OBSERVATOIRE REGIONAL DE LA SANTE
PROVENCE-ALPES-CÔTE D'AZUR

Proposition d'une démarche de production et d'analyse de cartes d'indicateurs sanitaires

- Décembre 2013 -

Ce projet a été réalisé grâce au soutien financier de l'Agence Régionale de Santé Provence-Alpes-Côte d'Azur.

Sommaire

1 Problématique et objectifs

Les données d'observation sanitaire font partie de la panoplie des informations qui sont habituellement utilisées pour étayer les décisions des pouvoirs publics en matière de planification sanitaire. Ces données sont depuis plusieurs années accessibles au travers d'outils cartographiques, que ces derniers soient utilisés lors d'études ad hoc (cf. étude réalisée par l'Agence régionale de santé –ARS-- en 2012 sur la prévalence des ALD en Provence-Alpes-Côte d'Azur -- Paca) ou dans des applications sur l'Internet dédiées aux professionnels afin de leur permettre une utilisation interactive pour documenter divers types d'indicateurs, tel SIRSéPACA, développé et alimenté par l'ORS Paca. Bien qu'ayant un potentiel illustratif important car elles permettent une visualisation directe de la fréquence des phénomènes de santé et de leurs variations territoriales, ce pour divers types d'indicateurs (taux buts, taux standardisés, indices comparatifs…), les cartes n'en posent pas moins des problèmes d'interprétation épineux.

1.1 Le choix de l'échelle d'analyse et le Modifiable Areal Unit Problem (MAUP)

Les épidémiologistes et géographes de la santé ont rarement l'occasion de choisir leur propre partition du territoire. Ils utilisent souvent des données agrégées à différentes échelles géographiques administratives (Ilots regroupés pour l'information statistique – Iris --, commune...), selon des critères liés au respect de l'anonymat ou de disponibilité des données, et souvent indépendants des objectifs de leur recherche. La littérature montre que la représentation cartographique ainsi que l'interprétation d'une carte sont très sensibles à l'échelle géographique [1]. C'est ce phénomène que les géographes appellent le Modifiable Areal Unit Problem (MAUP) [2], défini comme « la variabilité des résultats, sur les mêmes données et la même zone d'étude, selon la définition des unités spatiales d'analyse ». Le MAUP entraîne deux effets statistiques sur les valeurs des données spatiales [3]. Le premier effet statistique est un effet d'échelle (Figure 1, partie gauche) défini comme « les variations que subissent les données lorsque l'on change de niveau d'observation » [3]. Lors d'une analyse sur de petites unités géographiques, les indicateurs sanitaires sont estimés sur de petits effectifs et ont une variance importante (cf. partie 1.2) ; ceci entraîne une surestimation de l'hétérogénéité spatiale. Par contre, le fait d'agréger des unités géographiques a pour conséquence de lisser les données et donc de diminuer l'hétérogénéité spatiale. En plus du problème de l'effet d'échelle, de nombreuses études [2, 4-6] sur le MAUP ont montré que les résultats des analyses qui s'appuient sur un découpage administratif sont plus fortement dépendants du découpage utilisé. Le second, l'effet de zonage (Figure 1, partie droite), intervient quant à lui lorsque l'on change la forme de ces unités spatiales [7], et souligne ainsi « le rôle des normes des découpages territoriaux sur les résultats » [3].

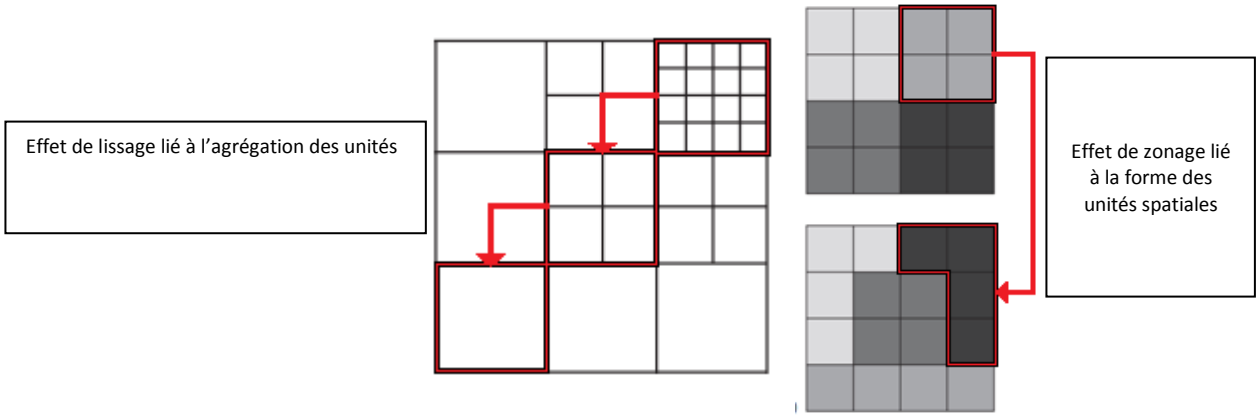


Figure 1. Illustration des effets du Modifiable Areal Unit Problem (adaptée de [1])

1.2 Les problèmes de fiabilité des indicateurs liés aux faibles effectifs dans des zones géographiques de petite taille

Après le choix de l'échelle, un second enjeu est d'obtenir une représentation cartographique fidèle des variations géographiques des indicateurs sanitaires avec comme objectif de séparer de réelles variations spatiales d'éventuellement variations induites par un « bruit statistique ». Dans les unités géographiques les moins peuplées, le nombre de cas d'une cause de décès ou d'une maladie peut être faible, ce qui induit une importante instabilité des indicateurs sanitaires calculés (indices comparatifs, taux standardisés…). Les valeurs les plus élevées et les plus faibles des indicateurs sont d'ailleurs plus fréquemment observées dans les zones géographiques de petite taille et ce problème est d'autant plus marqué que l'événement étudié est peu fréquent. La figure 2 illustre ce phénomène dans le cas du calcul d'un indice comparatif de morbidité (ou standardized morbidity ratio - SMR) : les valeurs extrêmes du SMR (>3), sont observées dans les unités géographiques présentant un faible nombre de cas attendus (c'est-à-dire le nombre de cas estimés sous une incidence de référence).

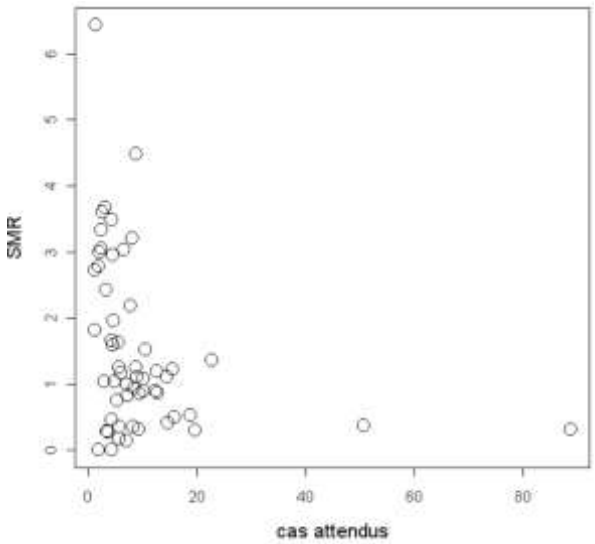


Figure 2. Illustration de l'instabilité d'un indice comparatif de morbidité (SMR base 1) selon le nombre de cas attendus sur chaque unité géographique (adaptée de [8])

1.3 L'identification de zones où le risque de maladie est plus élevé ou plus faible

Les représentations cartographiques d'indicateurs sanitaires prennent souvent un aspect plus ou moins hétérogène, juxtaposant des zones dans lesquelles les valeurs des indicateurs sont élevées et des zones où elles ne le sont pas. Un des enjeux est alors de mieux comprendre la distribution du risque de maladie dans l'espace : le risque est-il réparti de façon aléatoire sur le territoire ? Les unités spatiales adjacentes ou proches ont-elles des risques similaires ? Peut-on définir des agrégats d'unités géographiques avec niveaux de risque élevés ou faibles ? Ces questions renvoient notamment à la notion d'autocorrélation spatiale (ACS), définie comme la ressemblance des valeurs prises par un indicateur selon la proximité des unités géographiques. Les représentations cartographiques d'indicateurs sanitaires présentent souvent une ACS positive (Figure 3, partie gauche). En effet, les individus proches dans l'espace ont plus tendance à partager des caractéristiques socio-économiques similaires ou à être exposés aux mêmes facteurs environnementaux (polluants…) par exemple et ont donc plus de risque de développer les mêmes pathologies (allergies par exemple).

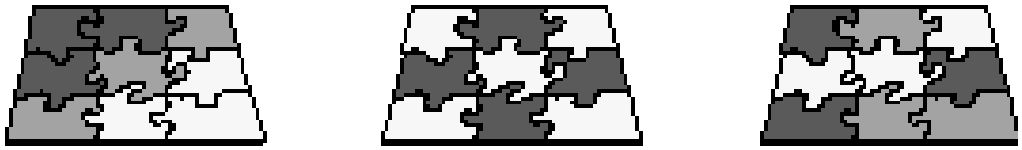


Figure 3. Illustration du phénomène d'autocorrélation spatiale (adaptée de [9])

1.4 Objectifs de l'étude

Dans ce contexte et afin de répondre à ces différentes problématiques, l'ARS Paca a confié à l'ORS Paca le développement d'une démarche systématique de production et d'analyse de cartes d'indicateurs sanitaires qui puisse être appliquée en routine. Cette démarche sera testée sur une base de données d'Affections de longue durée (ALD) constituée par l'ARS Paca et un indicateur de mortalité.

Sur la base d'une revue de la littérature, ce document recense un ensemble de techniques statistiques et géographiques d'analyse de cartes d'indicateurs sanitaires répondant aux différentes problématiques soulevées dans les parties précédentes. Etant donné la nature des données fournies par l'ARS (données agrégées à l'échelle des codes communes de la région Paca), ce document recense en particulier les méthodes écologiques, dans lesquelles les données ne sont disponibles qu'à l'échelle d'unités géographiques. Des méthodes adaptées à des données individuelles seront néanmoins présentées, mais non appliquées.

Enfin, une attention particulière a été portée à l'identification de méthodes implémentées dans les outils de traitement de données spatiales et de traitement statistique disponibles en libre accès sur Internet (Geoda, SaTScan) ou couramment utilisées (SAS et ArcGis).

2 Choisir l'échelle géographique d'analyse

2.1 Un choix qui dépend du type de données disponibles

Le choix de l'échelle géographique utilisée pour la représentation cartographique d'indicateurs sanitaires dépend en premier lieu du type de données disponibles. La production de cartes d'indicateurs sanitaires peut s'appuyer sur des données d'observation de différents types :

- Données ponctuelles (ex : ensemble des cas d'une pathologie géocodés à l'X/Y).
- Données agrégées :
    - Selon un découpage administratif (ex : nombre de cas d'une pathologie dans une commune).
    - Selon des zones d'études non administratives (ex : nombre de cas d'une pathologie dans un carreau de 200m*200m dans le cas de données carroyées).

En pratique, les données sanitaires disponibles en routine à partir des bases de données médico-administratives françaises (statistiques de décès, ALD, données sur les séjours hospitaliers issus du PMSI…) ne sont pas accessibles sous forme de données ponctuelles ou même de données carroyées, notamment en raison de problèmes liés à la confidentialité des données. Elles sont le plus souvent disponibles à l'échelle de zones administratives telles que les communes. A noter que l'utilisation de données ponctuelles ou agrégées à une échelle géographique très fine telle que les carreaux nécessite qu'un géocodage fiable des adresses de résidence des individus puisse être réalisé en amont.

2.2 Comment minimiser les effets du MAUP

Comme indiqué précédemment (cf. partie 1.1), les représentations cartographiques d'indicateurs sanitaires sont très sensibles à la façon dont le découpage géographique a été défini.

Un des principaux challenges en épidémiologie spatiale est de s'affranchir des limites administratives en travaillant par exemple sur des données carroyées. Bien qu'il n'existe aucune règle ni aucune convention internationale pour minimiser les effets du MAUP [10], Briggs [11] liste des objectifs à respecter lors de la définition des unités spatiales d'analyse :

1. Les unités spatiales doivent fournir un découpage homogène du territoire, pour faciliter à la fois la représentation visuelle, l'interprétation des données et l'analyse de tendances spatiales. Les techniques de carroyage (i.e. découpage du territoire en carreaux de tailles identiques) permettent par exemple de répondre à ce premier objectif.
2. La désagrégation des données doit être minimale. Ces méthodes impliquent en effet certaines approximations induisant de possibles sources d'erreurs.
3. L'échelle doit être suffisamment fine pour permettre l'observation de variabilités locales et assez large pour limiter l'instabilité des estimations. Si des techniques statistiques, comme les modèles de lissage (voir partie 3), peuvent être mises en œuvre pour corriger l'instabilité des estimations liées aux petits effectifs sur certains territoires, elles peuvent avoir pour conséquence de « sur-lisser » les données[1] et

[1] Sur-lissage : poids des données trop faible sur la représentation cartographique. Le poids du lissage est très important et toutes les valeurs sont ramenées à la moyenne sur l'ensemble du territoire (forte diminution de l'hétérogénéité spatiale)

donc de masquer les territoires avec des valeurs élevées [12]. En amont de toute représentation cartographique, il est ainsi utile de vérifier si l'échelle géographique est adaptée à l'indicateur sanitaire. Ce point est discuté plus en détails dans la partie 3.3.

Afin de répondre aux objectifs cités ci-dessus, une solution peut consister à construire de nouvelles zones d'études, ne suivant pas les frontières administratives, et ce grâce à des outils informatiques spécifiques (AZM, ZDes, AZTool..). Ces outils permettent d'agréger itérativement des unités spatiales (carreaux), de façon à former de plus larges zones, incluant une contrainte de contigüité [1, 13].

2.3 Choix de l'échelle dans le cadre du présent projet

Dans le cadre des travaux menés avec l'ARS, les indicateurs sanitaires (ALD, statistiques de mortalité) sont disponibles à l'échelle des codes communes Insee, un découpage administratif très irrégulier quant à la forme et à la taille des unités géographiques qui le composent. Les indicateurs sanitaires analysés dans le cadre de projet pouvant avoir des prévalences très faibles, il pourra être envisagé de travailler à une échelle géographique plus agrégée que la commune comme les espaces de santé de proximité (ESP).

Enfin, une des perspectives du projet pourrait être de réaliser un test expérimental sur une zone géographique restreinte (Marseille par exemple) avec des données « ALD » géocodées à l'X/Y ou carroyées[2].

[2] La CnamTS dispose d'une base « adresses » géocodée à l'X-Y (géocodage réalisé par l'Insee et revu chaque année) pour tous les assurés de France.

3 Réaliser des cartographies d'indicateurs sanitaires fiables en tenant compte de l'instabilité des indicateurs liée aux petits effectifs

3.1 Deux types de représentation cartographique envisageables

Comme souligné précédemment, les indicateurs sanitaires (taux standardisés, indices comparatifs) calculés sur des unités géographiques de petite taille (i.e. faiblement peuplées) sont instables. Sur les représentations cartographiques, les valeurs extrêmes faibles ou élevées des indicateurs sont ainsi souvent observées dans les territoires faiblement peuplés. Afin de contourner ce problème, deux solutions sont envisageables : 1) représenter sur une même carte la valeur de l'indicateur et sa significativité statistique ; 2) utiliser des méthodes de lissage afin de corriger les estimations sur les unités avec de petits effectifs.

