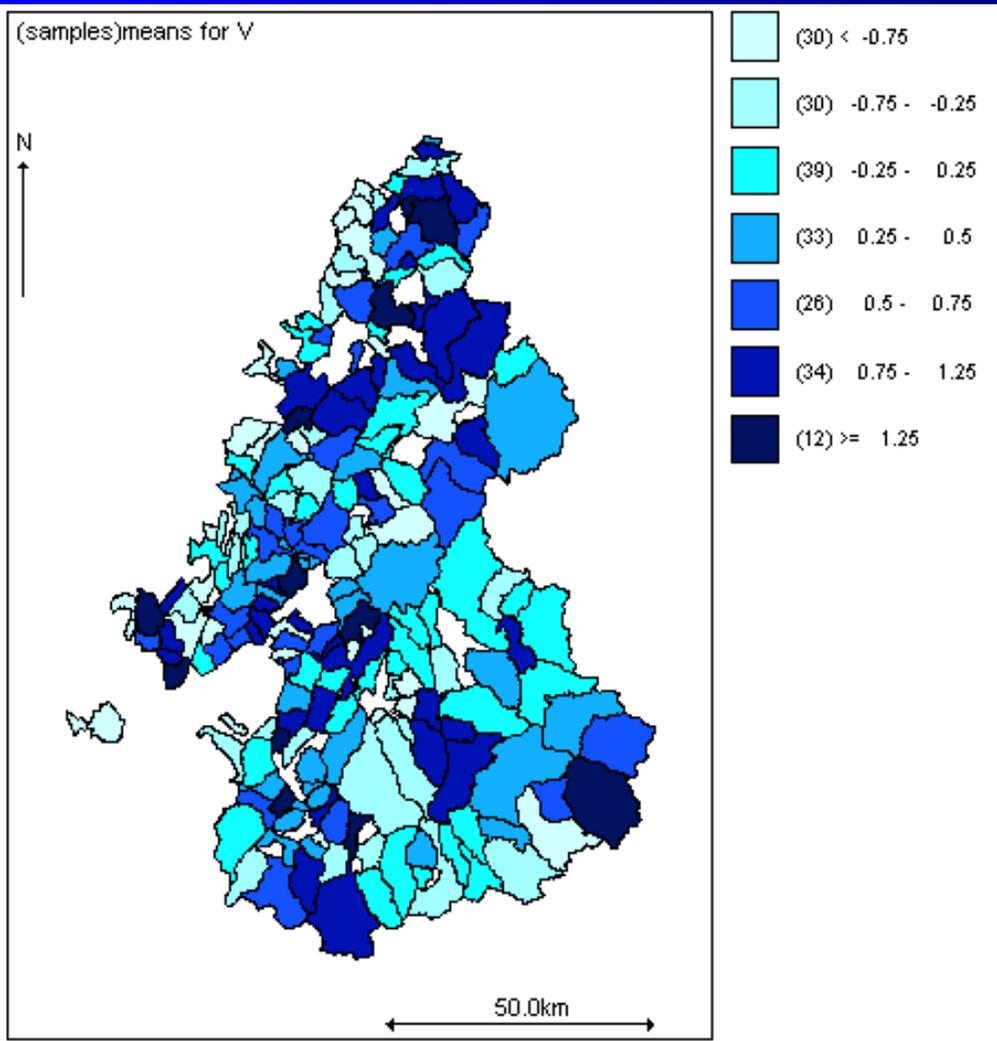
U avec modèle spatial **U avec modèle spatio-temp**
 (avec cov. et données manquantes) (avec cov. et données manquantes)



Recalcul du bruit

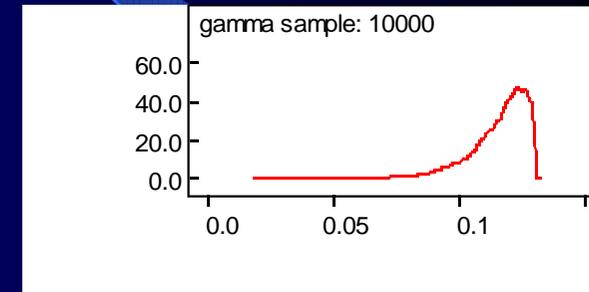
V avec modèle spatial
(avec cov. et données manquantes)

V avec modèle spatio-temporel
(avec cov. et données manquantes)



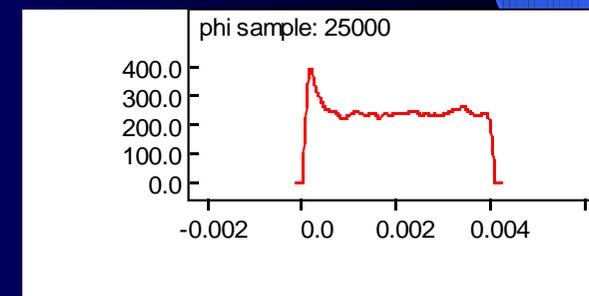
Recherche d'améliorations spatiales

CAR « propre » avec paramètre représentant le degré de dépendance spatiale



Paramètre de dépendance spatiale du « CAR propre »

