

## EXISTE-T-IL UNE HETEROGENEITE SPATIALE DES CAS NAIF D'ESB EN FRANCE ?\*

David Abrial<sup>1</sup>, Nathalie Lauvergne<sup>2</sup>, Eric Morignat<sup>2</sup>,  
Didier Calavas<sup>2</sup> et Christian Ducrot<sup>1</sup>

**RESUME :** Cet article présente l'étude spatiale d'un projet de recherche sur les bovins NAIF (nés après l'interdiction des farines de viande et d'os dans l'alimentation des bovins) atteints d'ESB, qui cherche à déterminer pourquoi des centaines de bovins ont été contaminés en France après l'interdiction des farines animales en juillet 1990. Pour comprendre l'origine de cas NAIF, deux voies de contamination sont explorées: le rôle des farines de viande et d'os par l'intermédiaire de contaminations croisées entre aliments pour bovins et aliments pour monogastriques (utilisation des farines animales autorisée jusqu'en novembre 2000) et celui des produits d'abattoir autorisés pour la consommation animale (graisses et phosphates bicalciques).

L'hypothèse sous-jacente de l'étude spatiale des cas d'ESB NAIF est que l'existence d'agrégats spatiaux de cas pourrait être mise en relation avec les pratiques des fabricants d'aliments du bétail (matières premières utilisées et outil de production) dans les zones en question, en relation avec les hypothèses étudiées. La première étape de cette analyse, présentée dans l'article, est de rechercher l'existence d'une hétérogénéité spatiale des cas d'ESB. Pour cela, deux ensembles de données sont utilisées (car la surveillance de la maladie est optimale et comparable entre zones géographiques) ; d'une part, celles de la région du Grand-Ouest de la France (Basse-Normandie, Pays de la Loire et Bretagne) au cours du deuxième semestre 2000, avec 84 cas d'ESB recensés, et d'autre part la France entière, entre juillet 2001 et décembre 2002, avec 378 cas. Les informations sur les cas d'ESB sont fournies par l'Agence française de sécurité sanitaire des aliments qui regroupe les données des différents programmes de surveillance et de dépistage de l'ESB (réseau de surveillance clinique, tests rapides systématiques à l'abattoir, tests réalisés à l'équarrissage).

L'analyse statistique (présentée en détail) associe une méthode de cartographie des taux de contamination par l'ESB avec estimation bayésienne et deux méthodes de recherche des agrégats de cas (Scan de Kulldorff et méthode de Besag et Newell). L'analyse des données du Grand Ouest à l'échelle cantonale met en évidence une hétérogénéité spatiale des cas. En revanche, l'analyse des données de la France entière, réalisée à ce jour à l'échelle départementale, montre une hétérogénéité non statistiquement établie. Compte tenu des hypothèses de l'étude, à savoir l'existence d'une hétérogénéité spatiale qui correspondrait aux zones de chalandise des usines d'aliments du bétail, une taille d'unité géographique intermédiaire entre le canton et l'arrondissement pourrait être mieux adaptée. Par ailleurs, la non indépendance des zones géographiques voisines peut être prise en compte par un modèle descriptif plus élaboré. C'est dans ce sens que l'étude est poursuivie actuellement.

\* Communication présentée lors des Journées AEEMA-AESA, 22-23 mai 2003

<sup>1</sup> Unité d'épidémiologie animale, INRA Theix, 63122 Saint Genès Champanelle, France

<sup>2</sup> Unité épidémiologie, AFSSA Lyon, 31 avenue Tony Garnier, 69364 Lyon cedex 07, France

**SUMMARY :** *This paper presents a spatial study of BSE cases in France ; it is part of a project called "bovins NAIF" (animals born after the feed ban) that aims at determining the reasons why hundreds of cattle were contaminated in France after the Meat and Bone Meal feed ban for cattle on July 1990. The objective of the study is to analyze if it does exist spatial clusters of cases that could be linked to the activity areas cattle feed suppliers; the hypotheses of the whole project are the possible contamination of cattle, either with the MBM via cross-contamination between feed for monogastric and those for cattle, or the contamination of abattoir products derived from cattle (fat and phosphates). For that, two sets of data are available (optimal and comparable monitoring between geographical areas); the first concerns the Western France (Basse-Normandie, Pays de la Loire and Bretagne) during the second half of 2000, with 84 BSE cases, the second is the whole French territory, between July 2001 and December 2002, with 378 cases. BSE data were obtained from the "Agence Française de Sécurité Sanitaire des Aliments". The statistical analysis (presented in detail) associates a method of cartography of the standardised contamination rates with Bayesian estimate and two methods of clustering (Scan of Kulldorff and method of Besag and Newell). The analysis of the data of the Western France highlights a spatial heterogeneity and the analysis in the whole France, shows a non significant heterogeneity at a large scale. However the later has been carried out at the "arrondissement" level. A smaller geographical scale might be more appropriate due to the hypotheses under consideration. So another analysis is under way using a different geographical scale and taking into account the non independence of contiguous geographic areas.*



## I - INTRODUCTION - LE PROJET GENERAL

### 1. CONTEXTE

Cet article s'inscrit dans le cadre d'un projet de recherche sur le rôle de l'alimentation dans la contamination par l'agent de l'encéphalopathie spongiforme bovine (ESB) des bovins nés après l'interdiction des farines de viande et d'os (FVO) dans l'alimentation des bovins (cas dits NAIF). L'ESB est une maladie dégénérative dont l'agent infectieux est étroitement associé à la protéine prion transconformée, communément nommée PrPres pour protéine prion résistante à la protéinase K [Prusiner, 1982]. L'agent infectieux de l'ESB a été recyclé dans les farines de viande et d'os qui étaient auparavant utilisées en France dans l'alimentation des bovins.

Malgré l'interdiction des FVO dans l'alimentation des bovins depuis juillet 1990 en France

[<http://www.agriculture.gouv.fr/esbinfo/esbinfo.htm>], 740 cas (90% de l'ensemble des cas – chiffre au 31/05/03) ont été détectés sur des animaux nés après cette date. Pour analyser l'origine de ces cas, plusieurs hypothèses de contamination alimentaires des bovins sont explorées dans ce projet de recherche : rôle des FVO par le biais des contaminations croisées entre aliments pour monogastriques (pour lesquels les FVO étaient autorisées

jusqu'en novembre 2000) et aliments pour bovins, rôle des phosphates bicalciques précipités d'os et des graisses d'origine animale autorisés dans l'alimentation animale jusqu'en novembre 2000 [AFSSA, 2001]. Le projet de recherche est divisé en deux études et le présent article concerne l'une d'entre elles qui repose sur une analyse spatiale des cas NAIF d'ESB en France.