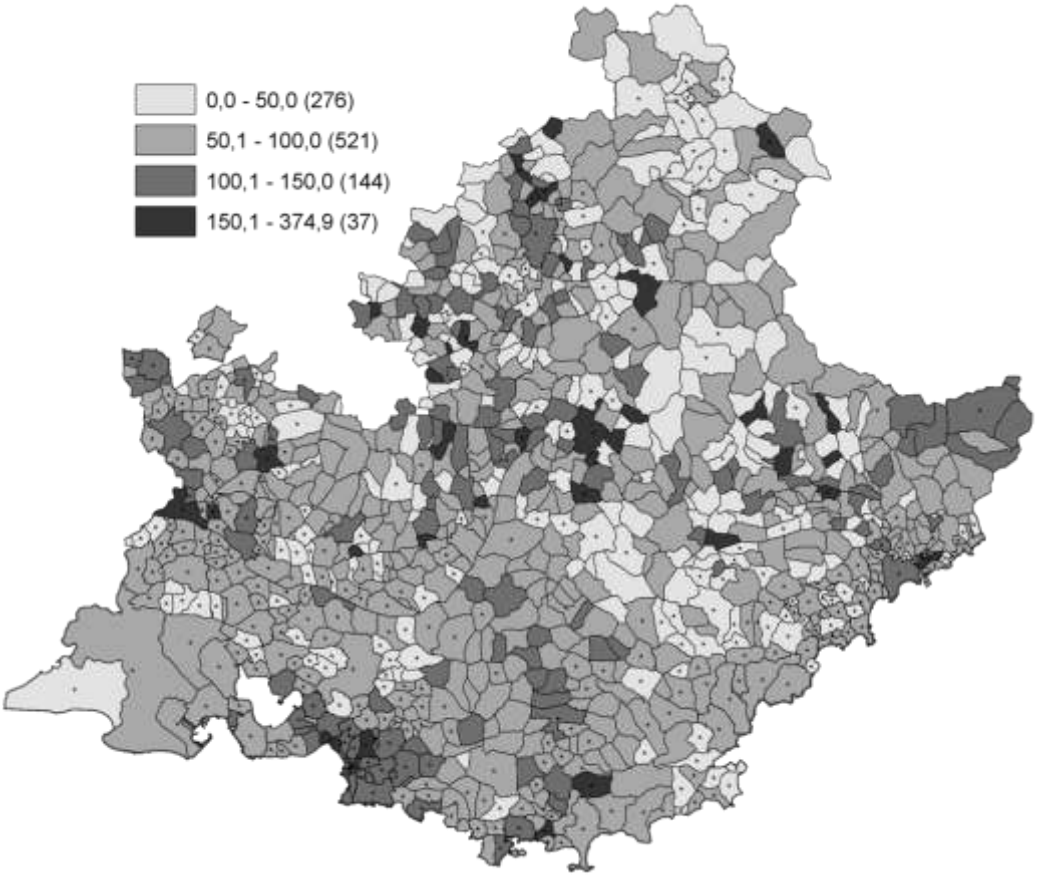
Le principal avantage de la première approche est qu'elle ne nécessite pas de traitement statistique élaboré. Les tests de significativité statistique, basés sur des tests du khi-deux dans le cadre d'un indice comparatif [14], peuvent être réalisés avec un simple tableur. En revanche, cette approche présente l'inconvénient de devoir représenter sur une même carte des plages de couleurs (pour représenter la valeur de l'indicateur) et des symboles (pour indiquer la significativité statistique), ce qui ne facilite pas la lecture de cette dernière. Ce problème est illustré sur la figure 4 où il est difficile d'identifier les territoires avec une valeur d'indice comparatif significativement différente de 100. De plus, le fait de cartographier des territoires avec des valeurs extrêmes non statistiquement significatives gêne la visualisation de la distribution spatiale de l'indice comparatif sur la région Paca.

Du fait de ces inconvénients, nous privilégierons donc dans ce projet la seconde approche, qui est la plus couramment utilisée dans les travaux de géographie de la santé. Les parties suivantes de ce rapport présentent en détail différentes méthodes de lissage utilisables.

3.2 Méthode de lissage : exemple de l'indice comparatif

Les représentations cartographiques d'indicateurs sanitaires lissés sont largement privilégiées aux représentations sur données non lissées car elles permettent d'atténuer le « bruit statistique » causé par l'instabilité des indicateurs et permettent de mieux apprécier la distribution d'un indicateur sanitaire dans l'espace [15, 16]. Ces méthodes dites de « disease mapping » sont particulièrement adaptées aux analyses réalisées à une échelle géographique fine sur des événements peu fréquents.

L'objectif d'une représentation cartographique d'un indicateur est souvent de comparer une valeur observée sur une unité géographique à la moyenne sur l'ensemble du territoire ou à d'autres unités géographiques. Ces comparaisons ne sont toutefois valides que si certains facteurs associés au risque de maladie ne diffèrent pas significativement entre ces unités géographiques. On sait, par exemple, que le sexe, l'âge ou la catégorie socioprofessionnelle sont des déterminants importants des maladies et il est rare que les unités géographiques soient comparables sur l'ensemble de ces facteurs. Avant de procéder au lissage, il est donc nécessaire de réaliser une standardisation des indicateurs sanitaires.



Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca
**Les communes marquées par une étoile ont un SMR statistiquement plus élevé ou plus faible que la valeur de référence (100).Test du khi-deux, seuil de significativité 5 %.*

Figure 4. Indice comparatif (base 100) de la prévalence de l'ALD 23 en région Paca au 31 décembre 2011 à l'échelle des communes avec test de significativité statistique*

L'indicateur standardisé souvent représenté sur les cartes de données sanitaires est l'indice comparatif[3] (ou Standardized mortality ratio (SMR) – également applicable à d'autres données que les statistiques de décès). Dans le cadre du SMR, on parle de standardisation indirecte. Sur chaque unité géographique, le SMR est défini comme le ratio entre un nombre de cas observés et un nombre de cas attendus (sous l'hypothèse d'une incidence de référence, généralement mesurée sur l'ensemble du territoire d'étude – ensemble de la région Paca par exemple). Si les variables de standardisation sont toutes catégorielles, on définit une strate comme le croisement des modalités de ces variables.
Prenons l'exemple d'une standardisation sur le sexe et l'âge qui sera appliquée sur les données du présent projet :

- Sexe en deux modalités : homme / femme
- Âge en cinq modalités : 0-29 / 30-49 / 50-64 / 65-79 / 80+

Au total, on dénombre dix strates (2*5).

[3] Même si l'on s'intéresse ici au cas particulier du SMR (indicateur le plus souvent utilisé pour décrire des données de mortalité ou de morbidité), l'ensemble des techniques de lissage présentées ci-après peuvent être adaptées au calcul de n'importe quel taux.

Pour procéder à la standardisation, il est nécessaire de disposer, pour chaque unité géographique et chaque strate, du nombre total de cas observés (nombre de cas pour une ALD donnée chez les hommes de moins de 30 ans par exemple) et de la taille de strate dans l'unité géographie (ex : nombre d'habitants de moins de 30 ans et de sexe masculin). A partir de ces données, il sera possible de calculer un taux d'incidence ou de prévalence de référence pour chaque strate (en règle générale calculé sur l'ensemble des unités géographiques) et donc d'estimer le nombre de cas attendus pour chaque unité géographique selon la formule :

$$E_i = \sum_{p=1}^{S} n_{ip} R_p$$

Avec :

- E_i = nombre de cas attendus sur l'unité géographique i
- S = nombre total de strates
- n_{ip}= taille de la strate p de l'unité géographique i
- R_p= taux d'incidence/prévalence de référence pour la strate p

Le SMR peut alors être défini comme l'estimateur du maximum de vraisemblance du risque relatif (RR) de l'unité géographique. Les variations du nombre de cas observés sont modélisées par une loi de Poisson :

$$Y_i \sim Poisson\ (E_i \theta_i)$$

Avec:

- Y_i = nombre de cas observés sur l'unité géographique i
- θ_i = risque relatif sur l'unité géographique i

L'estimateur de maximum de vraisemblance de θ_i :

$$\widehat{\theta_i} = SMR_i = \frac{Y_i}{E_i}$$

Avec

$$\boldsymbol{var(SMR_i) = \frac{Y_i}{E_i^{\,2}}}$$

Plus le nombre de cas attendus E_i est faible, plus la variance de l'indicateur est élevée (Figure 2). Des méthodes de lissage des SMR ont donc été développées pour produire des estimations des SMR plus fiables. On distingue deux grands types de lissage : les approches spatiales et les approches non spatiales.

3.3 Lissage et échelle géographique

En statistique, pour toute règle de décision ou test, deux types d'erreurs peuvent être commises :

- Un « faux positif » : par exemple, conclure à tort que le SMR d'une zone est plus élevé que la moyenne sur l'ensemble du territoire. On parle aussi d'une erreur de type I ou d'un manque de spécificité du test.
- Un « faux négatif » : par exemple, ne pas conclure que le SMR d'une zone est plus élevé que sur l'ensemble du territoire à tort. On parle alors d'une erreur de type II ou d'un manque de sensibilité du test.

Les objectifs des techniques de disease mapping vis-à-vis de ces deux types d'erreurs peuvent être résumés de la façon suivante :

- Maximiser la spécificité en corrigeant les valeurs extrêmes liées au problème des petits effectifs (ex : maximiser les chances de ne pas conclure que le SMR d'une zone est plus élevé que la moyenne sur l'ensemble du territoire quand cela est faux)
- Maximiser la sensibilité pour permettre de détecter sur la carte les zones avec de vrais risques élevées ou faibles (ex : maximiser les chances de conclure que le SMR d'une zone est plus élevé que la moyenne sur l'ensemble du territoire quand cela est vrai)

Un problème récurrent lorsque l'on lisse des données est le choix de l'échelle géographique de représentation qui va déterminer les effectifs attendus sur chaque unité géographique. Une diminution de l'échelle aura pour conséquence une diminution des effectifs sur chaque unité géographique et donc une baisse la sensibilité. Ainsi, bien que l'objectif soit en règle générale de représenter les données à l'échelle géographique la plus fine possible, le fait de travailler sur des unités géographiques avec de très faibles effectifs peut entrainer un « sur-lissage » des données et faire disparaitre toute hétérogénéité spatiale.
Des simulations réalisées sur des modèles de lissage spatiaux par Richardson [17] ont montré que pour des SMR observés se situant entre 200 et 300 et un nombre de cas attendus inférieur à 5, les modèles de lissage auront tendance à avoir une sensibilité inférieure à 75 % (i.e. 75% de conclure que le SMR est supérieur à 100 quand cela est vrai). Pour des SMR compris entre 150 et 200, l'effectif attendu minimal pour assurer une bonne sensibilité est de 20.
Ainsi avant de lisser les données, il est indispensable de vérifier si le nombre de cas attendus n'est pas trop faible, notamment sur les unités géographiques avec un SMR élevé. Si c'est le cas, le lissage aura pour conséquence de faire disparaître toute hétérogénéité spatiale. Afin d'augmenter le nombre de cas attendus sur ces territoires, des techniques d'agrégation comme celles listées dans la partie 2.2 peuvent être mises en œuvre. Une augmentation de l'échelle (i.e. baisse de la résolution) d'analyse peut être aussi envisagée (par exemple, passer du niveau communal à celui des ESP en Paca).

3.4 Méthodes de lissage non spatiales

Avec les méthodes de lissage non spatiales, les estimations des indicateurs sur les unités géographiques avec un faible effectif sont corrigées vers la moyenne de l'indicateur estimée sur l'ensemble du territoire étudié. Pour les unités avec une population importante, le poids des données sur l'unité sera important et l'indicateur lissé sera très voisin de sa valeur avant lissage. Par contre, sur les unités avec des effectifs faibles, le poids associé aux données de l'unité sera plus faible et le lissage sera plus important.

3.4.1 Lissage non spatial sur données agrégées

3.4.1.1 Objectif

L'objectif des méthodes de lissage non spatiales sur données agrégées est de corriger les estimations des indicateurs sanitaires sur les unités géographiques avec un faible effectif vers la moyenne estimée sur l'ensemble du territoire étudié.

3.4.1.2 Prérequis

Afin de lisser un indice comparatif, il faudra donc disposer au minimum :

- Du nombre de cas observés Y_i sur chaque unité géographique (ex : nombre de cas d'une ALD)
- Du nombre de cas attendus E_i sur chaque unité géographique (ex : nombre de cas attendus d'une ALD sous l'hypothèse d'une prévalence de référence calculée sur l'ensemble de la région Paca)

Dans le cadre plus général du lissage d'un taux, il faudra disposer, pour chaque unité géographique du nombre de cas (numérateur) ainsi que de l'effectif (dénominateur) pour chaque unité géographique.

3.4.1.3 Méthodes et logiciels

Trois techniques de lissage non spatiales sont fréquemment mises en œuvre pour lisser des indicateurs de mortalité/morbidité dans la littérature :

- Modèle Poisson-Gamma [18].
- Lissage empirique bayésien [19].
- Estimation complètement bayésienne : modèle poisson lognormal avec effet aléatoire [20].

Le modèle poisson lognormal se décompose ainsi :

$$Y_i|\theta_i \sim Poisson\ (E_i\theta_i)$$

$$\log(\theta_i) = \log\left(\frac{Y_i}{E_i}\right) = \log(SMR_i) = \beta_0 + U_i$$

Avec :

β_0 terme constant sur tout le territoire (moyenne globale)

$$U_i \sim Normale\ (0, \sigma_u^2)$$

U_i effet aléatoire qui permet de modéliser la variabilité extra-Poisson (variabilité inter-unités). Plus σ_u^2 est faible, plus les risques sur les unités géographiques ont des valeurs proches (i.e hétérogénéité spatiale faible).

Le lissage par le modèle poisson lognormal avec effet aléatoire, réalisable sous SAS via la procédure GLIMMIX, est l'approche la plus flexible car elle permet de facilement intégrer des covariables dans le modèle. Ces covariables, mesurées au niveau des unités géographiques (i.e. variables contextuelles), sont en général des facteurs d'exposition connus comme étant étroitement liés à l'événement étudié. Dans le cadre de l'estimation de la mortalité par maladies cardio-vasculaires, on pourrait par exemple ajuster sur la prévalence du tabagisme dans l'unité géographique si cette information était disponible à l'échelle géographique étudiée [21]. L'ajout de covariables dans un modèle de lissage permet d'améliorer la qualité d'ajustement du modèle ; il reste néanmoins optionnel.

3.4.1.4 Sorties

Ces techniques de lissage permettent de récupérer en sortie :

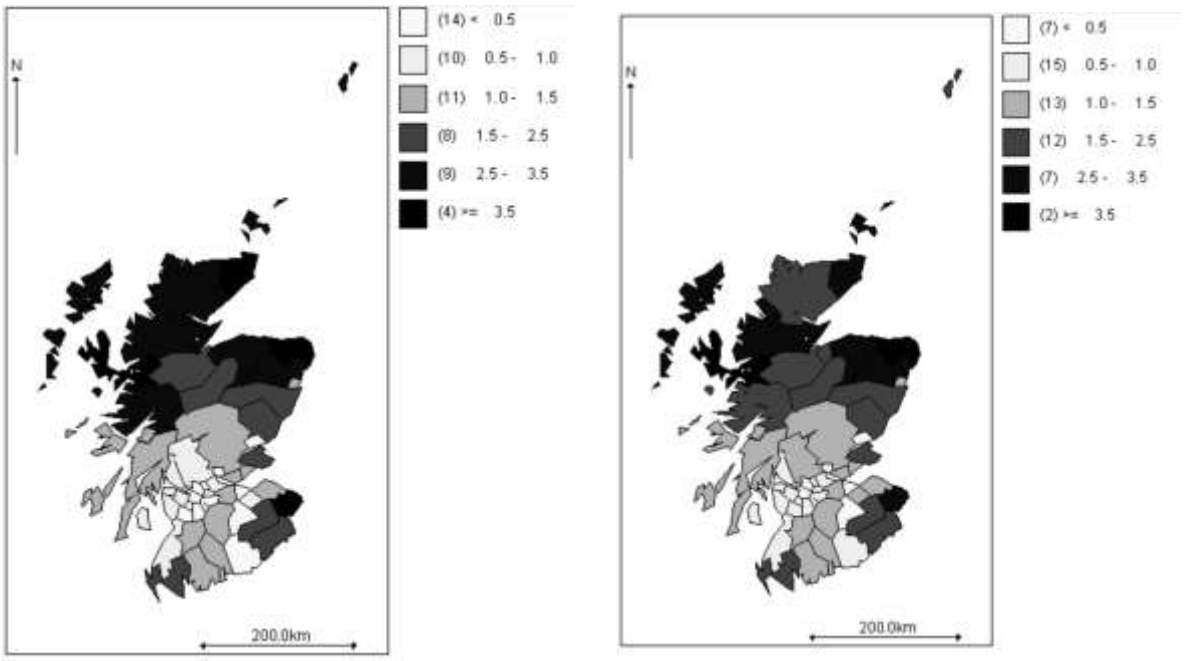
- Une valeur estimée du nombre de cas observés $\widehat{Y}_i$ sur chaque unité géographique. Le SMR lissé de l'unité géographique i se calcule alors selon la formule :

$$SMR'_i = \frac{\widehat{Y}_i}{E_i}$$

Cette valeur lissée du SMR est :

    - Standardisée sur les mêmes facteurs individuels (sexe, âge,…) que les SMR non lissés.
    - Eventuellement ajustée sur les covariables (mesurées au niveau contextuel) du modèle.
- Un test global de significativité (Wald) de l'hétérogénéité spatiale, éventuellement ajusté sur les covariables.
- Statistiques sur la qualité d'ajustement du modèle : log-vraisemblance, critère d'information d'Akaike (AIC) et critère d'information bayésien (BIC).
- Optionnel : estimation des effets des covariables (coefficient de régression) avec tests de significativité (Wald) associé.

Un exemple de lissage est illustré sur la figure 5 : avant lissage du SMR (figure de gauche), le nombre d'unités géographiques avec une valeur inférieure ou égale à 0,5 est de 14. L'effectif de cette catégorie de SMR diminue de moitié après lissage des données (figure de droite). Une diminution de l'effectif des unités géographiques avec des SMR supérieurs à 2,5 est aussi observée après lissage.