Modèles de krigeage au niveau des centroïdes : exponentiel ou sphérique

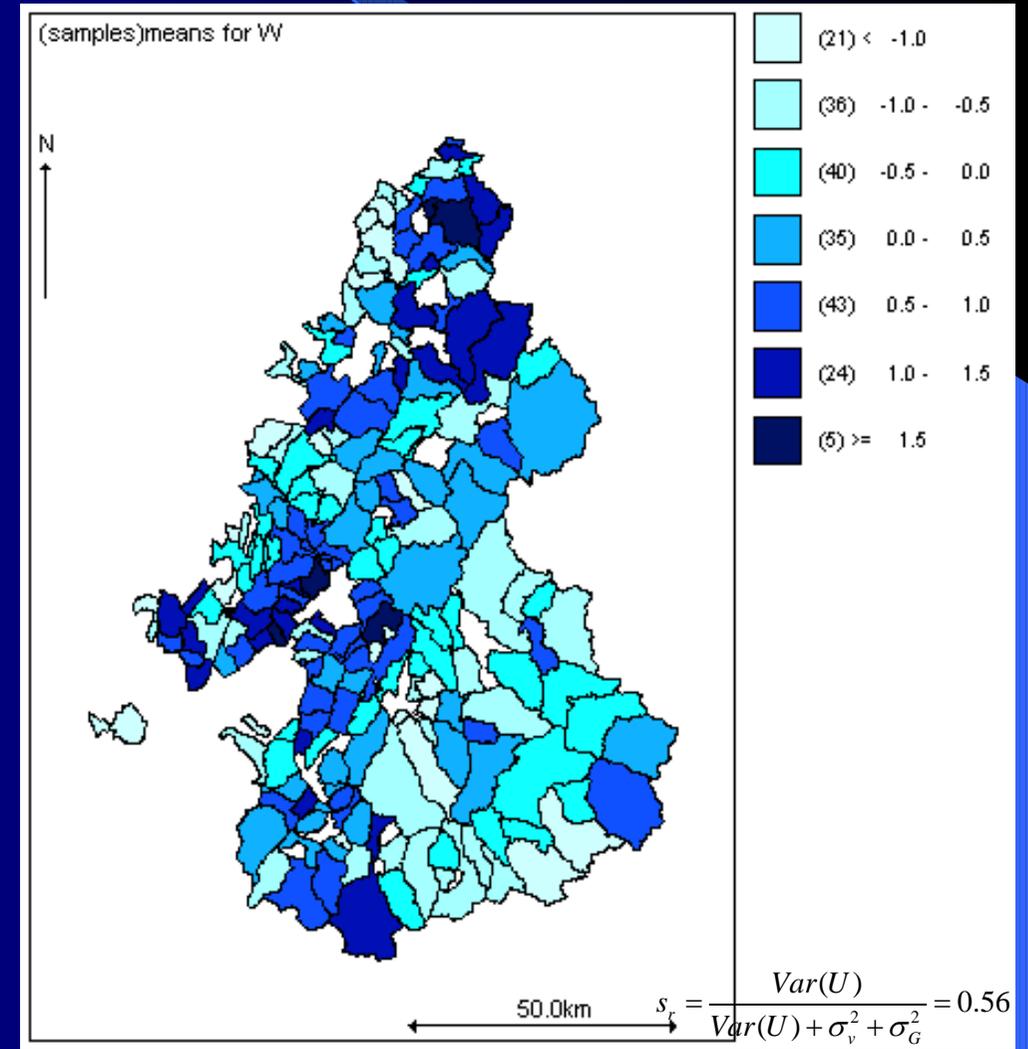
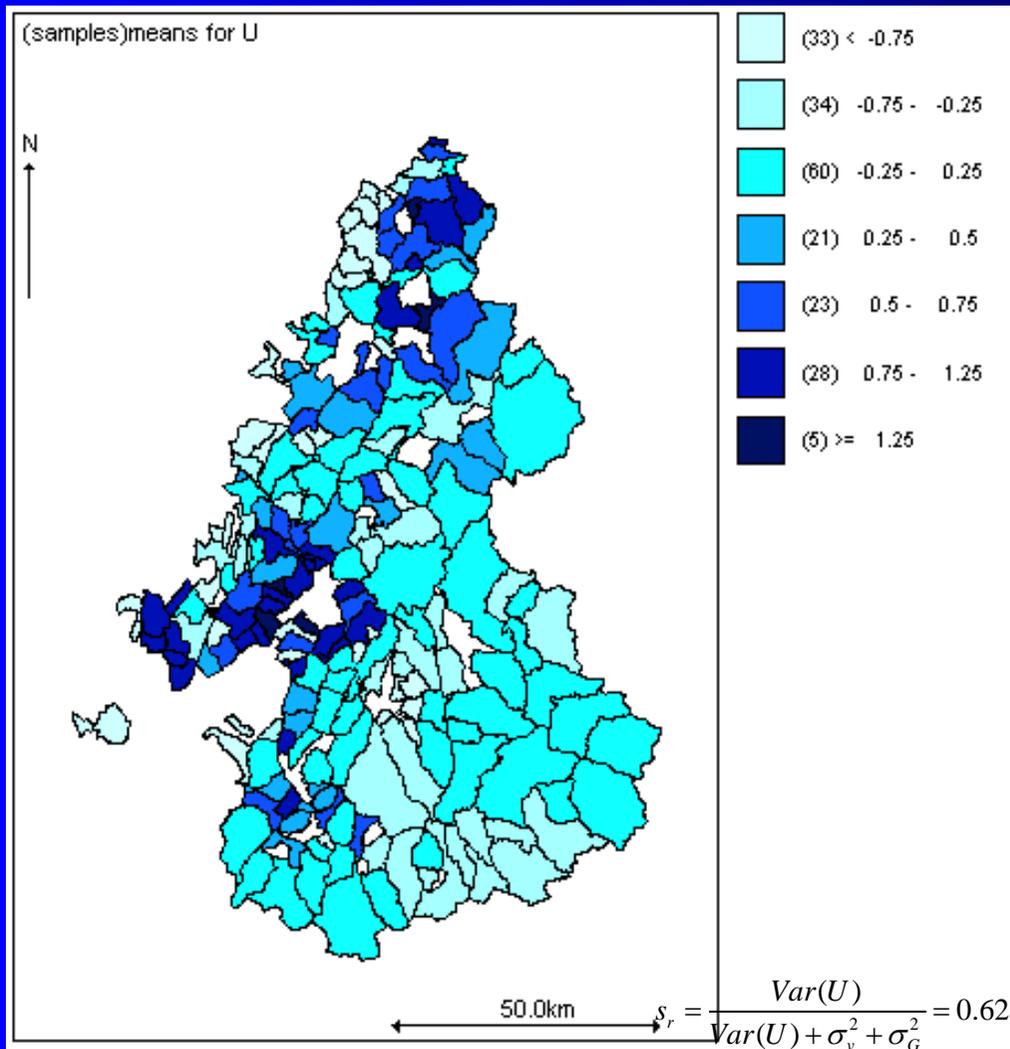