### 2. HYPOTHESES ET OBJECTIFS

L'étude « spatiale » a pour objectif final de mettre en évidence à l'échelle géographique une éventuelle relation entre le taux d'incidence des cas d'ESB et certaines caractéristiques des usines de fabrication d'aliment pour bovins en relation avec les hypothèses étudiées (utilisation de FVO, de phosphates bicalciques précipités d'os et de graisses animales d'une part, sécurisation des filières de fabrication d'autre part pour éviter les contaminations entre les fabrications pour monogastriques et celles pour bovins).

Cette analyse géographique repose sur l'hypothèse suivante. Si la contamination des bovins NAIF est liée aux hypothèses étudiées et si les pratiques des fabricants d'aliments du bétail ne sont pas homogènes vis-à-vis de ces hypothèses, le risque de contamination des

bovins est alors lié aux zones de livraison des fabricants. En d'autres termes, et c'est l'objet de la première phase de l'analyse, la répartition géographique des cas d'ESB ne devrait pas être aléatoire.

L'article présente les données utilisées, puis les différents types d'analyse retenus pour décrire l'hétérogénéité spatiale des cas d'ESB. Pour chaque analyse, les méthodes statistiques sont détaillées et les résultats présentés. Ensuite sont discutés les méthodes employées et les résultats obtenus.

## II – ORIGINE DES DONNEES

### 1. DONNEES EPIDEMIOLOGIQUES

Les données relatives aux cas d'ESB sont fournies par l'Agence française de sécurité sanitaire des aliments et la Direction générale de l'alimentation. Elles proviennent du réseau de surveillance clinique [Savey, 1991], du dépistage à l'abattoir par les tests rapides et du programme de tests en équarrissage [<http://www.agriculture.gouv.fr/esbinfo/esbinfo.htm>]. Afin de faire porter l'analyse sur des périodes et des zones géographiques pour lesquelles le dispositif de surveillance était efficace, exhaustif et homogène géographiquement [Calavas et Ducrot, 2003], deux ensembles d'observations ont été choisis.

Le premier concerne la région du Grand-Ouest pour la France (régions Basse-normandie, Pays de la Loire et Bretagne, soit 12 départements), la période allant de juillet à décembre 2000, et porte sur 84 cas d'ESB NAIF observés. Le second recouvre la France entière pour la période allant de juillet 2001 à décembre 2002, avec 378 cas NAIF d'ESB.

### 2. POPULATION BOVINE DE REFERENCE - CARTOGRAPHIE

Les données démographiques concernant la population de référence ou population à risque sont issues du Recensement agricole 2000 (CD-ROM édité par le Bureau des statistiques du ministère de l'agriculture et de la pêche). Les taux d'atteinte par l'ESB étant différents pour les vaches de race laitière (4,3 cas pour 100 000 vaches) et les vaches de race allaitante (0,6 cas pour 100 000 vaches), la distinction population bovine laitière/allaitante a été prise en compte dans nos analyses.

La population adulte de référence est de 2,6 millions de bovins dans le Grand-Ouest (1,8 millions de vaches laitières et 800 000 vaches allaitantes) et de 8,5 millions de bovins pour la France entière (4,2 millions de vaches laitières et 4,3 millions de vaches allaitantes).

Les cartes géographiques ont été construites sur la base des découpages administratifs que sont les 'cantons' et les 'départements' à partir du CD-ROM Geofla® 'France métropolitaine' (IGN® Paris, version 6, 2002). Le logiciel de représentation cartographique est ArcView®.

## III - METHODES ET PRINCIPAUX RESULTATS

### 1. CARTOGRAPHIE DES TAUX STANDARDISES

La cartographie des taux standardisés permet de mettre en évidence, au sein d'une population d'unités géographiques (UG), des différences de risque d'atteinte des bovins par l'ESB. C'est un ensemble de méthodes descriptives permettant de comparer graphiquement les UG d'un territoire étudié. Ces unités sont les individus statistiques et sont formées par un maillage régulier ou irrégulier. Le choix de ce maillage est d'une grande importance pour la précision des résultats et l'interprétation de l'analyse.

Un modèle probabiliste est appliqué à chacune des UG ; il indique quelle loi de distribution de probabilité suit le nombre de cas observés. On note  $n_{ug}$  le nombre d'UG composant le territoire étudié et  $y_i$  le nombre de cas d'ESB observés dans l'UG  $i$ . L'ESB étant une maladie rare, la loi de distribution de Poisson est la plus adéquate [Mollié, 1999]. Le modèle est de la forme :

$$y_i = \text{Poisson}(\lambda_i) \quad (1)$$

Le paramètre  $\lambda_i$  de la loi de Poisson est à la fois l'espérance mathématique et la variance du nombre de cas observés  $y_i$ .

Pour mettre en évidence des différences entre les UG, ces dernières sont comparées à une référence : une distribution homogène des taux de contamination, c'est-à-dire le niveau de contamination moyen de l'ensemble des UG. Ainsi, les nombres de cas observés sont comparés aux nombres de cas attendus si, pour un animal, la probabilité d'être contaminé est la même dans toutes les UG (hypothèse  $H_0$ ). Comme le paramètre  $\lambda_i$  est l'espérance mathématique du nombre de cas observés, si on note  $e_i$  le nombre de cas attendus sous  $H_0$ , il est possible d'écrire :

$$E(y_i) = \lambda_i = e_i \cdot r_i \text{ avec } r_i = 1 \text{ sous } H_0 \quad (2a)$$

Sous l'hypothèse alternative ( $H_1$ ), le nombre de cas observés est significativement différent du nombre de cas attendus :

$$E(y_i) = \lambda_i = e_i \cdot r_i \text{ avec } r_i \neq 1 \text{ sous } H_1 \quad (2b)$$

Le paramètre  $r_i$  est appelé Risque relatif de contamination (RRC) ; c'est une mesure de l'écart entre le nombre observé de cas et le nombre attendu sous  $H_0$  pour chacune des UG. A ce stade, les *nug* paramètres  $r_i$  doivent être estimés et plusieurs méthodes sont disponibles. Nous présentons dans les chapitres suivants la méthode du maximum de vraisemblance et la méthode empirique Bayésienne.