SMR non lissés **SMR lissés (modèle poisson gamma)**

Figure 5. Illustration d'un lissage d'un indice comparatif (adaptée de [8])

3.4.1.5 Intérêts et limites

Les méthodes non spatiales ont pour principal avantage d'être facilement mises en œuvre. Les prérequis (voir partie 3.3.1.2) sont les mêmes que pour le calcul d'un SMR non lissé et la plupart des modèles de lissage peuvent être facilement implémentés dans les logiciels habituels de traitement statistique de données.

Le principal inconvénient de l'approche non spatiale est lié au fait qu'elle étudie les variations géographiques des phénomènes en s'appuyant sur un territoire fragmenté sans prendre en compte la structure spatiale des données. Sur le plan statistique, une limite majeure des modèles non spatiaux est qu'ils postulent que deux individus résidant dans des unités géographiques différentes sont absolument indépendants même si les zones sont

proches ou adjacentes. En d'autres termes, ces modèles négligent complètement le problème de l'ACS. Actuellement, les techniques spatiales sont largement privilégiées aux méthodes non spatiales qui ont tendance à « sur-lisser » et à supprimer une hétérogénéité spatiale potentiellement informative [22].
Deuxièmement, les différences observées entre les valeurs avant et après lissage peuvent être une source de questionnement pour les utilisateurs des cartes ; il peut exister des réticences vis-à-vis de l'utilisation de cartes qui présentent des données lissés [23]. Il est ainsi conseillé de publier deux cartes : une carte avec les valeurs non lissées et une carte avec des valeurs issues d'un modèle de lissage [24].

3.4.2 Lissage non spatial sur données non agrégées

3.4.2.1 Objectif

Dans le cas où des données individuelles sont disponibles, l'approche multiniveau peut être envisagée. Cette approche, largement utilisée en France dans les travaux de Basile Chaix [25] en particulier, est plus puissante que les approches écologiques [26, 27] car elle permet de prendre en compte toute l'information disponible (individuelle et contextuelle) et notamment de distinguer la variabilité entre les unités géographiques (inter) et au sein des unités géographiques (intra).

3.4.2.2 Prérequis

Les prérequis de l'approche multiniveaux sont les suivants :

- Affecter tous les individus (cas et population à risque ou témoins) à leur unité géographique de résidence.
- Disposer de covariables, mesurées au niveau individuel, pour tous les individus (cas et population à risque ou témoins). Ces variables sont en règle générale identiques aux variables utilisées lors du processus de standardisation d'un taux standardisé ou d'un indice comparatif (sexe et âge par exemple).
- Optionnel : covariables mesurées au niveau de l'unité géographique.

3.4.2.3 Méthodes et logiciels

Cette approche consiste à mettre en œuvre des modèles multiniveaux logistiques sur deux niveaux.

$$Logit\left(P\left(y_{ij}=1\right)\right)=\beta_0+U_i+\sum_{p=1}^{P}\gamma_{p0}X_{pij}+\sum_{q=1}^{Q}\gamma_{0q}Z_{qj}$$

Avec :

$$Logit\left(P\left(y_{ij}=1\right)\right)=ln\left(\frac{P\left(y_{ij}=1\right)}{1-P\left(y_{ij}=1\right)}\right)$$

y_{ij} réalisation de la variable dépendante Y (ALD/sans ALD par exemple) pour l'individu i appartenant à l'unité géographique j
β_0 constante du modèle, i.e. la valeur moyenne sur tout l'échantillon
$U_i \sim Normale\ (0, \sigma_u^2)$ effet aléatoire qui permet de mesurer la variabilité inter-unités
γ_{p0} effet propre de la covariable mesurée au niveau individuel X_p (p=1...P)
γ_{0q} effet de la covariable au niveau de l'unité géographique Z_p (q=1...Q)

En estimant les résidus de niveau 2 « shruken residuals » $U_i$de ces modèles multiniveaux, il est possible de quantifier et de tester la variabilité inter-unités par l'intermédiaire d'indicateurs comme le Median odds ratio (MOR) ou l'Intra class coefficient (ICC) [26, 28]).

3.4.2.4 Sorties

L'approche multiniveaux logistique permet de récupérer en sortie :

- Une valeur estimée du nombre de cas observés $\widehat{Y_i}$ sur chaque unité géographique ajustée sur l'ensemble des covariables du modèle. Le SMR lissé de l'unité géographique i se calcule alors selon la formule :

$$SMR'_i = \frac{\widehat{Y_i}}{E_i}$$

  Cette valeur lissée du SMR est ajustée sur l'ensemble des covariables (mesurées au niveau individuel et contextuel).
- Un test global de significativité (Wald) de l'hétérogénéité spatiale ajusté sur l'ensemble des covariables.
- Estimation des effets de l'ensemble des covariables au niveau individuel (coefficient de régression, odds ratios) avec tests de significativité (Wald) associé.
- Estimation et comparaison de la variance inter et intra unités géographiques par l'intermédiaire d'indicateurs comme le Median odds ratio (MOR) ou l'Intra class coefficient (ICC) [27] qui mesurent la part des effets de contexte (i.e. effets contextuels) et des effets de composition (i.e. effets individuels).
- Statistiques sur la qualité d'ajustement du modèle : log-vraisemblance, critère d'information d'Akaike (AIC) et critère d'information bayésien (BIC).
- Optionnel : estimation des effets de l'ensemble des covariables au niveau contextuel (coefficient de régression, odds ratios) avec tests de significativité (Wald) associé.

3.4.2.5 Intérêts et limites

Les limites de cette approche sont les mêmes que celles discutées dans la partie 3.3.1.5.
Le principal avantage de la méthode multiniveaux est le fait de pouvoir estimer et comparer variance inter et intra unités géographiques. Dans le cadre des travaux menés avec l'ARS, l'approche multiniveaux ne peut être mise en œuvre puisque les données ne sont disponibles qu'à un niveau agrégé (communes).

3.5 Méthodes de lissage spatial

3.5.1 Lissage spatial sur données agrégées

3.5.1.1 Objectif

L'objectif des méthodes de lissage spatial sur données agrégées est de corriger les estimations des risques sur les unités géographiques en prenant en compte la structure spatiale des données.

3.5.1.2 Prérequis et choix de la structure de voisinage

Les prérequis sont similaires à ceux détaillées dans la partie 3.1.1.2. Afin de prendre en compte la spatialisation des données dans le processus de lissage, il est nécessaire de définir la structure de voisinage entre les unités géographiques. En analyse spatiale, on distingue différents types de structures, dont les plus couramment utilisés étant les suivantes :

- Structures basées sur des notions de distance :
    - Distance inverse
        - Exemple : inverse de la distance euclidienne entre les centroïdes (i.e. barycentres) des communes.
    - Zone d'influence
        - Exemple : les voisins d'une commune sont toutes les communes dont le centroïde est à une distance euclidienne inférieure à une valeur fixée (rayon de 20 km par exemple).
- Structures types « K Nearest Neighbors »
    - Exemple : les voisins d'une commune sont les K communes les plus proches (distance euclidienne entre les centroïdes).
- Structures basées sur des notions de contigüité :
    - Contigüité type « Rook » : les voisins sont l'ensemble des unités géographiques avec une frontière en commun avec l'unité considérée.
    - Contigüité type « Queen » : les voisins sont l'ensemble des unités géographiques avec une frontière ou un sommet en commun avec l'unité considérée.
    - Bishop, Gabriel connectivity…

A noter que pour les structures basées sur des notions de contigüité, il est possible de définir des structures dites du premier ordre ou d'ordres supérieurs. Pour une structure d'ordre 2 par exemple, toutes les unités voisines de l'ensemble des voisins d'ordre 1 d'une unité géographique sont considérées comme des voisins de cette unité. Il est aussi possible d'affecter des poids à chaque voisin, en utilisant par exemple la longueur de frontière partagée entre les deux unités.

De nombreux travaux ont montré que le choix de la structure de voisinage pouvait avoir un impact important sur la représentation cartographique de données lissées [29-32]. Une première étude [29] a analysé la sensibilité des résultats du modèle de lissage spatial de Besag, York et Mollié dit BYM sur trois structures de voisinage. L'analyse, réalisée sur une lattice irrégulière (découpage géographique communal par exemple), soulignait la forte variabilité des résultats du lissage selon la structure choisie ainsi qu'une meilleure adéquation du modèle de lissage avec des structures de contigüité par rapport aux structures de distance. Une autre analyse de sensibilité sur ce même modèle [30] montrait aussi des résultats très sensibles selon la structure choisie. Toujours sur une lattice irrégulière, cette étude insistait sur les bonnes performances des structures de distance quand l'objectif du lissage était de détecter les unités géographiques avec des valeurs élevées ou faibles. L'impact du choix de la structure de voisinage sur les résultats d'un lissage spatial étant variable selon le type de modèle choisi (BYM, Simultaneous Autoregressive Model, modèles linéaires généralisées mixte…) et le type de lattice (régulière type données carroyées ou irrégulière), il est fortement recommandé de tester plusieurs structures afin de vérifier la robustesse des résultats des modèles de lissage [30, 33, 34].

3.5.1.3 Méthodes et logiciels

Le modèle bayésien hiérarchique BYM [35] permet d'obtenir un lissage dit « mixte » : un compromis entre lissage global (obtenu par un modèle poisson lognormal par exemple) et lissage local (obtenu par une moyenne mobile sur une série temporelle par exemple). Le modèle se décompose ainsi :

$$Y_i|\theta_i \sim Poisson\ (E_i\theta_i)$$

$$\log(\theta_i) = \log\left(\frac{Y_i}{E_i}\right) = \log(SMR_i) = \beta_0 + U_i + V_i$$

Avec :

β_0 terme constant sur tout le territoire (moyenne)

$$U_i \sim Normale\ (0, \sigma_u^2)$$

U_i effet aléatoire qui contrôle la variabilité des SMR dans sa composante non spatiale. Une valeur faible de σ_u^2 indique des risques similaires entre unités géographiques (faible hétérogénéité spatiale).

V_i effet aléatoire qui contrôle la variabilité des SMR dans sa composante spatiale. Cet effet suppose que les unités géographiques proches ont tendance à avoir un risque similaire. On utilise le modèle gaussien autorégressif intrinsèque (ICAR) pour prendre en compte la structure spatiale des données :

$$(V_i|V_j = v_j, j \neq i) \sim Normale\ \left(\frac{\sum_{j\neq i} w_{ij} v_j}{\sum_{j\neq i} w_{ij}}, \frac{\sigma_v^2}{\sum_{j\neq i} w_{ij}}\right)$$

Avec :

w_{ij} qui définit la notion de voisinage entre les unités géographiques

σ_v^2 qui permet de contrôler la variabilité du risque dans sa composante spatiale

En d'autre termes, le modèle BYM suppose que la distribution conditionnelle de l'effet V_i dans l'unité géographique i suit une loi normale centrée sur la moyenne des effets de ses unités voisines et de variance inversement proportionnelle au nombre de voisins. Une valeur faible de σ_v^2 indique des risques similaires entre unités géographiques voisines.

L'implémentation du modèle BYM se fait par l'utilisation du logiciel WinBUGS [36]. De récents travaux ont proposé une implémentation du modèle BYW sous le logiciel SAS [37]. Cependant, l'implémentation sous ce logiciel, utilisant la procédure GLIMMIX, permet uniquement la prise en compte de structures de voisinage basées sur une notion de contigüité. Des modèles linéaires généralisées mixtes GLMM [38] peuvent être mis en œuvre afin de prendre en compte une structure basée sur une notion de distance.

3.5.1.4 Sorties

Le lissage spatial sur données agrégées permet d'obtenir des sorties statistiques similaires à l'approche non spatiale sur données agrégées. La liste complète des sorties est disponible dans la partie 3.3.1.4.

3.5.1.5 Intérêts et limites

En géographie de la santé, l'approche spatiale est de plus en plus privilégiée par rapport à l'approche non spatiale. Certains travaux ont montré, que même en l'absence d'ACS dans les résidus d'un modèle non spatial, une modélisation BYM n'entrainait pas de biais dans l'estimation des mesures d'associations avec les variables d'ajustement du modèle [39]. En d'autres termes, il est tout à fait possible de réaliser un lissage spatial sur des données qui ne présentent pas d'ACS.

Le lissage spatial présente toutefois certains inconvénients. En effet, il introduit de l'ACS dans les données. En d'autres termes, des indicateurs instables sont remplacés par des indicateurs spatialement corrélés. La représentation cartographique de données lissées peut ainsi faire apparaitre des agrégats d'unités géographiques avec des valeurs élevées ou faibles. Si l'objectif de l'analyse dépasse le simple cadre de la représentation cartographique, il faudra alors se tourner vers les techniques de détection de clusters (voir partie 4) qui ont justement pour objectif de mettre en évidence ces agrégats d'unités géographiques avec un niveau de risque similaire.

Une autre limite de ces méthodes est la complexité des modèles. Lorsque le nombre d'unités géographique est important, la convergence des modèles n'est pas assurée dans 100 % des cas et les temps de calculs peuvent être très longs.

Enfin, comme pour le lissage non spatial, il est conseillé de publier deux cartes : une carte avec les valeurs non lissées et une carte avec des valeurs issues d'un modèle de lissage.

3.5.2 Lissage spatial sur données non agrégées

3.5.2.1 Objectif

L'objectif des méthodes de lissage spatial sur données non agrégées est de corriger les estimations des risques sur les unités géographiques en prenant en compte la structure spatiale des données.

3.5.2.2 Prérequis

Les prérequis sont similaires à ceux détaillées dans la partie 3.2.2.2. Afin de prendre en compte la structure spatiale des données dans le processus de lissage, il est nécessaire de définir la structure de voisinage entre les unités géographiques (cf. partie 3.5.1.2).

3.5.2.3 Méthodes et logiciels

En modifiant la structure de variance-covariance des résidus de niveau 2 des modèles multiniveaux, il est possible de prendre en compte l'ACS [40]. Des modèles spatiaux mixtes ont ainsi été développés afin de prendre en compte le problème de l'ACS. Ces modèles sont similaires aux modèles BYM discutés précédemment à la différence qu'ici, l'analyse est réalisée sur des données sanitaires non agrégées (modèles sur variables dépendantes binaires et non plus poissoniennes). Ces modèles peuvent être implémentés sous SAS via la procédure GLIMMIX.

3.5.2.4 Sorties

Le lissage spatial sur données non agrégées permet d'obtenir des sorties statistiques similaires à l'approche non spatiale sur données non agrégées. La liste complète des sorties est disponible en partie 3.3.2.4.

3.5.2.5 Intérêts et limites

Les limites de cette approche sont les mêmes que celles discutées dans la partie 3.4.1.5.
Pour les mêmes raisons évoquées précédemment à propos de l'approche multiniveaux classique, l'approche spatiale mixte sur données non agrégées ne peut être mise en œuvre dans le présent projet (données agrégées).

3.6 Intérêt des méthodes de lissage dans le cadre du présent projet

Dans le cadre de l'analyse des données d'ALD et de mortalité de l'ARS, disponibles au niveau des communes de la région Paca, un lissage des données apparait adapté pour plusieurs raisons :

- L'analyse sera réalisée à un niveau géographique (commune[4]) caractérisé par une forte variabilité concernant les effectifs de population (et par conséquent de cas observés et attendus – Tableau 1).
- L'analyse porte sur des indicateurs sanitaires avec une prévalence/incidence faible.
- Les indicateurs sanitaires étudiés sont potentiellement entachés d'une forte ACS.

Tableau 1. Effectifs de population des 978 communes* en région Paca (Source RP 2010 Insee)

Moyenne (écart-type)	Minimum	Q1	Médiane	Q3	Max
5 009 (16 624)	2	231	794	3 486	343 304

* La commune de Marseille étant découpée en 16 arrondissements.

Pour chaque indicateur sanitaire étudié, et avant tout lissage, le nombre de cas attendus sur l'ensemble des communes sera étudié. En se basant sur les simulations de Richardson [17], on vérifiera si le nombre de communes avec un nombre de cas attendus inférieur à 20 et 5 n'est pas trop important. S'il s'avère que la majorité des communes ont un nombre de cas attendus trop faible, il sera envisagé de travailler à une échelle géographique plus agrégée comme les espaces de santé de proximité (ESP).
Dans le présent projet, il sera envisagé de tester les méthodes de lissage spatial sur deux structures spatiales : une structure basée sur les distances entre les centroïdes des communes via un modèle GLMM et une autre basée sur une notion de contigüité (structure « Rook » d'ordre 1) via un modèle BYM. Au total, trois lissages seront réalisés sur chaque indicateur :

- Lissage non spatial avec un modèle lognormal avec effet aléatoire
- Lissage spatial BYM avec structure de contigüité
- Lissage spatial GLMM avec structure de distance

Les résultats des trois modèles seront comparés à l'aide d'indicateurs de qualité d'ajustement comme le critère d'information d'Akaike (AIC). Au final, deux cartes seront retenues : une carte avec les valeurs non lissées et une carte avec des valeurs issues du « meilleur modèle » de lissage (i.e. modèle avec le plus faible AIC).