Paramètre de portée du variogramme exponentiel

Motifs spatiaux

U « CAR propre »

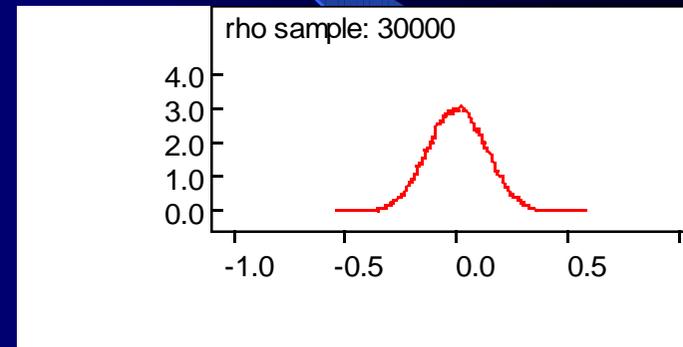
W=U+V : krigeage exponentiel + bruit



Recherche d'améliorations temporelles et spatio-temporelles

Modèle avec autorégression

$$G_t \sim N(\rho G_{t-1}, \sigma_G)$$



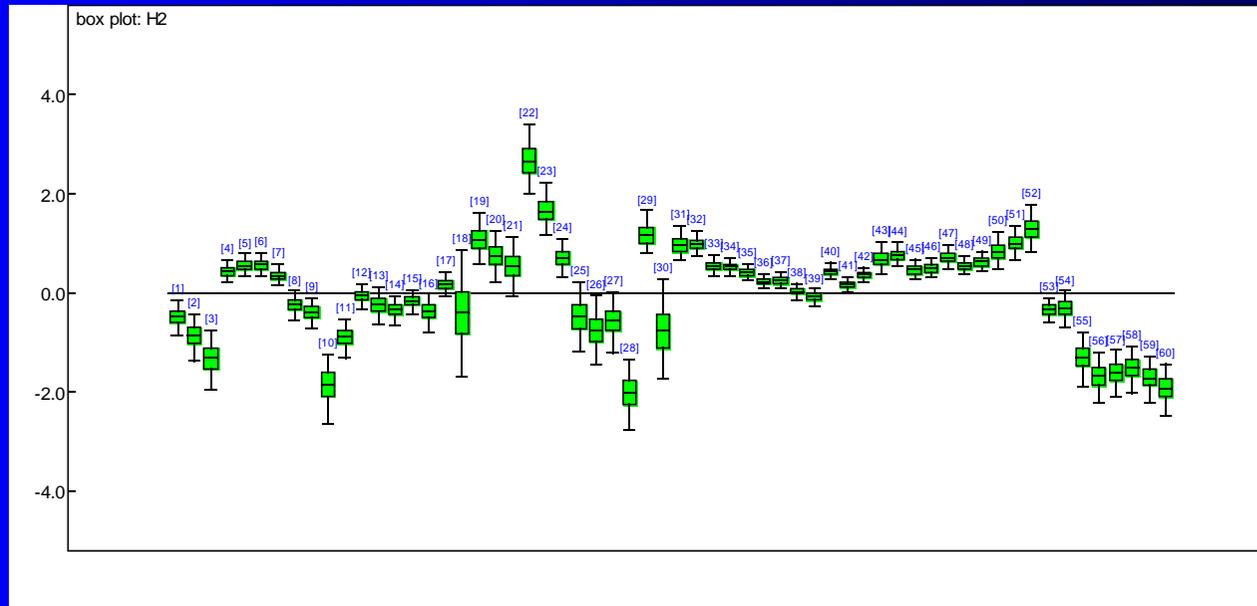
Modèle avec terme croisé

$$\ln(\lambda_{it}) = \ln(E_i') + v_i + u_i + G_t + H_{1i} \times H_{2t}$$

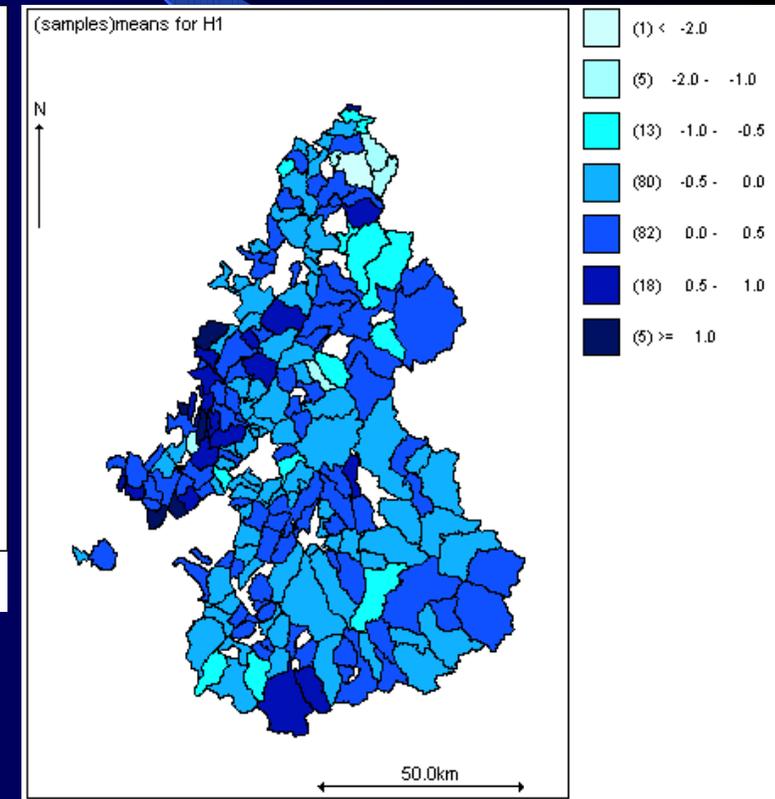
$$G_t \sim N(0, \sigma_G) \quad H_{1i} \sim N(0, \sigma_{H1}) \quad H_{2t} \sim N(0, 1)$$

Résultats terme croisé

Temporel « second ordre »



Spatial « second ordre »



Comparaison des modèles

	DIC
Modele	31568.600
Modele + temps	27121.800
Modele+temps+ CAR proper	27108.3
Modele+temps+ exponentiel	27454.200
Modele+temps+ spherique	27561.200
Modele+autorégression	27120.5
Modele+terme croisé	24510

- Apports majeurs : temps, terme croisé
- Apports mineurs : CAR propre
- Inefficace : autorégression sur t, modèles « géostatistiques »

Conclusions et perspectives

BILAN

- Séparation temps-espace :
« analyse de variance » en non Gaussien grâce à Winbugs
- Problème opérationnel :
Utilité mais travail avec des données très bruitées

PERSPECTIVES

- Interprétation des variations spatio-temporelles
- Passage de la commune au couloir : pb de taille de données et des zéros