### 1.1. METHODE DU MAXIMUM DE VRAISEMBLANCE

Avec un modèle de Poisson de paramètre  $e_i \cdot r_i$ , la fonction de vraisemblance est donnée par,

$$L(r_i) = e^{-e_i \cdot r_i} \cdot \frac{(e_i \cdot r_i)^{y_i}}{y_i!} \quad (3)$$

Dans cette écriture,  $y_i!$  est la factorielle du nombre de cas observés. Pour obtenir l'estimateur du maximum de vraisemblance du RRC, il faut annuler la dérivée de  $L$  par rapport à  $r_i$ , soit,

$$\begin{aligned} \frac{dL(r_i)}{dr_i} = 0 &\Rightarrow -e_i \cdot e^{-e_i \cdot r_i} \cdot \frac{(e_i \cdot r_i)^{y_i}}{y_i!} + e^{-e_i \cdot r_i} \cdot (e_i \cdot r_i)^{y_i} \cdot \frac{-y_i}{r_i \cdot y_i!} = 0 \\ &\Rightarrow \hat{r}_i = \frac{y_i}{e_i} \end{aligned} \quad (4)$$

Les nombres de cas attendus  $e_i$  sont obtenus en appliquant une probabilité d'être contaminé par l'ESB égale pour toutes les UG aux effectifs bovins de ces UG. Il faut noter que cet estimateur n'a de sens que pour les UG avec un effectif bovin non nul ( $e_i > 0$ ). Enfin, l'écart type de cet estimateur est donné par,

$$\sigma_{r_i} = \frac{\sqrt{y_i}}{e_i} \quad (5)$$

Cette méthode du maximum de vraisemblance a été appliquée aux données du Grand-Ouest [Abrial *et al.*, 2003], étude en amont de l'analyse pour la France entière (la carte de risque n'est pas présentée ici).

Une telle carte est facile à lire puisque elle indique par un niveau de grisé l'importance du risque dans chaque UG par rapport au risque attendu sous l'hypothèse d'homogénéité du taux de contamination. Dans le cas de la région Grand Ouest, le risque observé est de 0,001 à 22,1 fois plus important que le risque attendu.

Cependant, l'interprétation épidémiologique de cette carte n'est pas évidente car la précision de l'estimateur de  $r_i$  n'est pas la même dans tous les cantons. Au niveau de chaque UG, cette précision dépend de l'effectif de bovins à risque. Comme la démographie est hétérogène entre les cantons, la précision du RRC est également hétérogène. Ainsi, la comparaison du RRC entre les UG doit être formulée avec prudence car le risque peut être sur ou sous évalué, on parle dans ce cas de surdispersion du risque [Clayton *et al.*, 1987]. Pour pallier ce problème de surdispersion, l'estimation du RRC doit être lissée par diverses méthodes ; l'une d'entre-elles est la méthode Bayésienne empirique.

### 1.2. METHODE BAYESIENNE EMPIRIQUE

Pour limiter les effets de la surdispersion du RRC, une méthode d'estimation bayésienne peut être utilisée. La méthode empirique fournit un estimateur du RRC lissé globalement par l'ajout de termes additifs, communs à toutes les UG, au numérateur et au dénominateur du rapport (4), soit,

$$r_i = \frac{y_i + \nu}{e_i + \alpha} \quad \text{et} \quad \sigma_{r_i} = \frac{\sqrt{y_i + \nu}}{e_i + \alpha} \quad (6)$$

Les paramètres  $\alpha$  et  $\nu$  proviennent d'une information *a priori* (aspect bayésien) sur la distribution de probabilité de  $r_i$  à travers l'ensemble des UG. La loi *a priori* couramment

utilisée en cartographie des taux standardisés est la loi gamma de moyenne  $\mu=v/\alpha$  et de variance  $\tau^2=v/\alpha^2$ , soit,

$$g(r_i|\alpha, v) = \frac{\alpha^v \cdot r_i^{v-1} \cdot e^{-\alpha r_i}}{\Gamma(v)} \quad (7)$$

L'avantage d'un tel choix est que la loi gamma est la distribution conjuguée de la loi de Poisson (1) [Robert, 1992]. Ainsi, il est possible d'en déduire la loi de distribution conditionnelle de  $y_i$  sachant  $\alpha$  et  $v$ . C'est une loi binomiale négative dont la fonction de vraisemblance est donnée par,

$$L(y_i, e_i, \mu, \tau^2) = \sum_{i=1}^{nug} \ln \left[ \frac{\Gamma\left(y_i + \frac{\mu^2}{\tau^2}\right)}{\Gamma\left(\frac{\mu^2}{\tau^2}\right)} \right] + \frac{\mu^2}{\tau^2} \ln\left(\frac{\mu}{\tau^2}\right) - \left(y_i + \frac{\mu^2}{\tau^2}\right) \ln\left(e_i + \frac{\mu}{\tau^2}\right) \quad (8)$$

Cette fonction permet d'obtenir les estimations de  $\mu$  et  $\tau^2$  donc de  $\alpha$  et  $v$ . La méthode la plus simple est de fixer la moyenne  $\mu$  à 1 [Mollié et Richardson, 1991] et de faire varier la valeur de  $\tau^2$  pour déterminer le maximum de (8). Ainsi, on obtient les estimateurs empiriques de  $\alpha$  et  $v$ ,

$$L(y_i, e_i, \hat{\mu}=1, \hat{\tau}^2) = \max L(y_i, e_i, \mu=1, \tau^2) \quad (9)$$

$$\hat{\alpha} = \frac{\hat{\mu}}{\hat{\tau}^2} \text{ et } \hat{v} = \frac{\hat{\mu}^2}{\hat{\tau}^2}$$

Une carte du risque de contamination par l'ESB du Grand-Ouest (Figure 1) a été obtenue par cette méthode.

Le RRC varie de 0,80 à 2,18. L'étendue de ce RRC est évidemment plus étroite que celle obtenue par la méthode du maximum de vraisemblance (0,001 à 22,1). Ainsi, la surdispersion est atténuée et la comparabilité des UG est plus crédible car la précision de l'estimateur (6) est plus homogène.