[4] En raison de la faiblesse des effectifs observée pour certains indicateurs sanitaires à l'échelle des communes, les analyses ont également été réalisées à l'échelle des ESP.

4 Détecter des zones où le risque de maladie est plus élevé ou plus faible (clusters)

4.1 Trois familles de méthodes

Dans le cadre d'une analyse écologique, sur données agrégées, un cluster est défini comme un agrégat d'unités géographiques adjacentes avec un risque similaire (risque élevé ou faible). Sur données ponctuelles, un cluster sera défini comme un agrégat de personnes. De nombreuses méthodes ont été développées pour identifier la présence de clusters d'unités géographiques [41]. On distingue deux grandes familles de techniques de détection de clusters :

- L'approche globale : permet d'évaluer, sans identifier les clusters, si les cas sont distribués aléatoirement sur l'ensemble du territoire.
- L'approche locale : permet d'identifier des agrégats d'unités géographiques.

A noter que la détection d'un cluster significatif n'implique pas une tendance globale au clustering significative et vice versa [42]. Les deux approches sont donc complémentaires.
Enfin, une troisième famille de méthodes de clustering comprenant les tests focalisés ou tests de concentration, permettent l'examen de l'existence d'agrégats en référence à un point spécifique. Ces tests nécessitent une mesure du facteur de risque sur le territoire (par exemple la distance à un point source comme facteur d'exposition). Ces méthodologies ne rentrant pas dans le cadre des objectifs du présent projet, elles ne seront pas présentées dans ce rapport.
Les analyses de clusters peuvent aussi être classées selon le type de données qu'elles permettent d'étudier, agrégées ou non agrégées. On se focalisera ici sur les méthodes adaptées à des données agrégées.

4.2 Méthodes globales

4.2.1 Objectif

On distingue deux types de méthodes globales de détection de clusters. Certaines méthodes permettent d'évaluer la dispersion des cas sur l'ensemble du territoire (i.e. mesure de l'hétérogénéité globale en termes de surdispersion) : est-ce que la dispersion des SMR sur l'ensemble du territoire est trop importante pour être compatible avec les fluctuations aléatoires d'une loi de Poisson ? D'autres méthodes ont pour objectif de mesurer le degré d'ACS sur l'ensemble du territoire : est-ce que les unités géographiques proches ont des SMR similaires ? Près d'une centaine de méthodes existent pour répondre à ces questions [41].

4.2.2 Prérequis

Afin d'évaluer la dispersion des cas sur le territoire, il suffit de disposer de la valeur de l'indicateur sur chaque unité géographique. Dans le cadre du SMR, il faudra disposer pour chaque unité géographique du nombre de cas attendus et observés. Pour mesurer l'ACS, il sera aussi nécessaire de définir la structure de dépendances entre les unités géographiques (voir partie 3.1.1.2).

4.2.3 Méthodes et logiciels

Le test d'homogénéité de Potthoff et Whittinghill [43] et la statistique de Moran [44] sont très souvent utilisés en géographie de la santé [45].
Le test de Potthoff et Whittinghill n'est implémenté que dans le logiciel R. Dans sa version simplifiée (approximation asymptotique par une loi normale), il est toutefois très aisé de calculer cette statistique qui permet de mesurer le degré de dispersion des données.
La statistique de Moran résume le degré de ressemblance des unités géographiques voisines. Elle prend des valeurs situées dans l'intervalle [-1 ; 1] et s'interprète comme un coefficient de corrélation de Pearson ou de Spearman. Le lissage spatial introduisant de l'ACS dans les données, la mesure de l'ACS peut ainsi se faire sur des données non lissées ou des données lissées par des méthodes non spatiales [46]. La statistique de Moran est notamment implémentée dans les logiciels SAS (Procédure VARIOGRAM) et Arcgis.
D'autres tests de global clustering existent : la statistique de Tango, la statistique de Besag-Newell… Parmi l'ensemble des méthodes proposées, la statistique de Tango est la plus puissante statistiquement [47], mais elle n'est disponible que dans un package du logiciel R.

4.2.4 Sorties

- Statistique de test (p_valeur) de Potthoff et Whittinghill. L'hypothèse nulle du test est l'homogénéité de l'indicateur sur l'ensemble du territoire.
- La valeur de l'indice de Moran et test de significativité. Le test permet de comparer la valeur estimée de l'ACS à la valeur nulle.

4.2.5 Intérêts et limites

La principale limite du test de Potthoff et Whittinghill est qu'il n'est implémenté, dans sa version exacte, que dans le logiciel R.
A la différence des autres indicateurs de global clustering, la statistique de Moran est disponible en routine dans la plupart des logiciels statistiques et SIG (SAS, Arcview, R ….).
Une des limites de, la statistique de Moran est qu'elle ne prend pas en compte l'hétérogénéité des effectifs de population. Des versions alternatives de cette statistiques ont néanmoins été proposées pour en tenir compte [42].
Comme la majorité des approches spatiales, la mesure de l'ACS est affectée par le MAUP qui peut avoir un impact sur la puissance des tests réalisés [48-50]. Enfin, la valeur de la statistique de Moran est sensible au choix de la structure de voisinage. Le choix de la structure de visionnage optimale est un sujet très discuté dans la littérature et aucune recommandation n'existe à l'heure actuelle sur ce sujet [51-54].

4.3 Méthodes locales

4.3.1 Objectif

L'objectif des méthodes locales de clustering est de localiser des agrégats d'unités géographiques avec des niveaux de risque significativement plus faibles ou plus élevés que sur le reste du territoire.

4.3.2 Prérequis

On se place ici dans le cadre standard de données sanitaires agrégées de comptage (ex : nombre de cas d'une ALD dans une commune). Afin d'identifier des clusters, il faudra au minimum disposer :

- Du nombre de cas observés Yi sur chaque unité géographique (ex : nombre de cas d'une ALD).
- Du nombre de cas attendus Ei sur chaque unité géographique (ex : nombre de cas attendus d'une ALD sous l'hypothèse d'une prévalence de référence mesurée sur l'ensemble de la région Paca).

On pourra aussi disposer de covariables mesurées au niveau des unités géographiques (i.e. variables contextuelles). L'ajout de covariables dans le modèle permettra alors d'identifier des clusters qui ne seront pas expliqués par ces covariables. Enfin, comme toute approche spatiale, il faudra définir une structure de voisinage entre les unités géographiques.

4.3.3 Méthodes et logiciels

Parmi les méthodes de détection de clusters, les *Local indicators of spatial associations* LISA [51], le *Getis-Ord Gi* et la statistique de scan (scan statistic) [47, 55, 56] sont les plus populaires.

Les LISA, développés par Anselin, sont une extension locale de l'indice de Moran et permettent de mesurer le degré de ressemblance d'une unité spatiale avec ses voisines. Les travaux d'Anselin sur les LISA se place dans la continuité de celui de Getis et Ord qui avaient déjà cherché à mettre en place des indicateurs locaux d'association spatiale, mais sans lien avec les indicateurs globaux existants. Les LISA et le Getis-Ord Gi sont implémentés dans la plupart des systèmes d'information géographique.

La statistique de scan est une méthode très puissante qui s'applique aussi bien sur des données agrégées que non agrégées. Une fenêtre, de forme circulaire ou elliptique et centrée sur une unité géographique, balaye la zone d'étude (Figure 6). Pour chaque taille de la fenêtre, un test est réalisé. L'hypothèse alternative de ce test est qu'il existe un risque plus élevé à l'intérieur de la fenêtre par rapport à l'extérieur de la fenêtre. Une fois toutes les tailles testées, la fenêtre se déplace sur une autre unité géographique. Au final, la fenêtre qui correspond au maximum de vraisemblance est le cluster le plus probable (i.e. cluster principal) et une statistique de test (p_valeur), calculée à partir de simulations de Monte Carlo, est assignée à ce cluster. D'autres clusters dits clusters secondaires peuvent aussi êtres identifiés par le modèle. La statistique de scan est implémentée dans le logiciel ad-hoc SaTSCan.

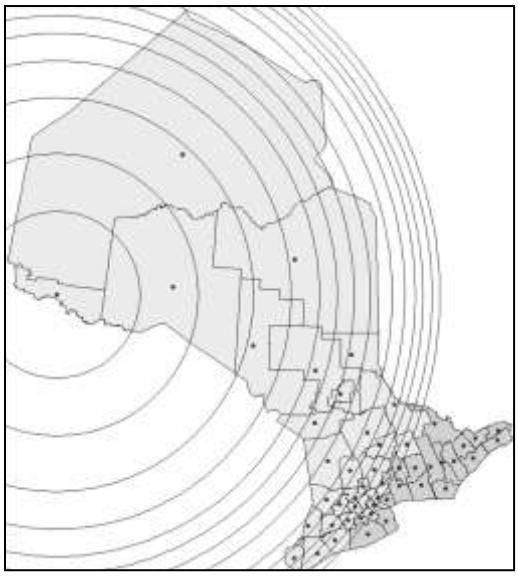


Figure 6. Illustration de fenêtres circulaires de tailles variables (adaptée de [57])

4.3.4 Sorties

En sortie, la statistique de scan permet de cartographier les clusters identifiés par le modèle (avec ajustement possible sur les covariables du modèle – voir Figure 7). La table de sortie donne les informations suivantes :

- Nombre total de clusters identifiés : un cluster principal (i.e. celui avec la p_valeur la plus faible) et des clusters secondaires.
- Composition des clusters : liste des unités géographiques rattachées à chaque cluster.
- Mesure de la significativité de chaque cluster (p_valeur).
- Mesure du « risque » sur chaque cluster. Dans le cadre de données de comptage, le logiciel renvoie une estimation du risque relatif (i.e. risque dans le cluster versus risque en dehors du cluster).

Figure 7.Statistique de scan : exemple type de sortie (adaptée de [57])

4.3.5 Intérêts et limites

Le principal avantage de la méthode de scan est sa souplesse. Elle s'adapte aussi bien à des données ponctuelles qu'agrégées et permet l'ajustement sur des covariables (catégorielles exclusivement) ce qui n'est pas le cas pour les autres méthodes de détection de clusters souvent utilisées en épidémiologie spatiale ; on citera par exemple les *Local indicators of spatial associations* LISA [51] ou le *Getis-Ord Gi* [58]. De plus, si ces méthodes ont l'avantage d'être implémentées dans la plupart des SIG, elles ne permettent pas de prendre en compte l'hétérogénéité des effectifs entre les unités géographiques.

Une des limites de la statistique de scan est le fait qu'elle ne soit implémentée que sous le logiciel SaTScan. Les résultats ne sont pas directement visualisables dans le logiciel. Il est donc nécessaire d'importer les sorties du logiciel SatSCan dans un SIG afin de pouvoir visualiser les clusters identifiés par le modèle (Figure 7). De plus, il est impossible d'imposer un nombre maximal de clusters identifiables par le modèle. Il est néanmoins possible de contrôler la taille maximale d'un cluster en terme de surface (ex : rayon maximal d'un cluster circulaire fixée à 10km) ou en terme d'effectif (ex : un cluster ne peut contenir plus de 50 % de la population de la région Paca). Enfin, le logiciel ne gère que les matrices de voisinage basées sur des distances (clusters circulaires ou elliptiques) et il n'est pas possible de construire des clusters à partir d'une notion d'adjacence.

4.4 Intérêt des méthodes de clustering dans le cadre du présent projet

Tout comme les méthodes de lissage, la détection de clusters est particulièrement adaptée à l'analyse de données mesurées sur de petites unités géographiques. Dans le cadre du présent projet, la détection de clusters de communes pourrait être intéressante à plusieurs titres.

Dans un premier temps, deux approches globales, par l'intermédiaire du test de Potthoff et Whittinghill et de la statistique de Moran, permettraient de vérifier si les risques sont distribués de manière homogène sur le territoire et/ou si unités géographiques voisines ont tendance à présenter des niveaux de risque similaires. Afin de contrôler l'hétérogénéité des effectifs entre les communes, la mesure de l'ACS devra être effectuée sur des données lissées par une approche non spatiale. La mesure de l'ACS étant sensible au choix de la structure de voisinage, deux calculs pourraient être réalisés : sur une structure d'adjacence simple type Rook 1 et sur une structure basée sur les distances entre les centroïdes des communes.

Dans un second temps, l'approche locale permettrait, sur une sélection d'indicateurs, de vérifier certaines hypothèses soulevées par la représentation cartographique. La détection de cluster pourrait alors se faire en deux étapes :

- Recherche de clusters avec un modèle vide, sans covariable.
- Recherche de clusters avec un modèle ajusté sur des variables potentiellement liées à l'événement étudié. Dans le cadre de l'ALD diabète type 2 par exemple, l'ajustement sur le niveau socio-économique de l'unité géographique pourrait être envisagé. Pour l'ALD 23 « affections psychiatriques de longue durée », l'impact de l'ajustement sur la distance au centre médico-psychologique le plus proche pourrait être étudié.

5 Démarche proposée dans le cadre du présent projet

La démarche proposée dans le cadre du présent projet se décompose sept étapes (Figure 8). Elle peut permettre de répondre à deux objectifs : d'une part, **réaliser des cartes d'indicateurs sanitaires fiables en tenant compte de l'instabilité des indicateurs liée aux petits effectifs (étapes 2 à 4)** et, d'autre part, **détecter des zones où le risque de maladie est plus élevé ou plus faible (étapes 5 à 7)**. Ces deux types analyses peuvent être réalisés de manière totalement indépendante dans le sens où il est tout fait possible de se limiter à une représentation cartographique lissée ou à une analyse de clusters. Toutefois, dans plupart des cas, il est préférable, dans un premier temps, de cartographier l'indicateur sanitaire afin de bien appréhender la distribution des valeurs sur l'ensemble du territoire et de pouvoir générer certaines hypothèses. Concernant l'analyse de clusters, il est tout à fait possible de se limiter à une analyse globale ou locale. Les différentes étapes de la démarche proposée sont décrites ci-dessous.

1. **Evaluer si l'échelle géographique est adaptée à l'indicateur sanitaire.** Dans cette première étape, le nombre de communes avec un nombre de cas attendus inférieur à 5 et à 20 est étudié. S'il s'avère que la majorité des communes ont un nombre de cas attendus trop faible, une analyse à un niveau géographique plus agrégé (ESP par exemple) sera envisagée.

2. **Cartographier l'indicateur non lissé.** Cette étape permet d'obtenir une première représentation cartographique des données et de faire le diagnostic du problème des valeurs extrêmes, souvent estimées sur des unités géographiques avec de petits effectifs.

3. **Lisser les données selon trois approches.** Dans cette seconde étape, une approche non spatiale (modèle Poisson lognormal avec effet aléatoire : étape 3a) et deux approches spatiales (BYM : étape 3b ; et GLMM : étape 3c) sont testées. La sensibilité des résultats au type de structure de voisinage choisi (contigüité ou distance) est aussi vérifiée lors de ces étapes. Le modèle avec l'AIC le plus faible (i.e. meilleure qualité d'ajustement) est ensuite sélectionné (étape 3d).

4. **Cartographier l'indicateur lissé.** Deux représentations cartographiques sont proposées (cartographie non lissée et celle issue du « meilleur modèle » de lissage).

5. **Analyser la tendance globale au clustering.** La dispersion des données sur l'ensemble du territoire d'analyse est évaluée à l'aide du test de Potthoff et Whittinghill, réalisé sur des données non lissés. L'autocorrélation spatiale est estimée à l'aide de l'indice de Moran. Ce dernier ne prenant pas en compte l'hétérogénéité des effectifs entres les communes, le calcul est réalisé sur des données lissées par un modèle de Poisson lognormal avec effet aléatoire (i.e. lissage non spatial) sur deux structures de voisinage différentes (contigüité ou distance).

6. **Analyser les tendances locales au clustering.** La recherche de clusters de communes de forte prévalence (« high risk clusters ») est effectuée par la statistique de scan. Un premier modèle « vide », sans covariable, est testé (étape 6a). Afin de supprimer

l'effet de certains facteurs contextuels (niveau de précarité socio-économique, offre de soins…), il pourra éventuellement être envisagé de réaliser des modèles ajustés (étape 6b).

7. **Cartographier les clusters, comparer les modèles et interpréter les résultats.** Les clusters (avant et après ajustement) sont cartographiés et les résultats des modèles comparés.

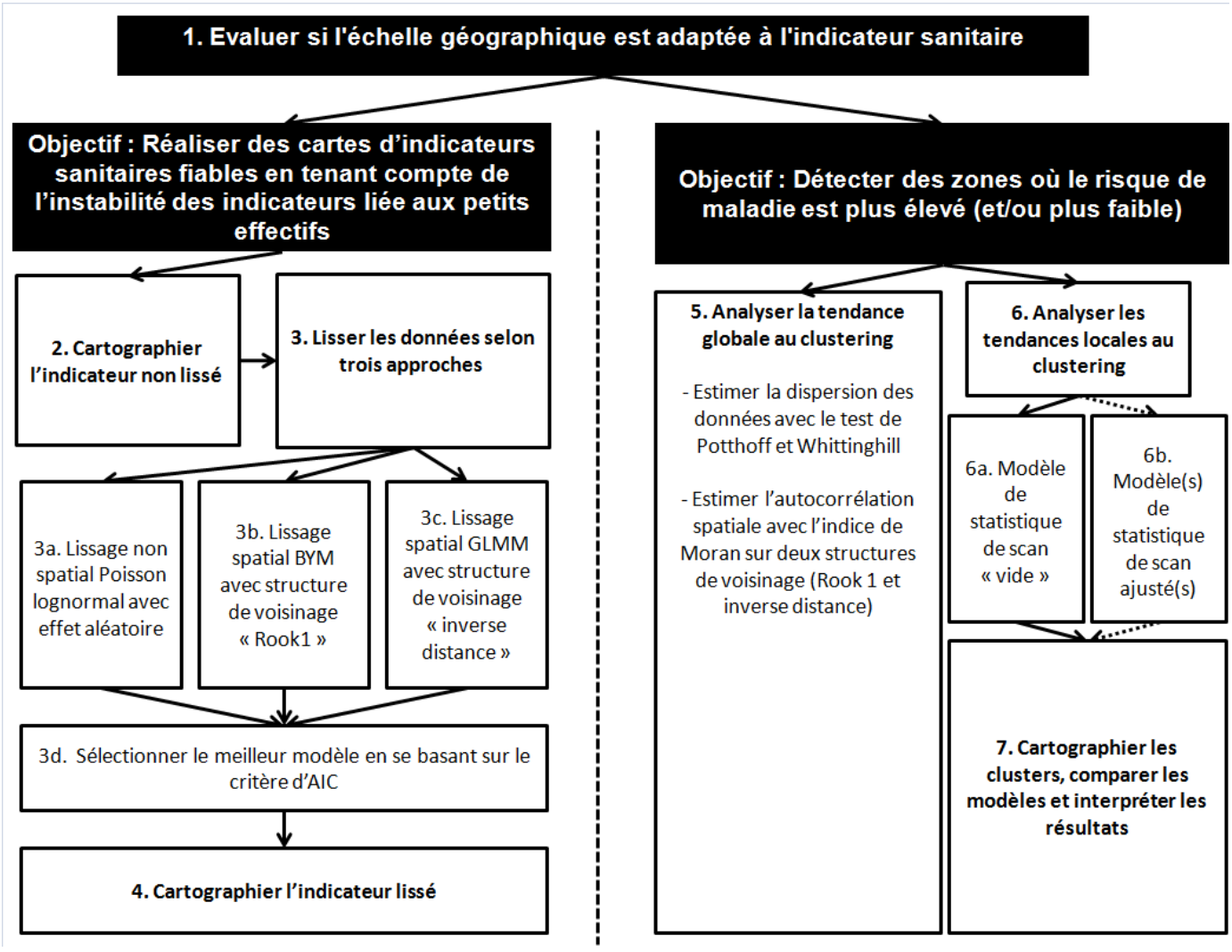


Figure 8. Démarche globale proposée pour la production et l'analyse de cartes d'indicateurs sanitaires

6 Synthèse des résultats sur les indicateurs d'affections de longue durée et de mortalité

La démarche proposée a été testée sur sept indicateurs d'ALD et un indicateur de mortalité. Le Tableau 2 propose une synthèse des résultats obtenus. L'ensemble des résultats détaillés sont disponibles dans les annexes de ce rapport. En accord avec l'ARS et sur la base d'hypothèses spécifiques, l'analyse locale de clusters par la statistique de scan n'a été testée que sur deux indicateurs (ALD 23 et ALD 8 E11) et ne sera pas discutée dans cette partie.
Pour les indicateurs avec une prévalence inférieure à 2 % (Tableau 2), la part de communes avec un nombre de cas attendus très faible (<5) pouvait être très importante. Pour les ALD pour cancers du poumon par exemple, cette part atteignait 78 %. Le niveau communal apparaissait donc trop fin pour représenter ces indicateurs sanitaires. Du fait d'effectifs trop faibles, on ne pouvait rejeter l'hypothèse d'homogénéité des SMR entre les communes (p_valeurs des tests de Potthoff et Whittinghill supérieures à 0,05) et, du fait du poids du lissage, il était très délicat d'obtenir une estimation fiable de l'autocorrélation spatiale. Une seconde approche, au niveau des espaces de santé de proximité (ESP) a donc était proposée pour ces indicateurs.

Pour les trois indicateurs pour lesquels une représentation cartographique au niveau communal apparaissait adaptée (ALD 5 et 13, ALD 23 et ALD 8 E11), le modèle BYM s'est révélé être le « meilleur modèle de lissage ». Les représentations cartographiques après lissage permettaient de mieux apprécier la distribution de ces indicateurs sanitaires sur l'ensemble de la région Paca. L'autocorrélation spatiale était significativement positive pour les trois indicateurs et le test de Potthoff et Whittinghill indiquait des variations du SMR significatives entre communes pour les ALD 23 et 8 E11.

Tableau 2. Synthèse des résultats sur les indicateurs d'ALD et de mortalité au niveau communal (n = 978)

	Taux de prévalence pour 100 000 habitants	SMR non lissé (Min-Max)	SMR lissé (Min-Max)	Modèle retenu	AIC (Min-Max)	Hétérogénéité spatiale	Indice de Moran
ALD 5 et 13 "maladies cardiaques »†	3 499	0-241	59-146	BYM	205-324	p=0,07	0,32 (p<0,01)
ALD 23 "troubles psychiatriques"†	2 428	0-375	35-350	BYM	1253-1408	p<0,01	0,22 (p<0,01)
ALD 8 E11 "diabète de type 2"†	3 107	0-210	45-207	BYM	770-933	p<0,01	0,35 (p<0,01)
ALD cancers du sein chez la femme†	1 977	0-456	72-124	Non spatial	878-886	p=0,36	0,23 (p<0,01)
ALD cancers du côlon†	266	0-888	85-129	GLMM	1864-1895	p=0,48	0,06 (p<0,01)
ALD cancers de la vessie†	181	0-1 047	74-132	BYM	2300-2317	p=0,44	0,09 (p<0,01)
ALD cancers du poumon†	124	0-1 448	85-113	Non spatial	2569-2661	p=0,48	0,02 (p<0,30)
Mortalité par cancers chez les moins de 65 ans entre 2006 et 2011 ††	82	0-809	42-156	BYM	1800-1822	p=0,42	0,13 (p<0,01)

† Nombre de cas d'ALD en région Paca au 31 décembre 2011 à l'échelle des communes. Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca. Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).

†† Mortalité par cancers en région Paca entre 2006 et 2011 à l'échelle des commune chez les moins de 65 ans. Standardisation selon l'âge (0-29/30-49/50-64) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca. Source : Inserm CépiDC

Références

1. Sabel, C.E., et al., *Creation of synthetic homogeneous neighbourhoods using zone design algorithms to explore relationships between asthma and deprivation in Strasbourg, France.* Social Science & Medicine. **91**(0): p. 110-121.
2. Fotheringham, A.S. and D.W.S. Wong, *The modifiable areal unit problem in multivariate statistical analysis.* Environment and Planning A, 1991. **23**(7): p. 1025-1044.
3. Oliveau, S., *Autocorrélation spatiale : leçons du changement d'échelle.* Espace Géographique, 2010: p. 51-64.
4. Amrhein, C.G., *Searching for the elusive aggregation effect: evidence from statistical simulations.* Environment and Planning A, 1995. **27**(1): p. 105-119.
5. Martin, D., *An assessment of surface and zonal models of population.* International Journal of Geographical Information Systems, 1996. **10**(8): p. 973-989.
6. Holt, D., D.G. Steel, and M. Tranmer, *Area homogeneity and the modifiable areal unit problem.* Geographical Systems, 1996. **3**: p. 181 - 200.
7. Flowerdew, R., D.J. Manley, and C.E. Sabel, *Neighbourhood effects on health: does it matter where you draw the boundaries?* Soc Sci Med, 2008. **66**(6): p. 1241-55.
8. World Health, O., et al., *Atlas of cancer in Scotland, 1975-1980 : incidence and epidemiological perspective*. 1985, Lyon : New York: International Agency for Research on Cancer ; Distributed in the United States by Oxford University Press.
9. Hansen, H.S. *Avenue - a powerful environment for developing spatial data analysis tools*. [cited; Available from: http://proceedings.esri.com/library/userconf/europroc97/11technology/t2/t2.htm.
10. Openshaw, S., *The Modifiable Areal Unit Problem*. 1984: Elsevier Science Geo Abstracts.
11. Briggs, D., D. Fecht, and K. De Hoogh, *Census data issues for epidemiology and health risk assessment: experiences from the Small Area Health Statistics Unit.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 2007. **170**(2): p. 355-378.
12. Elliott, P. and D. Wartenberg, *Spatial epidemiology: current approaches and future challenges.* Environ Health Perspect, 2004. **112**(9): p. 998-1006.
13. Parenteau, M.-P. and M. Sawada, *The modifiable areal unit problem (MAUP) in the relationship between exposure to NO2 and respiratory health.* International Journal of Health Geographics. **10**(1): p. 58.
14. Bouyer, J., *Epidémiologie: principes et méthodes quantitatives*. 2009: Lavoisier.
15. Kafadar, K., *Smoothing geographical data, particularly rates of disease.* Stat Med, 1996. **15**(23): p. 2539-60.
16. Waller, L.A.a.C.A.G., *Applied spatial statistics for public health data*, ed. Wiley-Interscience. 2004.
17. Richardson, S., et al., *Interpreting posterior relative risk estimates in disease-mapping studies.* Environ Health Perspect, 2004. **112**(9): p. 1016-25.
18. Guihenneuc-Jouyaux, C., *[Statistical modelization of geographic variations: a major challenge in epidemiology and statistics].* Rev Epidemiol Sante Publique, 2002. **50**(5): p. 409-12.
19. Clayton, D. and J. Kaldor, *Empirical Bayes estimates of age-standardized relative risks for use in disease mapping.* Biometrics, 1987. **43**(3): p. 671-81.
20. Lunn, D.J., et al., *WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility.* Statistics and Computing, 2000. **10**(4): p. 325-337.

21. Li, H., B.I. Graubard, and M.H. Gail, *Covariate adjustment and ranking methods to identify regions with high and low mortality rates.* Biometrics. **66**(2): p. 613-20.
22. Auchincloss, A.H., et al., *A review of spatial methods in epidemiology, 2000-2010.* Annu Rev Public Health. **33**: p. 107-22.
23. Waller, L.A. and C.A. Gotway, *Applied Spatial Statistics for Public Health Data*. 2004: Wiley.
24. Anca Carrington, M.R., Bruce Mitchell, Patrick Heady,Nargis Rahman, *Smoothing of Standardised Mortality Ratios: A Preliminary Investigation*, in *Methodological Series No 35*. 2007, National Statistics.
25. Lynch, K.F., et al., *Context and disease when disease risk is low: the case of type 1 diabetes in Sweden.* Journal of Epidemiology and Community Health. **64**(9): p. 789-795.
26. Merlo, J., et al., *A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena.* J Epidemiol Community Health, 2006. **60**(4): p. 290-7.
27. Merlo, J., et al., *A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon.* J Epidemiol Community Health, 2005. **59**(6): p. 443-9.
28. Larsen, K. and J. Merlo, *Appropriate Assessment of Neighborhood Effects on Individual Health: Integrating Random and Fixed Effects in Multilevel Logistic Regression.* American Journal of Epidemiology, 2005. **161**(1): p. 81-88.
29. Conlon, E.M. and L.A. Waller, *Flexible spatial hierarchical models for mapping disease rates.* In Proceedings of A.S.A. Section on Statistics and the Environment: Washington, DC, 1999: p. 82 - 87.
30. Earnest, A., et al., *Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models.* International Journal of Health Geographics, 2007. **6**(1): p. 54.
31. Best, N., S. Richardson, and A. Thomson, *A comparison of Bayesian spatial models for disease mapping.* Stat Methods Med Res, 2005. **14**(1): p. 48.
32. Brett, C. and J. Pinkse, *Those Taxes are all over the Map! A Test for Spatial Independence of Municipal Tax Rates in British Columbia.* International Regional Science Review, 1997. **20**(1-2): p. 131-151.
33. Wakefield, J., *Statistical Methods in Spatial Epidemiology: AB Lawson. Chichester, UK: Wiley, 2001, pp.277, Â£55.00 (HB). ISBN: 0471975729.* International Journal of Epidemiology, 2001. **30**(6): p. 1504-1505.
34. Stakhovych, S. and T.H.A. Bijmolt, *Specification of spatial models: A simulation study on weights matrices.* Papers in Regional Science, 2009. **88**(2): p. 389-408.
35. Besag, J., J. York, and A. MolliÃ©, *Bayesian image restoration, with two applications in spatial statistics.* Annals of the Institute of Statistical Mathematics, 1991. **43**(1): p. 1-20.
36. Lunn, D., et al., *WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility.* Statistics and Computing, 2000. **10**(4): p. 325-337.
37. Rasmussen, S., *Modelling of discrete spatial variation in epidemiology with SAS using GLIMMIX.* Comput Methods Programs Biomed, 2004. **76**(1): p. 83-9.

38. Mohebbi, M., R. Wolfe, and D. Jolley, *A poisson regression approach for modelling spatial autocorrelation between geographically referenced observations.* BMC Medical Research Methodology. **11**(1): p. 133.
39. Latouche, A., et al., *Robustness of the BYM model in absence of spatial variation in the residuals.* International Journal of Health Geographics, 2007. **6**(1): p. 39.
40. Havard, S., et al., *Social inequalities in residential exposure to road traffic noise: an environmental justice analysis based on the RECORD Cohort Study.* Occup Environ Med. **68**(5): p. 366-74.
41. Kulldorff, M., *Tests of Spatial Randomness Adjusted for an Inhomogeneity.* Journal of the American Statistical Association, 2006. **101**(475): p. 1289-1305.
42. Lan, H., et al., *Evaluating spatial methods for investigating global clustering and cluster detection of cancer cases*. Vol. 27. 2008, Chichester, ROYAUME-UNI: Wiley. 32.
43. Potthoff, R.F. and M. Whittinghill, *Testing for homogeneity. II. The Poisson distribution.* Biometrika, 1966. **53**(1): p. 183-90.
44. Goodchild, M.F., *Spatial autocorrelation*. 1986: Geo Books.
45. Dufour B, H.P., *La surveillance épidémiologique en santé animale*. Quae éditions ed. 2007.
46. Anselin, L., *Exploring Spatial Data with GeoDaTM : A Workbook*, C.f.S.I.S. Science, Editor. 2005.
47. *Introduction aux statistiques spatiales et aux systèmes d'information géographique en santé environnement*. 2011, InVS.
48. Waller, L.A., E.G. Hill, and R.A. Rudd, *The geography of power: statistical performance of tests of clusters and clustering in heterogeneous populations.* Stat Med, 2006. **25**(5): p. 853-65.
49. Viel, J.-F., N. Floret, and F. Mauny, *Spatial and space-time scan statistics to detect low rate clusters of sex ratio.* Environmental and Ecological Statistics, 2005. **12**(3): p. 289-299.
50. Gaudart, J., et al., *[Spatial cluster detection without point source specification: the use of five methods and comparison of their results].* Rev Epidemiol Sante Publique, 2007. **55**(4): p. 297-306.
51. Anselin, L., *Local Indicators of Spatial Association—LISA.* Geographical Analysis, 1995. **27**(2): p. 93-115.
52. Qi, Y. and J. Wu, *Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices.* Landscape Ecology, 1996. **11**(1): p. 39-49.
53. Bellec, S., D. Hemon, and J. Clavel, *Answering cluster investigation requests: the value of simple simulations and statistical tools.* Eur J Epidemiol, 2005. **20**(8): p. 663-71.
54. Wakefield, J. and P. Elliott, *Issues in the statistical analysis of small area health data.* Stat Med, 1999. **18**(17-18): p. 2377-99.
55. Kulldorff, M., et al., *An elliptic spatial scan statistic.* Stat Med, 2006. **25**(22): p. 3929-43.
56. Kulldorff, M. and N. Nagarwalla, *Spatial disease clusters: detection and inference.* Stat Med, 1995. **14**(8): p. 799-810.
57. Patrick DeLuca, M.A., *Cluster Analysis using SaTScan*, A. Conference, Editor. 2007: Ottawa.

58. Getis, A. and J.K. Ord, *The Analysis of Spatial Association by Use of Distance Statistics.* Geographical Analysis, 1992. **24**(3): p. 189-206.

Annexes

Annexe 1. ALD 23 - Affections psychiatriques de longue durée

Le taux de prévalence brut de l'ALD 23 est de 2 428 pour 100 000 habitants en région Paca.