Il est possible de déterminer pour chaque UG de la carte si le RRC estimé est significativement supérieur ou inférieur au risque de contamination attendu sous l'hypothèse d'homogénéité. En d'autres termes, tester si  $r_i$  est différent de la valeur 1.

$$H_0 : r_i = 1$$

Pour tester cette hypothèse au seuil 5%, il suffit de déterminer pour chaque valeur  $r_i$ , un intervalle de confiance à partir d'une distribution gamma de moyenne  $\hat{r}_i$  et de variance  $\sigma_{\hat{r}_i}^2$  (voir équation 6). Si cet intervalle ne comprend pas la valeur 1, le RRC est significativement différent de 1. Sur la carte du Grand-Ouest (figure 1), un seul canton (en noir), dans le département de la Sarthe (n°72), présente un RRC significatif.

Avec la même méthode, nous proposons une carte du RRC par l'ESB pour la France avec comme UG les 96 départements pour la période juillet 2001 à décembre 2002 (figure 2).

### 1.3. MISE EN EVIDENCE D'UNE HETEROGENEITE GEOGRAPHIQUE GLOBALE

La méthode des taux standardisés permet de visualiser des différences de RRC entre les UG. Savoir si ces différences relèvent d'une hétérogénéité statistiquement significative nécessite un test que le modèle probabiliste ne fournit pas. Martuzzi et Hills [1995] proposent un test de rapport de vraisemblance pour tester l'hypothèse,

$$H_0 : \tau^2 = 0 \text{ « homogénéité spatiale du RRC »}$$

Sous  $H_0$ , le logarithme de la vraisemblance est,

$$L_0 = \sum_{i=1}^{nug} y_i \cdot \ln(\hat{\mu}) - \hat{\mu} \cdot \sum_{i=1}^{nug} e_i \quad (10)$$

et sous l'hypothèse alternative, le logarithme de la vraisemblance  $L_1$  est donné par le maximum de la fonction (8) – calculée en remplaçant  $\mu$  et  $\tau^2$  par leurs estimations  $\hat{\mu}$  et  $\hat{\tau}^2$ . Le rapport de vraisemblance  $2(L_1 - L_0)$  suit une loi de chi-deux à un degré de liberté.

Figure 1

Carte des taux standardisés de contamination par l'ESB dans les 526 cantons du Grand-Ouest de la France pour la période août à décembre 2000 : 84 cas d'ESB pour une population de 2,6 millions de bovins. Standardisation sur la distinction races laitières / races allaitantes, méthode empirique bayésienne, lissage global.

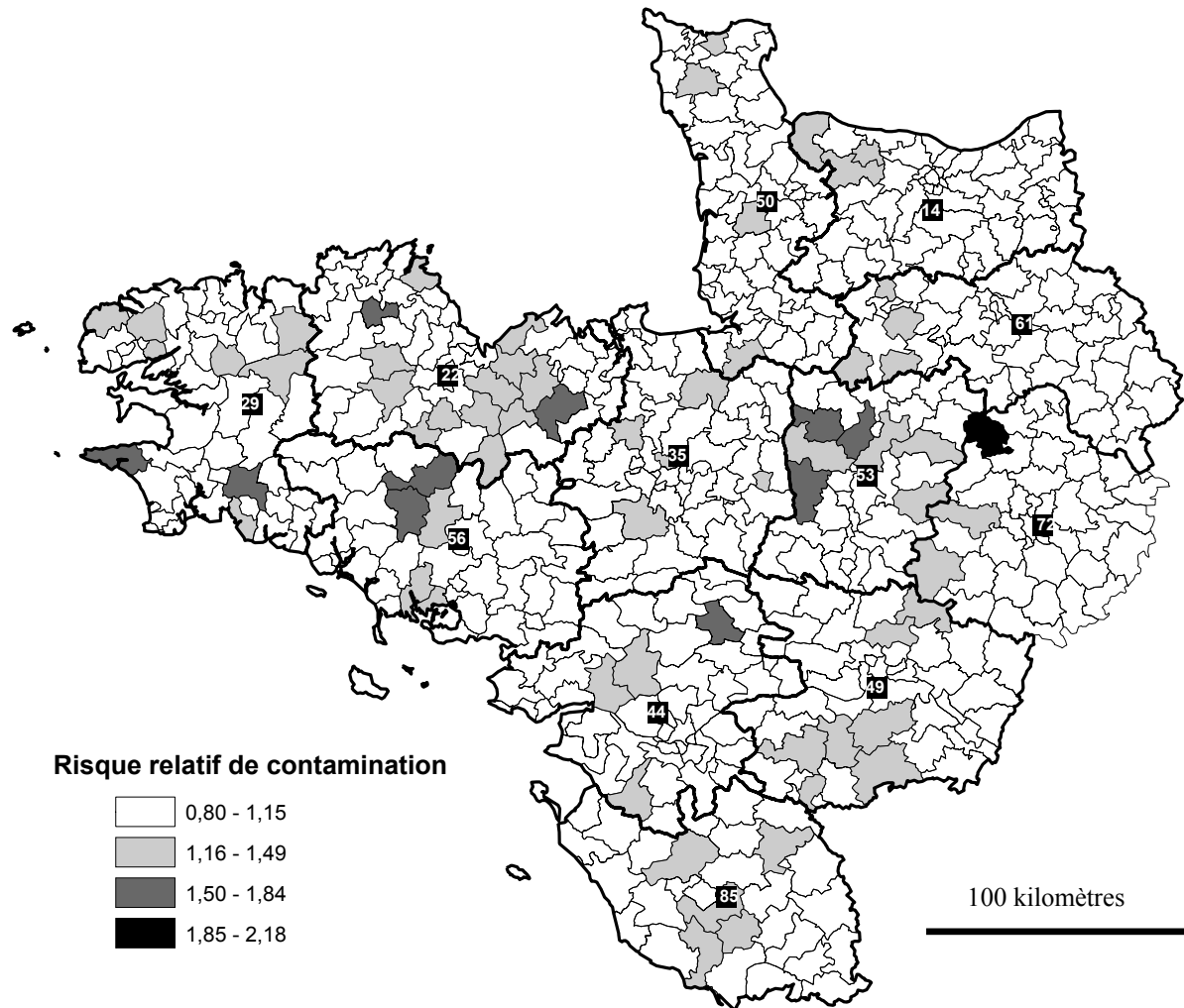
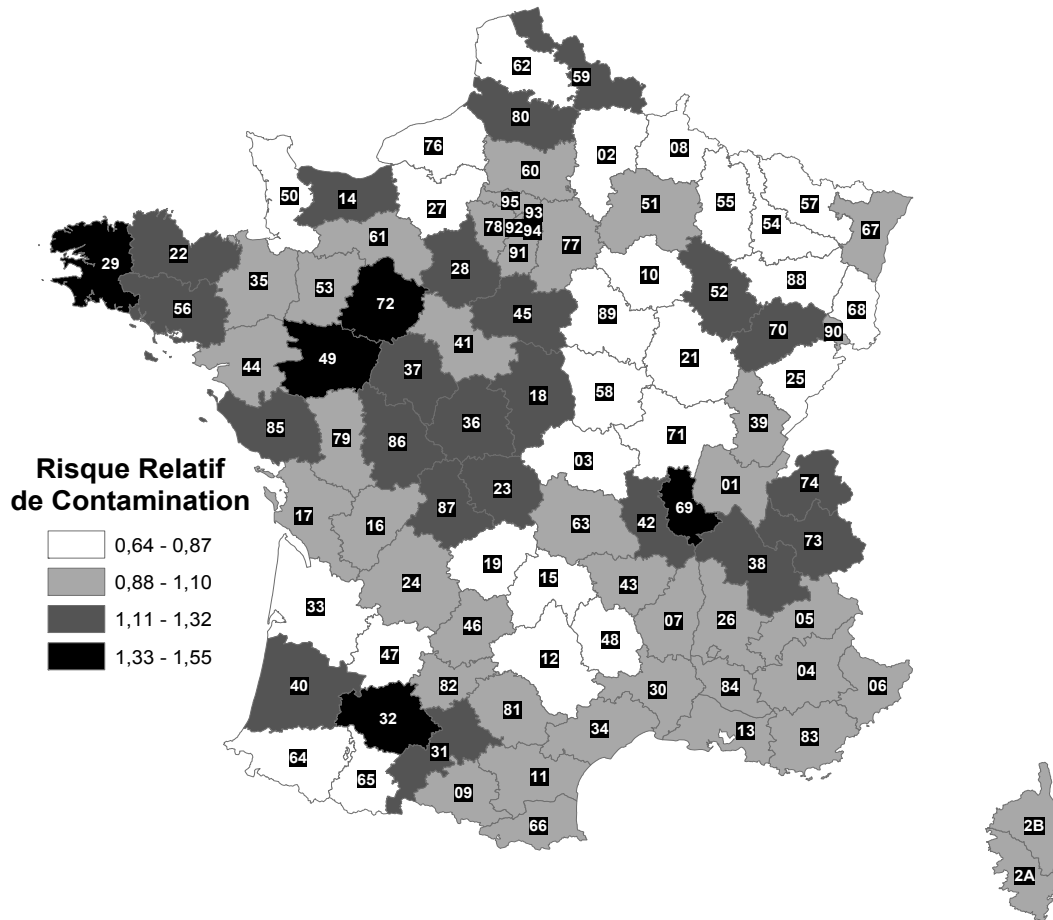