- Lissage

En se basant sur le critère d'AIC, le modèle de lissage spatial BYM est retenu comme « meilleur » modèle (tableau 1). L'effet du lissage sur la représentation cartographique de la prévalence de l'ALD 23 est modéré (figure 1). Le niveau communal apparait adapté à la représentation cartographique de la prévalence de cet indicateur sanitaire. Une version de la carte avec une méthode de discrétisation mieux adaptée à la distribution des SMR lissés est présentée en figure 2.

Tableau 1. Synthèse des résultats des modèles de lissage

	SMR		
	Min-Max	Ecart interquartile (Q3-Q1)	AIC
Données non lissées	0-375	45	NA
Lissage non spatial	34-359	22	1395
Lissage spatial BYM	35-350	24	1253
Lissage spatial GLMM	31-363	24	1408

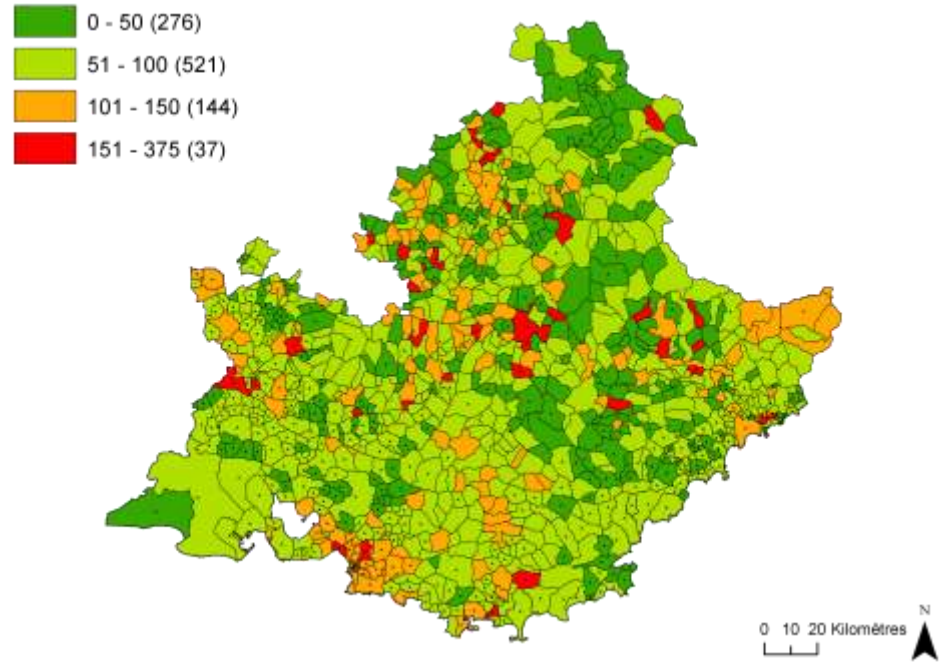
NA : non applicable pour les données non lissées.

- Tendance globale au clustering

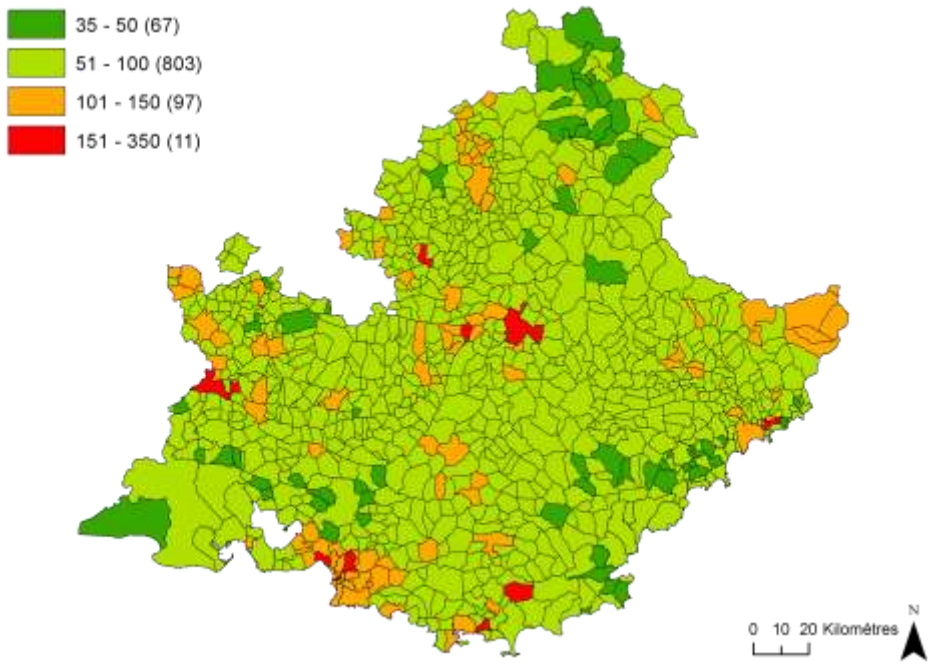
Le test de Potthoff et Whittinghill permet de conclure à une surdispersion des données ($p<0,01$). Le SMR de d'ALD 23 n'est donc pas distribué de manière homogène sur l'ensemble du territoire.

Afin de contrôler l'instabilité des estimations des SMR, le calcul de l'indice de Moran est réalisé sur des données lissées par une méthode non spatiale (modèle poisson lognormal avec effet aléatoire). En se basant sur une structure de contigüité, l'indice de Moran est significativement supérieur à zéro ($I = 0,20$ – $p<0,01$). La modification de la structure de voisinage (distance entre les centroïdes des communes) n'a quasiment pas d'incidence sur la valeur de l'indice de Moran ($I = 0,22$ – $p<0,01$). On peut donc conclure à la présence d'une ACS positive modérée (i.e. les communes adjacentes ont tendance à avoir des valeurs de SMR similaires).

SMR non lissés



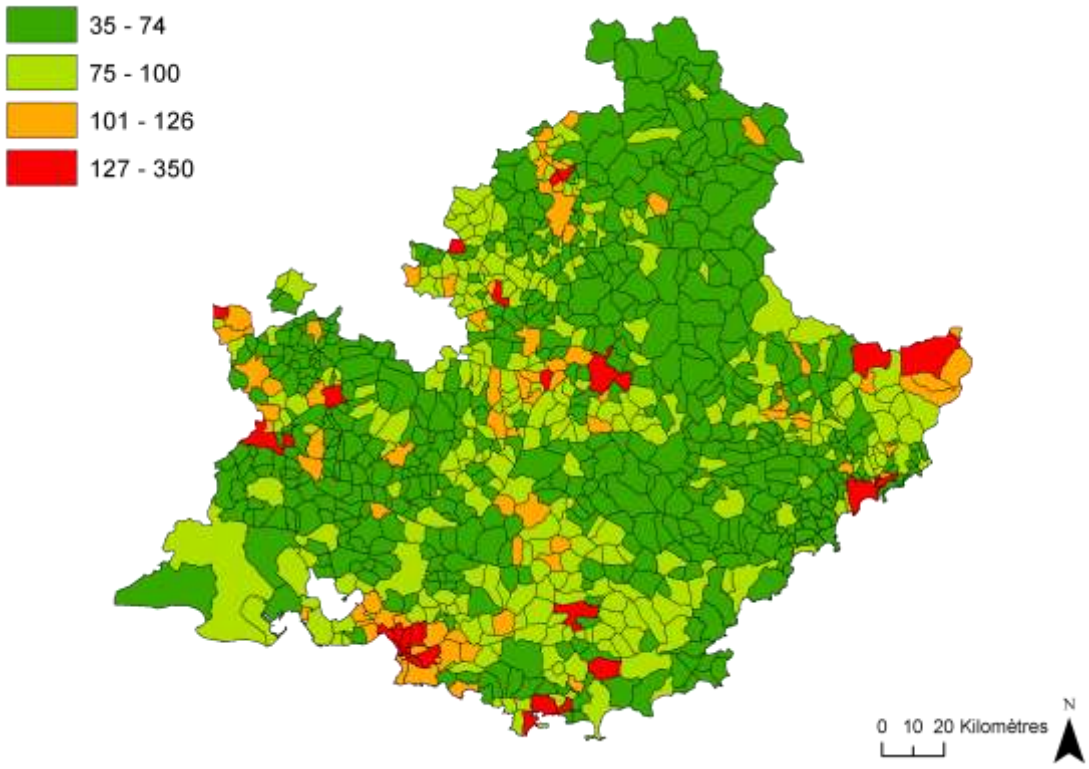
SMR lissés (lissage spatial BYM – matrice de contigüité Rook 1)



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 1. SMR* (base 100) de l'ALD 23 en région Paca au 31 décembre 2011 à l'échelle des communes

SMR lissés (lissage spatial BYM – méthode de discrétisation « écart-type* »)

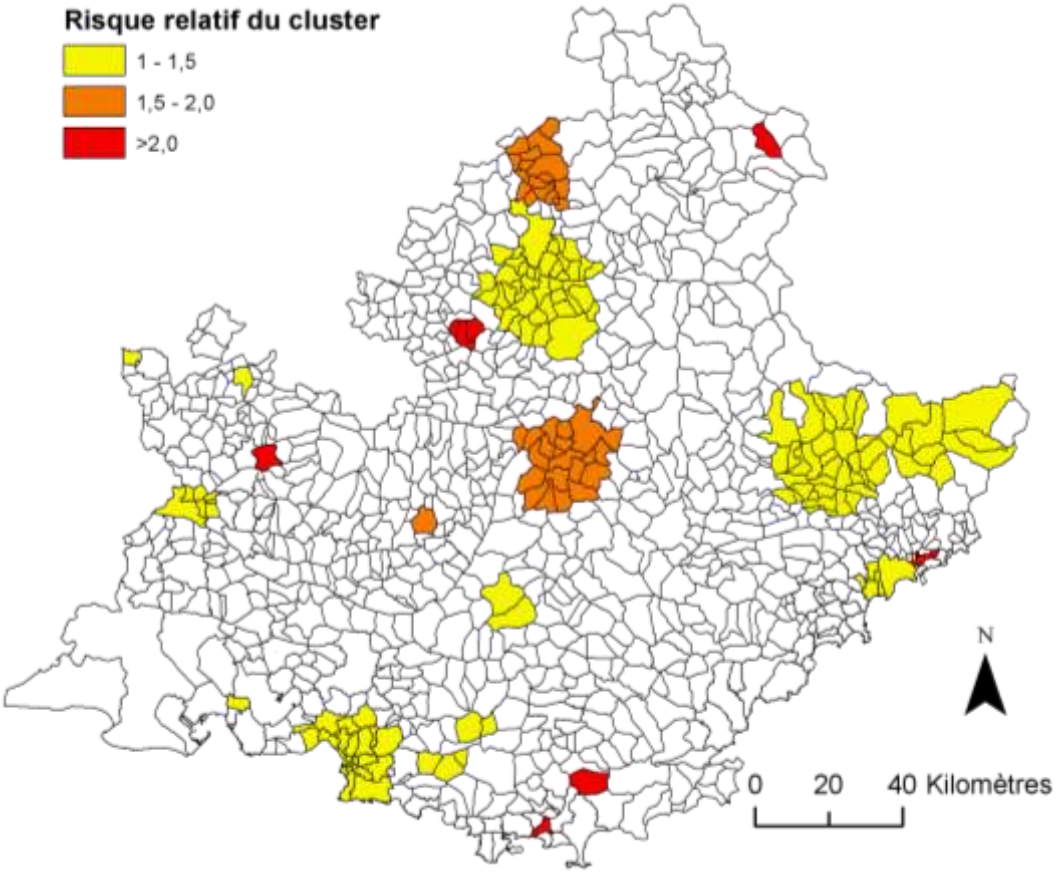


Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : <100-écart-type, [100-écart-type ; 100], [100 ; 100+écart-type], > 100+écart-type.*
*** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 2. SMR (base 100) de l'ALD 23 en région Paca au 31 décembre 2011 à l'échelle des communes**

- Tendances locales au clustering

L'analyse de cluster met clairement en évidence un risque plus élevé d'ALD 23 dans la commune de Marseille et certaines communes limitrophes. L'ajustement sur l'offre de soin en psychiatrie et le niveau de précarité socio-économique des communes fait diminuer le risque relatif de ce cluster qui reste néanmoins le cluster principal (i.e. le plus probable) du fait d'une forte puissance statistique (densité de population très élevée dans les unités géographiques qui composent ce cluster).
Après ajustement sur l'offre de soin en psychiatrie et le niveau de précarité socio-économique des communes (figure 3), on détecte aussi un cluster situé dans le département des Alpes-de-Haute- provence autour de la commune de Dignes-les-Bains avec un risque d'ALD 23 plus élevé que dans le reste de la région.
Des analyses de sensibilité sur les paramètres d'entrée du modèle (taille et forme des clusters, construction des variables d'ajustement) pourraient être envisagées afin d'étayer ces résultats.



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 3. Clusters d'ALD 23 en région Paca au 31 décembre 2011 à l'échelle des communes – Statistiques de scan – Modèle ajusté sur la capacité totale en hospitalisation complète des grandes structures psychiatriques et l'indice de désavantage social

Annexe 2. ALD 8 E11 – Diabète de type 2

Le taux de prévalence brut de l'ALD 8 E11 (diabète de type 2) est de 3 107 pour 100 000 habitants en région Paca.

- Lissage

Tableau 1. Synthèse des résultats des modèles de lissage

	SMR Min-Max	SMR Intervalle Interquartile (Q3-Q1)	AIC
Données non lissées	0-210	39	
Lissage non spatial	42-207	19	921
Lissage spatial BYM	45-207	22	770
Lissage spatial GLMM	42-207	19	933

NA : non applicable pour les données non lissées.

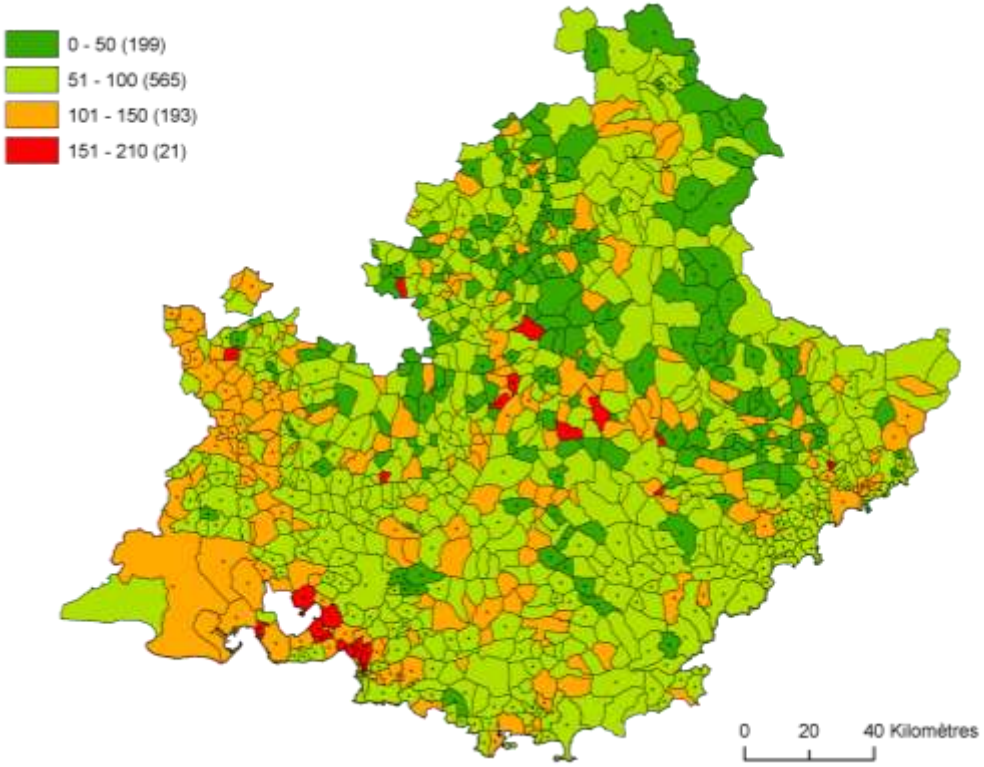
En se basant sur le critère d'AIC, le modèle de lissage spatial BYM est retenu comme « meilleur modèle » (tableau 1). L'effet du lissage sur la représentation cartographique de la prévalence de l'ALD 8 E11 est modéré (figure 1). Le niveau communal apparait adapté à la représentation cartographique de la prévalence de cet indicateur sanitaire.
Une version de la carte avec une méthode de discrétisation mieux adaptée à la distribution des SMR lissés est présentée en figure 2.