Figure 2

Carte des taux standardisés de contamination par l'ESB dans les 96 départements de France pour la période juillet 2001 à décembre 2002 : 378 cas d'ESB pour une population de 8,5 millions de bovins. Standardisation sur la distinction races laitières/ races allaitantes, méthode empirique bayésienne, lissage global.



La carte du Grand-Ouest, découpée en 526 cantons, présente une hétérogénéité spatiale dans la distribution du RRC. Pour les 84 cas d'ESB observés en 2000, le détail du test est présenté dans les lignes suivantes.

$$\hat{\mu} = 1 ; \hat{\tau}^2 = 0,49$$

$$L_0 = -84,00 ; L_1 = -79,18$$

$$2(L_1 - L_0) = 9,64$$

$$p = 1 - \chi_{9,64(1,ddl)}^2 = 0,002$$

Les 378 cas d'ESB, observés entre juillet 2001 et décembre 2002 sur la France découpée en 96 départements, ne présentent pas un RRC significativement hétérogène ( $p=0,08$ ).

## 2. MISE EN EVIDENCE DES AGREGATS DE CAS

Si l'analyse des taux standardisés permet d'analyser l'hétérogénéité globale, elle ne permet pas toujours de préciser les zones à risque car les UG avec un RRC élevé peuvent être isolées au milieu de zones avec un risque moindre. Ainsi, la lecture des cartes est parfois difficile. Pour compléter l'analyse, il est nécessaire de localiser avec précision les zones à risque par l'analyse des agrégats de cas ou étude des « clusters ». Pour l'analyse des cas d'ESB, deux méthodes d'analyse des agrégats ont été choisies et sont présentées dans les paragraphes suivants.

## 2.1. METHODE DE KULLDORF

Le balayage spatio-temporel de Kulldorff [1997] a deux avantages. Premièrement, c'est une méthode générale qui s'applique à toutes les données présentées sous forme de points géographiques et/ou temporels comme les cas d'ESB. Deuxièmement, cette méthode est très simple d'utilisation et est disponible dans un logiciel du domaine public SaTScan (Kulldorff M. and Information Management Services, Inc. SaTScan v. 3.0: Software for the spatial and space-time scan statistics. Bethesda, MD: National Cancer Institute, 2002). La version 3 est téléchargeable à l'adresse <http://srab.cancer.gov/satscan/download.html>.

La méthode de Kulldorf est basée sur la construction d'un grand nombre de cercles de centre et de diamètre variables sur le territoire étudié. Ces cercles sont centrés tour à tour sur les sommets d'un maillage couvrant tout le territoire, maillage dont le choix influence la précision de localisation des agrégats. Ainsi, un maillage (régulier ou irrégulier) avec des sommets très proches les uns des autres donnera la meilleure précision. Le diamètre des cercles est augmenté continûment de zéro jusqu'à une limite fixée par la population contenue dans le cercle, cette dernière ne devant pas dépasser une fraction de la population totale du territoire, généralement 50%. Ainsi, un très grand nombre de cercles sont construits et chacun est candidat pour être un agrégat de cas.