- Tendance globale au clustering

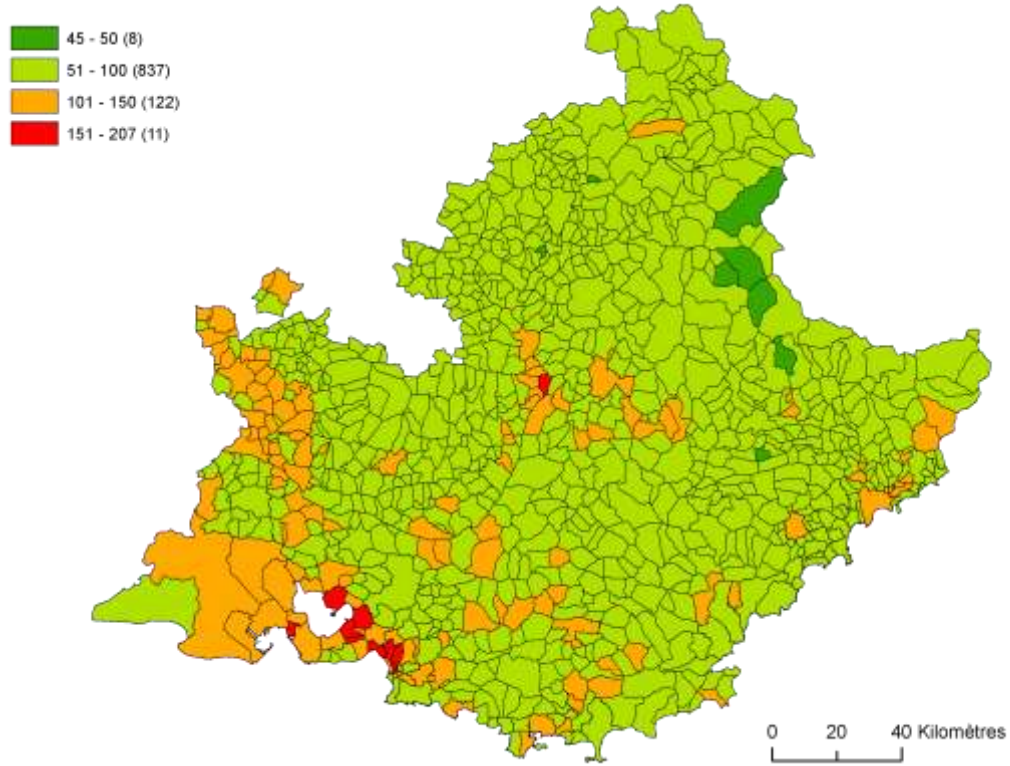
Le test de Potthoff et Whittinghill permet de conclure à une surdispersion des données ($p<0,01$). Le SMR de d'ALD E11 n'est pas distribué de manière homogène sur l'ensemble du territoire.

Afin de contrôler l'instabilité des estimations des SMR, le calcul de l'indice de Moran est réalisé sur des données lissées par une méthode non spatiale (modèle poisson lognormal avec effet aléatoire). En se basant sur une structure de contigüité, l'indice de Moran est significativement supérieur à zéro ($I = 0,35$ – $p<0,01$). La modification de la structure de voisinage (distance entre les centroïdes des communes) fait légèrement diminuer la valeur de l'indice de Moran ($I = 0,30$ – $p<0,01$).

SMR non lissés



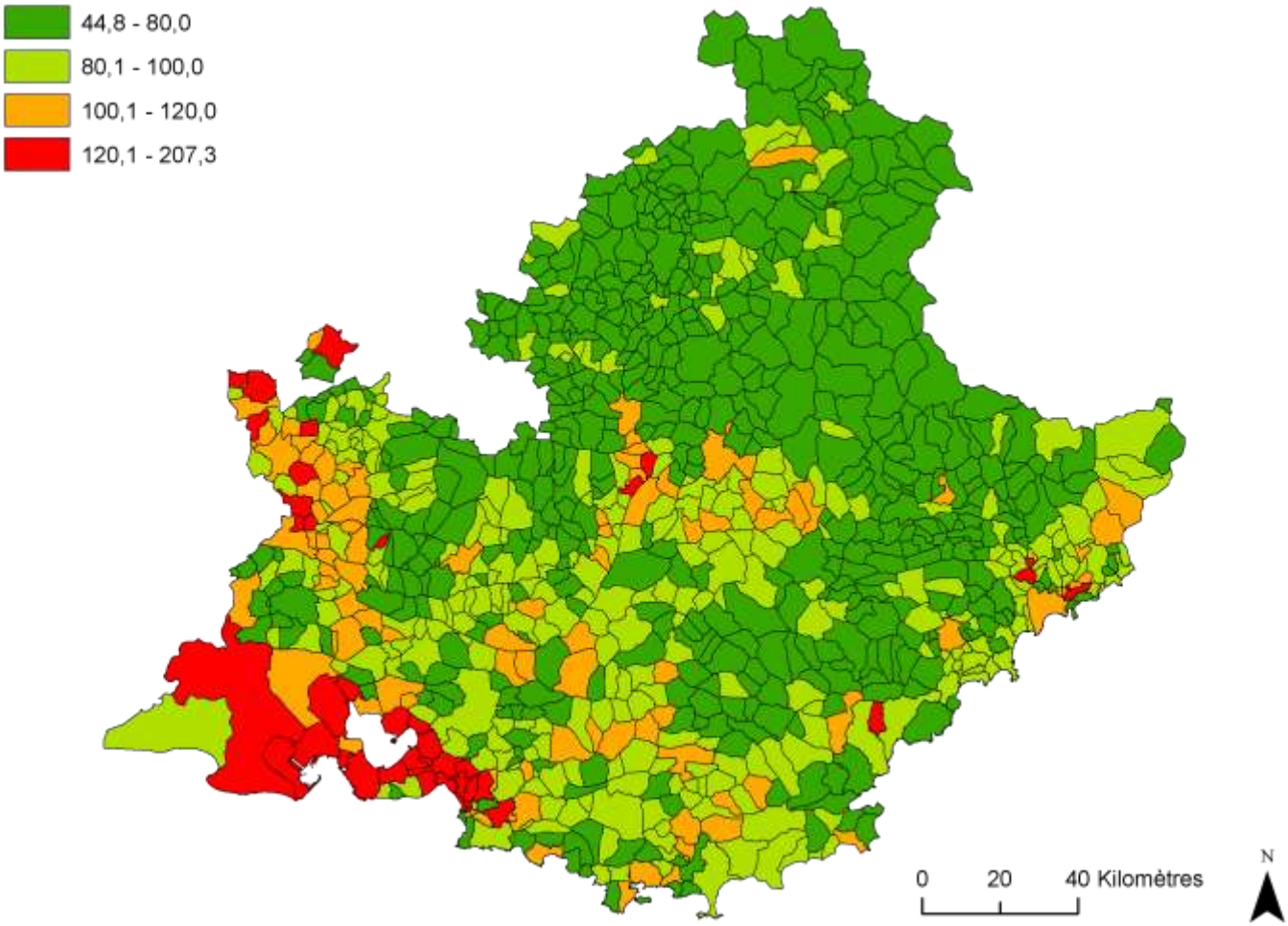
SMR lissés (lissage spatial BYM – matrice de contigüité Rook 1)



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 1. SMR* (base 100) de l'ALD 8 E11 en région Paca au 31 décembre 2011 à l'échelle des communes

SMR lissés (lissage spatial BYM –méthode de discrétisation « écart-type »*)



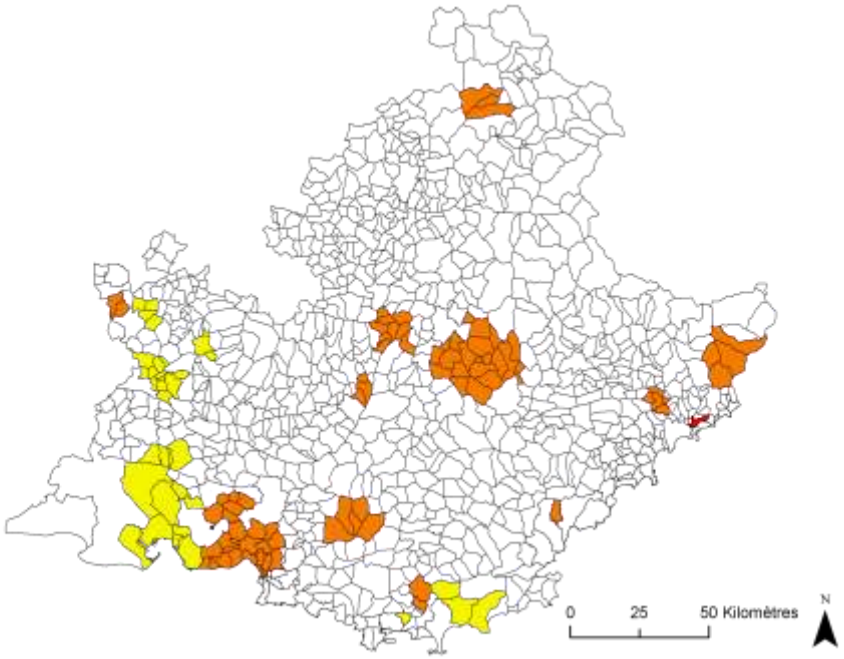
Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : <100-écart-type, [100-écart-type ; 100], [100 ; 100+écart-type], > 100+écart-type.*
*** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 2. SMR (base 100) de l'ALD 8 E11 en région Paca au 31 décembre 2011 à l'échelle des communes**

- Tendances locales au clustering

L'analyse locale de clusters met en évidence la présence d'un cluster principal de communes avec un niveau de risque significativement plus élevé que dans le reste de la région (RR>1,5). Ce cluster est composé des arrondissements et des communes situés au nord de la ville de Marseille (figure 3). Comme pour l'ALD 23, ce cluster est identifié comme le cluster principal (i.e. celui avec la p_valeur la plus faible) du fait d'une densité de population très importante dans les unités géographiques qui le composent (plus forte puissance statistique que dans le reste de la région).
Dans les départements des Bouches- du-Rhône et du Vaucluse, l'ajustement sur le niveau de précarité socio-économique des populations a pour conséquence la disparition de certains clusters et la baisse des niveaux de risque de diabète observés dans la plupart des clusters. Dans les autres départements de la région, cet ajustement fait apparaitre de nouveaux clusters et a tendance à augmenter les niveaux de risque observé dans le modèle non ajusté.

Statistiques de scan : clusters de l'ALD 8 E11



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 3. Clusters d'ALD 8 E11 en région Paca au 31 décembre 2011 à l'échelle des communes – Statistiques de scan - Modèle ajusté sur l'indice de désavantage social

Annexe 3. ALD 5 et 13 - Insuffisance cardiaque grave, troubles du rythme graves, cardiopathies valvulaires graves, cardiopathies congénitales graves et maladie coronarienne

Le taux de prévalence brut des ALD 5 et 13 est de 3 499 pour 100 000 habitants en région Paca.

- Lissage

Tableau 1. Synthèse des résultats des modèles de lissage

	SMR		
	Min-Max	Intervalle interquartile (Q3-Q1)	AIC
Données non lissées	0-241	34	
Lissage non spatial	56-144	11	324
Lissage spatial BYM	59-146	16	205
Lissage spatial GLMM	58-144	10	318

NA : non applicable pour les données non lissées.

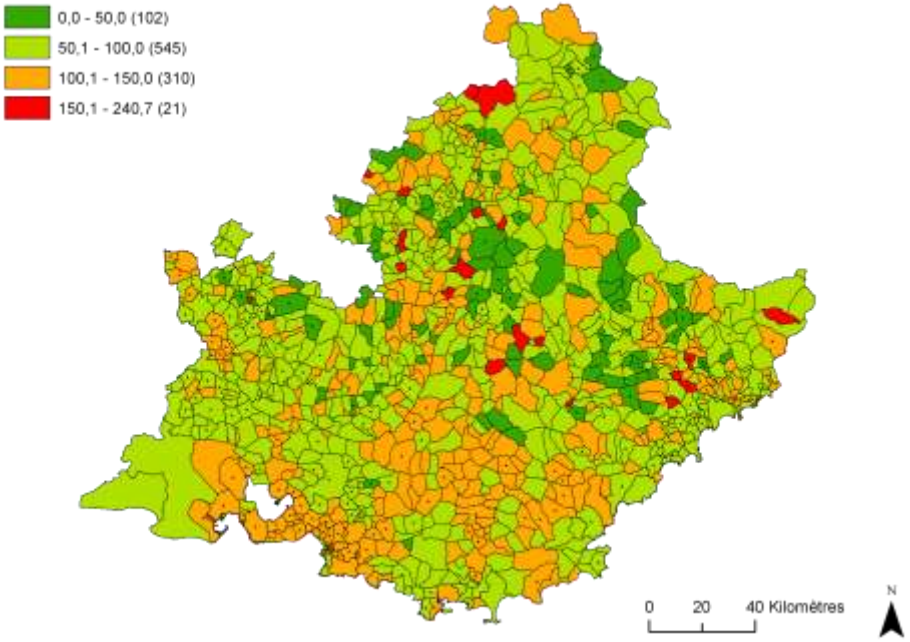
En se basant sur le critère d'AIC, le modèle de lissage spatial BYM est retenu comme « meilleur » modèle (tableau 1). L'effet du lissage sur la représentation cartographique de la prévalence des ALD 5 et 13 est assez important (figure 1). Une version de la carte avec une méthode de discrétisation mieux adaptée à la distribution des SMR lissés est présentée en figure 2.

- Tendance globale au clustering

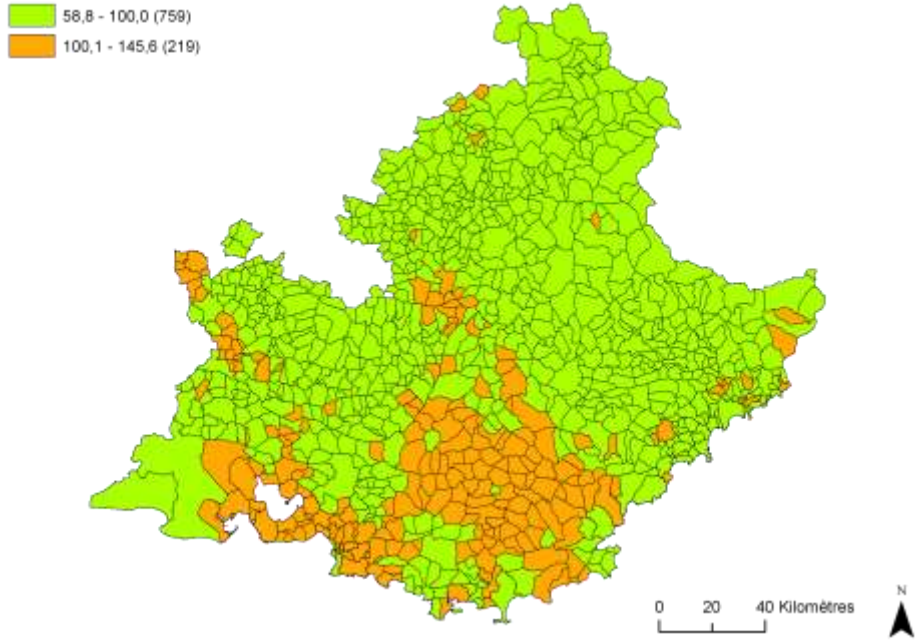
Le test de Potthoff et Whittinghill ne permet pas de conclure à une surdispersion des données ($p = 0,07$) sur l'ensemble de la région.

Afin de contrôler l'instabilité des estimations des SMR, le calcul de l'indice de Moran est réalisé sur des données lissées par une méthode non spatiale (modèle poisson lognormal avec effet aléatoire). En se basant sur une structure de contigüité, l'indice de Moran est significativement supérieur à zéro ($I = 0,32$ – $p<0,01$). La modification de la structure de voisinage (distance entre les centroïdes des communes) fait légèrement diminuer la valeur de l'indice de Moran ($I = 0,25$ – $p<0,01$).On peut donc conclure à la présence d'une ACS positive (i.e. les communes adjacentes ont tendances à avoir des valeurs de SMR similaires).