On note  $p$  la probabilité d'être contaminé pour tout individu appartenant à un cercle donné  $i$  et  $q$  la probabilité d'être contaminé pour tous les individus extérieurs à ce cercle. L'hypothèse nulle est que le nombre de cas est homogène partout sur le territoire, l'hypothèse alternative implique que la probabilité  $p$  est supérieure à  $q$ , c'est-à-dire que le cercle est un agrégat significatif.

$$H_0 : p = q$$

$$H_1 : p > q$$

Un rapport de vraisemblance est alors utilisé sur les données observées, sa forme est la suivante,

$$\delta_{obs} = \frac{\max[L_1(i, p, q)]}{\max[L_0(i, p, q)]} \quad (11)$$

Le maximum de vraisemblance (sur l'ensemble de tous les cercles construits) sous  $H_1$  est donné par,

$$\max[L_1(i, p, q)] = \left(\frac{y_i}{n_i}\right)^{y_i} \left(\frac{n_i - y_i}{n_i}\right)^{n_i - y_i} \left(\frac{Y - y_i}{N - n_i}\right)^{Y - y_i} \left(\frac{N - n_i - Y + y_i}{N - n_i}\right)^{N - n_i - Y + y_i} \quad \text{si } \frac{y_i}{n_i} > \frac{Y - y_i}{N - n_i}$$

$$\max[L_1(i, p, q)] = \frac{Y^Y (N - Y)^{N - Y}}{N^N} \quad \text{si } \frac{y_i}{n_i} \leq \frac{Y - y_i}{N - n_i}$$

et la vraisemblance sous  $H_0$  par,

$$\max[L_0(i, p, q)] = \frac{Y^Y (N - Y)^{N - Y}}{N^N}$$

Dans cette écriture,  $y_i$  est le nombre de cas observés et  $n_i$  est le nombre d'individus dans le cercle  $i$ . Ces derniers sont déterminés en sommant les points géographiques pondérés par les effectifs des cas et de population contenus dans le cercle étudié. La pondération peut être attribuée par les effectifs des communes ou autres UG auxquelles appartiennent les points. Enfin,  $Y$  est le nombre total de cas observés et  $N$  l'effectif total de la population. Le cercle avec la plus forte valeur de vraisemblance (auquel on attribue le rapport maximum  $\delta_{obs}$ ) est considéré comme l'agrégat le plus probable. Il est également possible de sélectionner le second, voire le troisième agrégat le plus probable.

Dans une seconde étape, le logiciel détermine la significativité ( $p$ -value) du ou des agrégats probables par une méthode de Monte Carlo [Dwass, 1957]. Pour cela, il génère aléatoirement un grand nombre de données en tirant au hasard  $Y$  individus parmi la population de  $N$  individus. Les individus tirés sont considérés comme des cas et le rapport maximum  $\delta_{simul}^j$  (11) est calculé pour chaque tirage aléatoire  $j$  (le logiciel effectue par défaut  $nbs=999$  simulations). SaTScan détermine ensuite le rang  $R$  du rapport maximum de vraisemblance  $\delta_{obs}$  des données réellement observées dans l'ensemble des rapports simulés  $\delta_{simul}^j$  triés par ordre décroissant. La  $p$ -value est alors donnée par,

$$p\text{-value} = \frac{R}{1 + nbs} \quad (12)$$

Sur la carte Grand-Ouest pour la période allant de juillet à décembre 2000 (figure 3, carte A), les deux agrégats circulaires tracés (45km de rayon et 18 cas chacun) sont les plus probables ( $\delta_{obs} = 8,48$  et  $\delta_{obs} = 6,33$ ), mais sont non significatifs ( $p$ -values 0,052 et 0,058). Ces valeurs sont à la limite du seuil couramment employé de 5%, ce qui peut être dû au faible nombre de cas. Ces deux agrégats sont aussi mis en évidence [Abrial et



al., 2003] en utilisant une méthode complémentaire, celle de Besag et Newell [1991]. Pour la France entière, période juillet 2001 à décembre 2002 (figure 3, carte B), on observe deux agrégats significatifs : 84 km de rayon pour 46 cas et 116 km pour 45 cas.

## 2.2. METHODE DE BESAG ET NEWELL

La méthode de Besag et Newell [1991] est destinée à la détection des agrégats de maladies rares pour des données (nombre de cas) par UG de petites tailles en utilisant une relation de voisinage entre ces unités. La méthode prend en compte pour chaque UG le nombre d'unités voisines nécessaires pour obtenir un nombre  $k$  préalablement fixé de cas. Ce nombre détermine la taille des agrégats (en nombre de cas) et doit être fixé avant le test, et plusieurs valeurs de  $k$  doivent être testées. Un cluster est valide s'il est significatif pour au moins trois valeurs consécutives de  $k$ .

Comme pour la cartographie des taux standardisés, la zone étudiée est découpée en  $nug=526$  cantons, on note  $y_i$  le nombre de cas et  $n_i$  le nombre de bovins dans l'UG  $i$ . Les cas observés sont pris en compte les uns après les autres : on note  $A_0$  l'UG du cas concerné. Les autres UG sont alors triées de sorte que la distance entre leurs centroïdes et celle de  $A_0$  soit croissante, ces UG voisines sont notées  $A_1, A_2, \dots, A_i, \dots, A_{nug-1}$ . Les nombres de cas observés cumulés, notés  $D_i$ , sont ensuite calculés pour chacune des UG voisines de  $A_0$ . Le cas étudié dans  $A_0$  n'est pas pris en compte pour  $D_i$ , d'où la valeur  $-1$  de la formule suivante,

$$D_i = \left( \sum_{j=0}^i y_j \right) - 1 \text{ avec } i=1, \dots, nug-1 \quad (13)$$

Lors de l'étape suivante, la population à risque est prise en compte par le calcul des effectifs cumulés  $u_i$  de bovins dans chaque unité  $A_i$ ,

$$u_i = \left( \sum_{j=0}^i n_j \right) - 1 \text{ avec } i=1, \dots, nug-1 \quad (14)$$

On définit ensuite la variable aléatoire  $M$  ( $M < nug$ ) indice de l'UG  $A_M$  de sorte que les unités  $\{A_0, A_1, A_2, \dots, A_M\}$  contiennent au moins  $k$  cas observés,

$$M = \min\{i \text{ de sorte que } D_i \geq k\} \quad (15)$$

$M$  est une distance, en nombre d'UG, contenant au moins  $k$  cas (en plus du cas contenu dans  $A_0$ ). Ainsi, une petite valeur observée  $m$  de  $M$  indique la présence d'un agrégat significatif autour de l'unité  $A_0$ . On peut noter que  $M > m$  si on observe moins de  $k$  individus atteints parmi les  $u_m$  individus. La probabilité d'observer exactement  $s$  cas parmi  $u_m$  peut être approximée pour une maladie rare par un terme 'Poissonien' (entre parenthèses dans l'écriture 16). Ainsi, pour une valeur observée  $m$  de  $M$ , la  $p$ -value du test est donnée par,

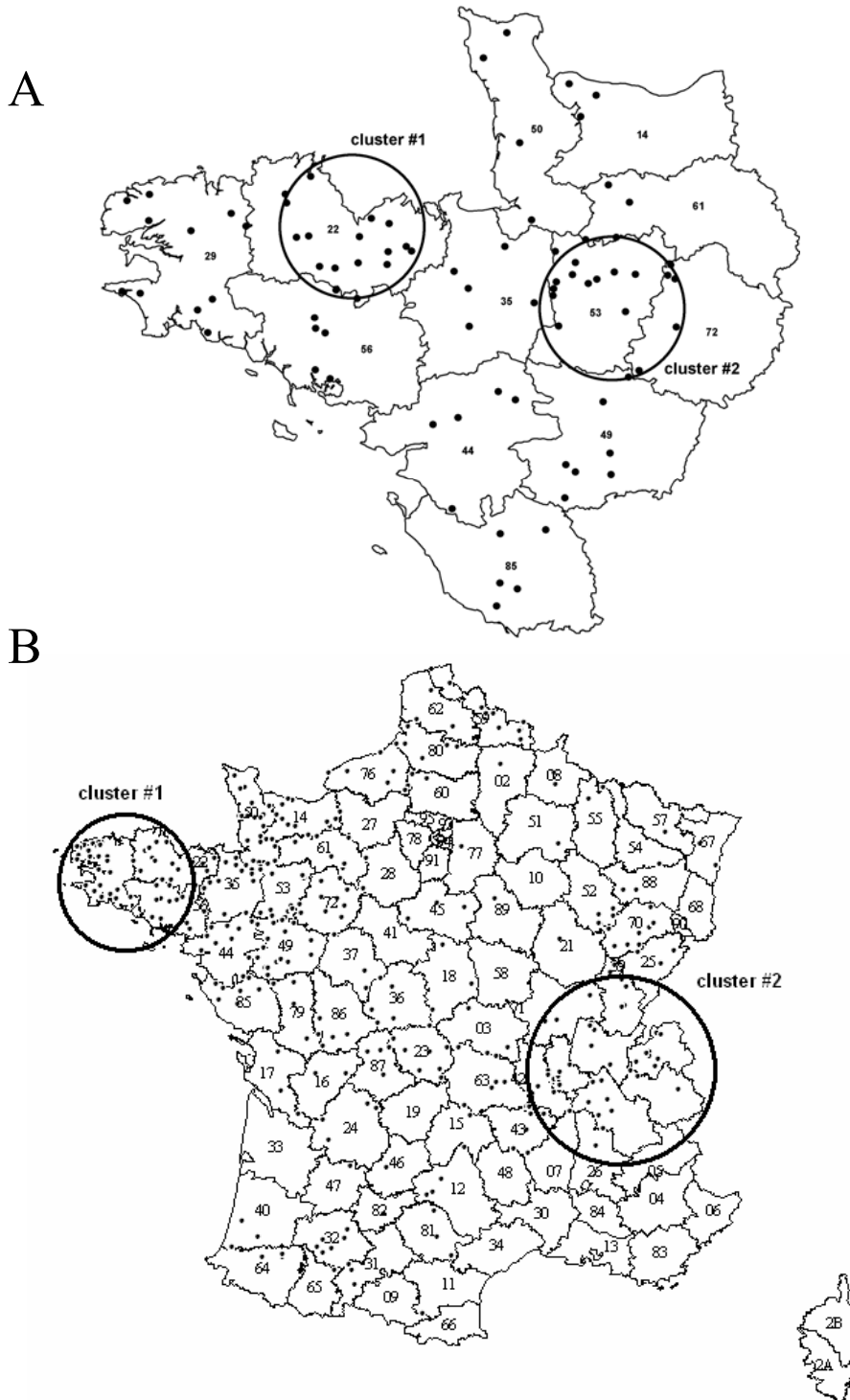
$$P(M \leq m) = 1 - \sum_{s=0}^{k-1} \left( e^{-u_m p} \cdot \frac{(u_m p)^s}{s!} \right) \text{ avec}$$

$$p = \frac{\sum_{i=1}^{nug} y_i}{\sum_{i=1}^{nug} n_i} \quad (16)$$

Cette méthode a été utilisée pour les 84 cas NAIF d'ESB du Grand-Ouest, elle confirme l'existence des deux agrégats mis en évidence par la méthode de Kulldorff (figure 2).

Figure 3

Cartes des agrégats (clusters) des cas d'ESB par la méthode de Kulldorff dans le Grand-Ouest de la France pour la période août à décembre 2000 : 84 cas d'ESB (carte A) et sur la France entière pour la période juillet 2001 à décembre 2002 : 378 cas d'ESB (carte B). Carte A : cluster #1 : 18 cas, 45 km de rayon, p-value 0,052 ; cluster #2 : 18 cas, 44 km de rayon, p-value 0,058. Carte B : cluster #1 : 46 cas, 84 km de rayon, p-value 0,001 ; cluster #2 : 45 cas, 116 km de rayon, p-value 0,008.



## IV - DISCUSSION - CONCLUSION

La première phase de l'étude spatiale a eu pour objectif de rechercher l'existence d'une hétérogénéité dans la distribution géographique des cas d'ESB. Les méthodes utilisées, la cartographie des taux standardisés et les deux méthodes des agrégats (Kulldorff et Besag & Newell) permettent d'affirmer que les cas NAIF ne sont pas répartis au hasard sur le territoire, et cela sur les deux ensembles d'observations : Grand-Ouest de la France entre juillet et décembre 2000 (84 cas NAIF) et France entière entre juillet 2001 et décembre 2002 (378 cas NAIF). Deux zones géographiques présentent un risque de contamination significativement élevé par rapport au reste de la France, les régions Bretagne (11% du cheptel français) et Rhône-Alpes (5,3%). Cette hétérogénéité spatiale des cas NAIF d'ESB a été mise en évidence en considérant des zones et des périodes pour lesquelles la surveillance de l'ESB était optimale et comparable entre régions, et n'est donc pas due à de possibles variations géographiques dans la qualité de la surveillance. Par ailleurs, elle a été analysée en tenant compte de la structure démographique de la population bovine et des différences de risque d'atteinte entre bovins laitiers et allaitants. Ce résultat nous permet d'envisager dans la deuxième phase du projet, l'analyse de l'hétérogénéité spatiale observée en relation avec les hypothèses de contamination alimentaire.

Les méthodes présentées dans cet article sont bien adaptées à l'épidémiologie descriptive. Leur utilisation requière cependant quelques précautions. D'abord, les UG doivent être choisies de sorte que la précision géographique soit adaptée à la problématique, tout en conservant une bonne lisibilité pour faciliter l'interprétation des cartes. Les limites administratives, communément utilisées pour définir les UG, présentent deux inconvénients. C'est d'abord un découpage arbitraire. Ensuite, la forte variabilité de leurs surfaces, l'irrégularité des frontières et de voisinage font des limites administratives des UG peu comparables entre elles. Aussi, en ce qui concerne l'ESB, un découpage par maillage régulier (carrés ou hexagones) est envisagé pour l'étude sur la France entière. La difficulté consiste ici à déterminer la taille des mailles la plus pertinente dans le contexte de notre étude, à savoir les zones de livraison des fabricants d'aliments du bétail.

Autre point, la standardisation des nombres de cas attendus est basée sur les structures

démographiques de la population ; à ce stade, les facteurs pris en compte doivent être en accord avec les hypothèses biologiques de l'étude. Si un facteur démographique important est omis, la carte mettra en évidence des contrastes artificiellement créés par ce facteur. Au contraire, si trop de facteurs sont considérés, les effectifs peuvent être trop faibles ou manquants pour certaines modalités, avec pour conséquence des imprécisions sur les estimateurs. Pour notre part, nous avons considéré comme facteur clé lié au risque d'atteinte la distinction entre types de production laitière et allaitante. Un autre facteur potentiellement important est la structure d'âge de la population, liée à la probabilité de dépister l'ESB. Etant donné qu'elle est peu variable entre régions, et que le seul facteur qui l'influence fortement est le type de production, déjà pris en compte dans l'analyse, nous avons négligé ce paramètre.

Il est aussi important de tester plusieurs formes de lissage du risque (RRC) et différentes méthodes d'estimation des paramètres du modèle probabiliste (méthodes de vraisemblance, itératives ou de Monte Carlo) et de comparer les cartes obtenues. Un tour d'horizon complet des méthodes de cartographie est présenté par Colonna [1997] et Mollié [1999]. Dans cet article, nous avons réalisé un lissage global du RRC en considérant que les UG sont indépendantes les unes des autres. Or, notre hypothèse d'une source alimentaire de contamination (zones de livraison d'aliments du bétail) milite en faveur d'une non indépendance des UG voisines. Tout comme pour d'autres sujets, il pourrait être envisagé une non indépendance entre UG liée aux échanges commerciaux (transports d'aliments et de bétail) ou aux flux écologiques (réseau hydrique, vents, migrations de la faune, ...). Pour tenir compte de cette non indépendance, nous allons compléter l'analyse en intégrant une structure de voisinage pour l'étude sur la France entière. Pour arriver à ce résultat, le voisinage doit être défini par le nombre de voisins à considérer (notions de distances), les hiérarchies et la nature des flux entre UG.

Les méthodes présentées dans cet article en relation avec l'analyse épidémiologique de l'ESB en France sont relativement simples à mettre en œuvre. Leur utilisation sur des données régulièrement mises à jour peut être un outil intéressant pour la surveillance de l'évolution des maladies.

## BIBLIOGRAPHIE

- Abrial D., Calavas D., Lauvergne N., Morignat E., Ducrot C. - Descriptive spatial analysis of BSE in Western France. *Veterinary Research*, 2003, **6**(34), sous presse.
- AFSSA (Agence Française de Sécurité Sanitaire des Aliments) - Les risques sanitaires liés aux différents usages des farines et graisses d'origine animale et aux conditions de leur traitement et de leur élimination, 2001, <http://www.afssa.fr/ftp/basedoc/farines.pdf>
- Besag J., Newell J. - The detection of clusters in rare diseases. *Journal of Royal Statistics Society A*, 1991, **154**(1), 143-155.
- Calavas D., Ducrot C. - L'ESB en France - synthèse sur l'évolution de l'épizootie à partir des données disponibles au 1<sup>er</sup> janvier 2003, 2003, <http://www.afssa.fr/ftp/actu/RapportESB040203.pdf>
- Colonna M. - Estimations bayésiennes empiriques de risques relatifs : principes et exemples d'application. *Revue d'Epidémiologie et de Santé publique*, 1997, **45**, 142-149.
- Dwass M. - Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*, 1957, **12**, 1961-1986.
- Kulldorf M. - Spatial disease clusters: detection and inference. *Statistics in Medicine*, 1995, **14**, 799-810.
- Martuzzi M., Hills M. - Estimating the degree of heterogeneity between event rates using likelihood. *American Journal of Epidemiology*, 1995, **141**(4), 369-374.
- Mollié A. - Bayesian and Empirical Bayes approaches to disease mapping, in: Lawson A., Biggeri A., Bohning D., Disease Mapping and Risk Assessment, 1999, ed. Wiley, Aberdeen, pp. 15-29.
- Prusiner SB. - Novel proteinaceous infectious particles cause scrapie. *Science*, 1982, **216**(4542), 136-44.
- Robert C. ~ L'analyse statistique Bayésienne, 1992, ed. Economica, Paris.
- Savey M., Belli P., Coudert M. - Le réseau d'épidémiosurveillance de la BSE en France : principes - premiers résultats. *Epidémiologie et Santé Animale*, 1991, **19**, 49-61.