SMR non lissés



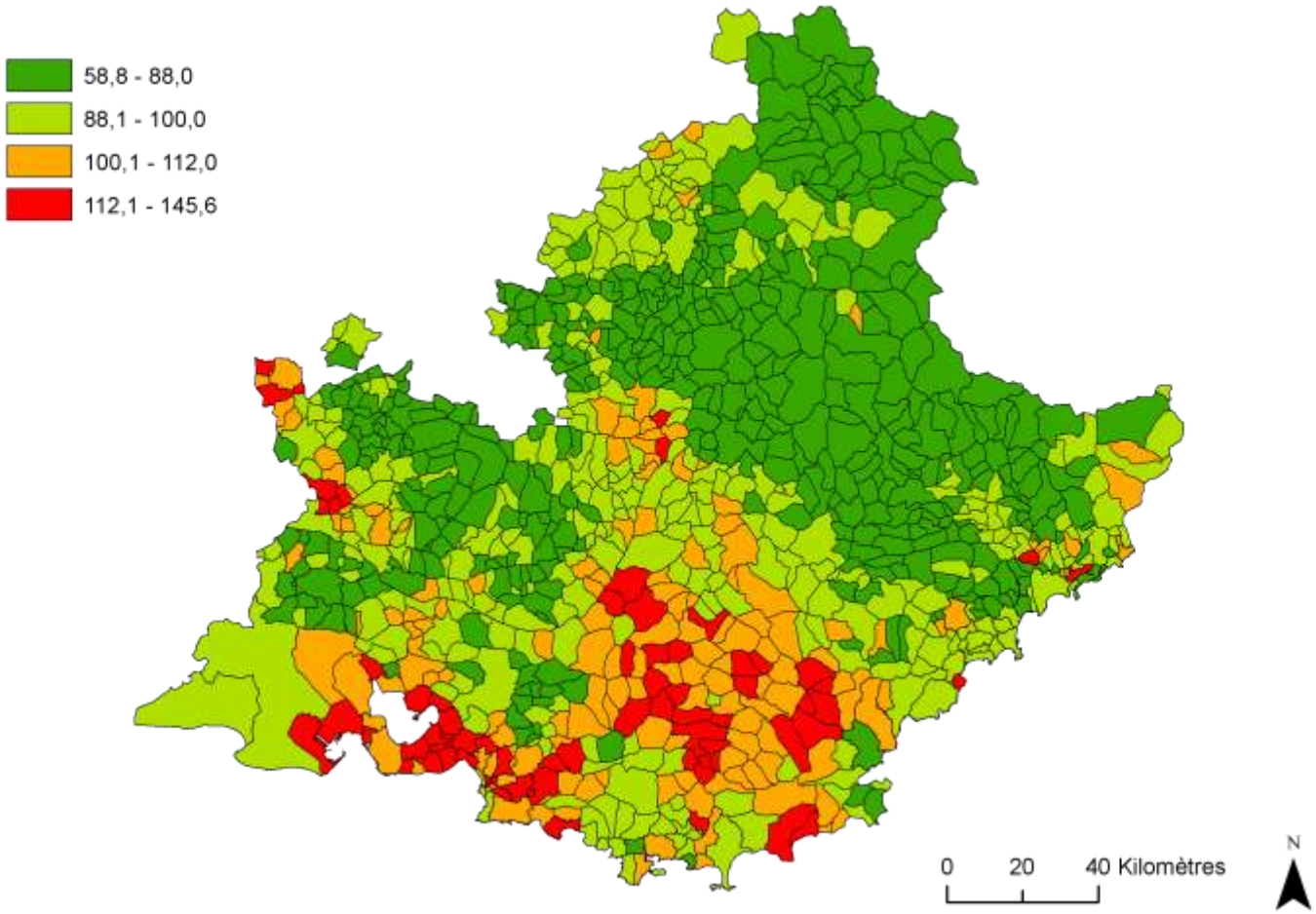
SMR lissés (lissage spatial BYM)



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 1. SMR* (base 100) des ALD 5 et 13 en région Paca au 31 décembre 2011 à l'échelle des communes

SMR lissés (lissage spatial BYM – méthode de discrétisation « écart-type* »)



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : <100-écart-type, [100-écart-type ; 100], [100 ; 100+écart-type], > 100+écart-type.*
*** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 2. SMR (base 100) des ALD 5 et 13 en région Paca au 31 décembre 2011 à l'échelle des communes**

Annexe 4. ALD 30 C50 - Tumeurs malignes du sein chez la femme

Le taux de prévalence brut de l'ALD 30 C50 chez la femme est de 1 977 pour 100 000 habitants en région Paca.

- Lissage

Tableau 1. Synthèse des résultats des modèles de lissage

	SMR		
	Min-Max	Ecart interquartile (Q3-Q1)	AIC
Données non lissées	0-456	50	
Lissage non spatial	72-124	4	878
Lissage spatial BYM	68-127	12	886
Lissage spatial GLMM	69-126	5	882

NA : non applicable pour les données non lissées.

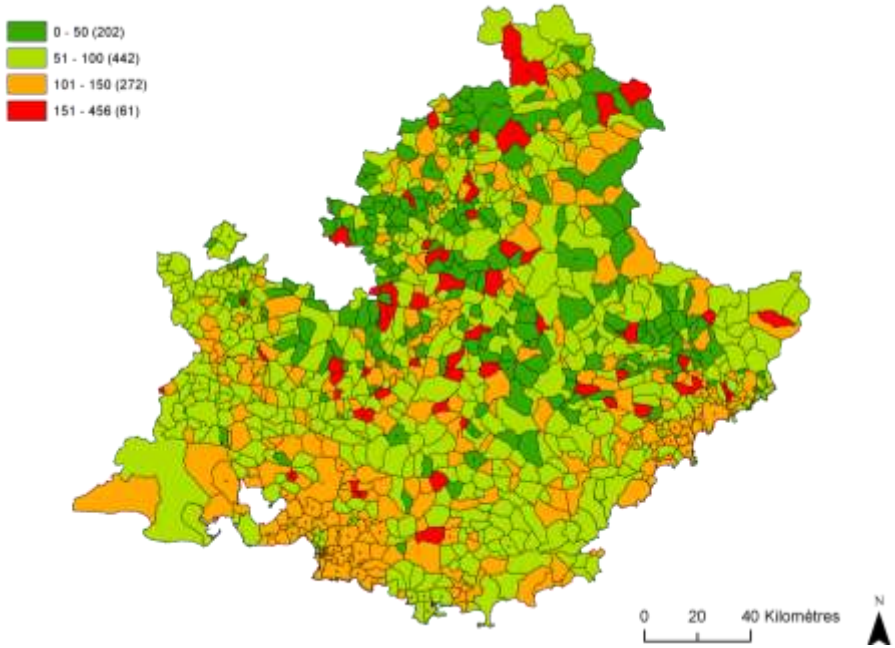
En se basant sur le critère d'AIC, le modèle de lissage non spatial est retenu comme « meilleur » modèle (tableau 1). Il apparait cependant que l'échelle communale ne soit pas adaptée pour représenter cet indicateur sanitaire. Pour un grand nombre de communes, le poids du lissage est trop important (figure 1).
Afin de donner plus de poids aux données, il apparait nécessaire d'augmenter l'échelle d'analyse. Une démarche au niveau des espaces de santé de proximité (ESP) a été testée. Toutefois, il apparait que les SMR non lissés au niveau ESP ne sont pas affectés par le problème de l'instabilité liée aux petits effectifs. Un lissage des données n'est donc pas nécessaire à ce niveau géographique (figure 2).

- Tendance globale au clustering

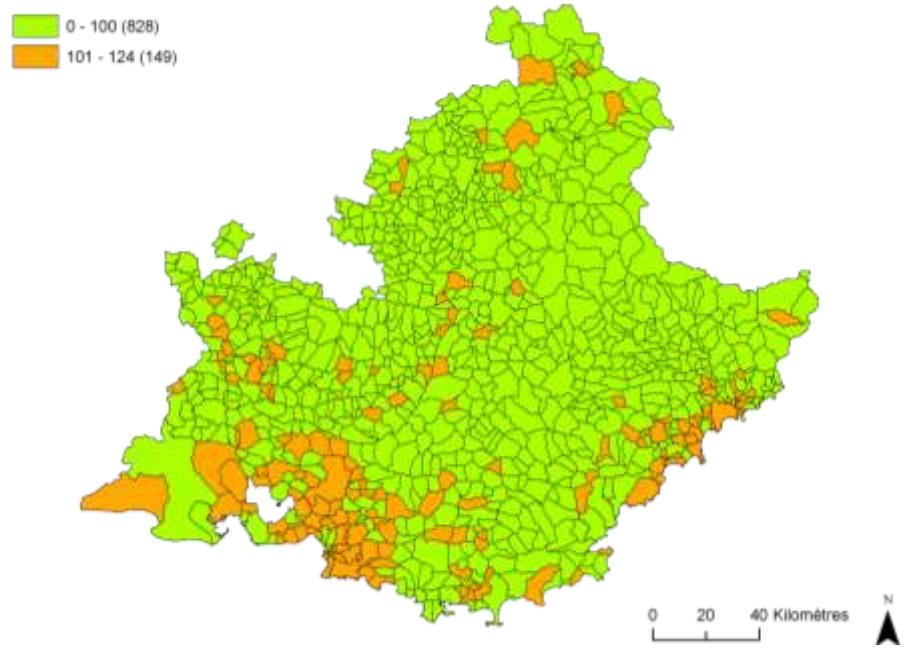
Le test de Potthoff-Whittinghill sur l'ensemble des SMR souligne une hétérogénéité spatiale non statistiquement significative ($p = 0,36$).

Afin de contrôler l'instabilité des estimations des SMR, le calcul de l'indice de Moran est réalisé sur des données lissées par une méthode non spatiale (modèle poisson lognormal avec effet aléatoire). En se basant sur une structure de contigüité, l'indice de Moran est significativement supérieur à zéro ($I = 0,23$ – $p<0,01$). La modification de la structure de voisinage (distance entre les centroïdes des communes) fait diminuer la valeur de l'indice de Moran ($I = 0,16$) mais ce dernier reste significativement supérieur à zéro ($p<0,01$).
On peut donc conclure à la présence d'une ACS positive modérée.

SMR non lissés



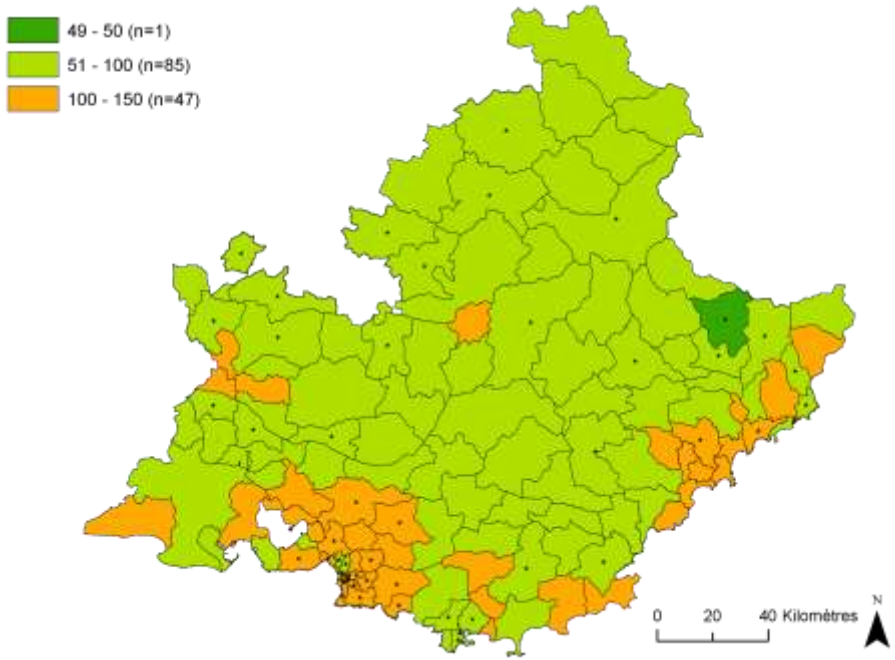
SMR lissés (lissage non spatial)



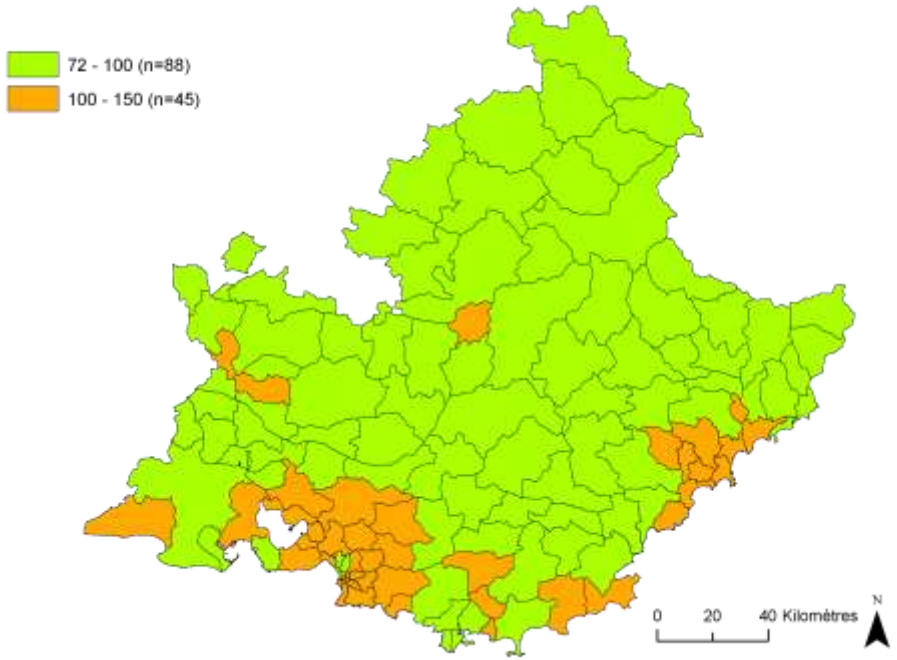
Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+). La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 1. SMR* (base 100) de l'ALD 30 C50 en région Paca au 31 décembre 2011 à l'échelle des communes

SMR non lissés



SMR lissés (lissage spatial BYM)



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
Les communes marquées par une étoile ont un SMR statistiquement plus élevé ou plus faible que la valeur de référence (100). Test du khi-deux, seuil de significativité 5 %.
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+). La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 2. SMR* (base 100) de l'ALD 30 C50 en région Paca au 31 décembre 2011 à l'échelle des espaces de santé de proximité

Annexe 5. ALD 30 C34 - Tumeurs malignes des bronches et du poumon

Le taux de prévalence brut de l'ALD 30 C34 est de 124 pour 100 000 habitants en région Paca.

- Lissage

Tableau 1. Synthèse des résultats des modèles de lissage

	SMR		
	Min-Max	Intervalle interquartile (Q3-Q1)	AIC
Données non lissées	0-1448	113	NA
Lissage non spatial	85-113	1	2569
Lissage spatial BYM	79-114	9	2661
Lissage spatial GLMM	79-118	2	2576

NA : non applicable pour les données non lissées.

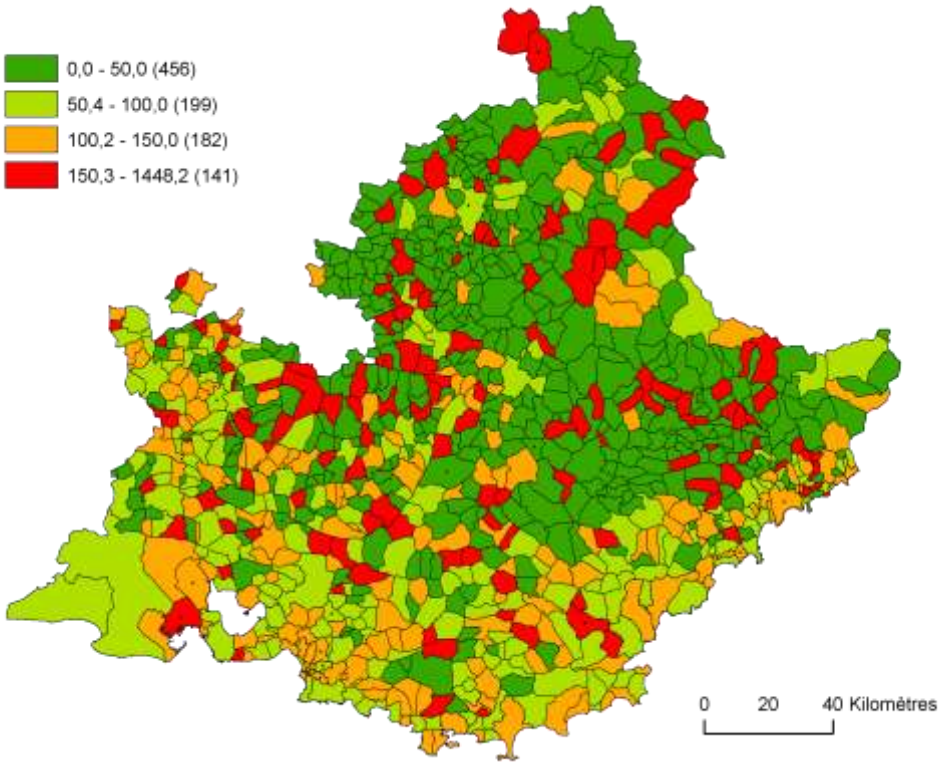
En se basant sur le critère d'AIC, le modèle de lissage non spatial est retenu comme « meilleur » modèle (tableau 1). Il apparait cependant que l'échelle communale ne soit pas adaptée pour représenter cet indicateur sanitaire. Pour un grand nombre de communes, le poids du lissage est trop important (figure 1). Afin de donner plus de poids aux données, il apparait nécessaire d'augmenter l'échelle d'analyse. Une démarche au niveau des espaces de santé de proximité (ESP) est présentée en figure 2.

- Tendance globale au clustering

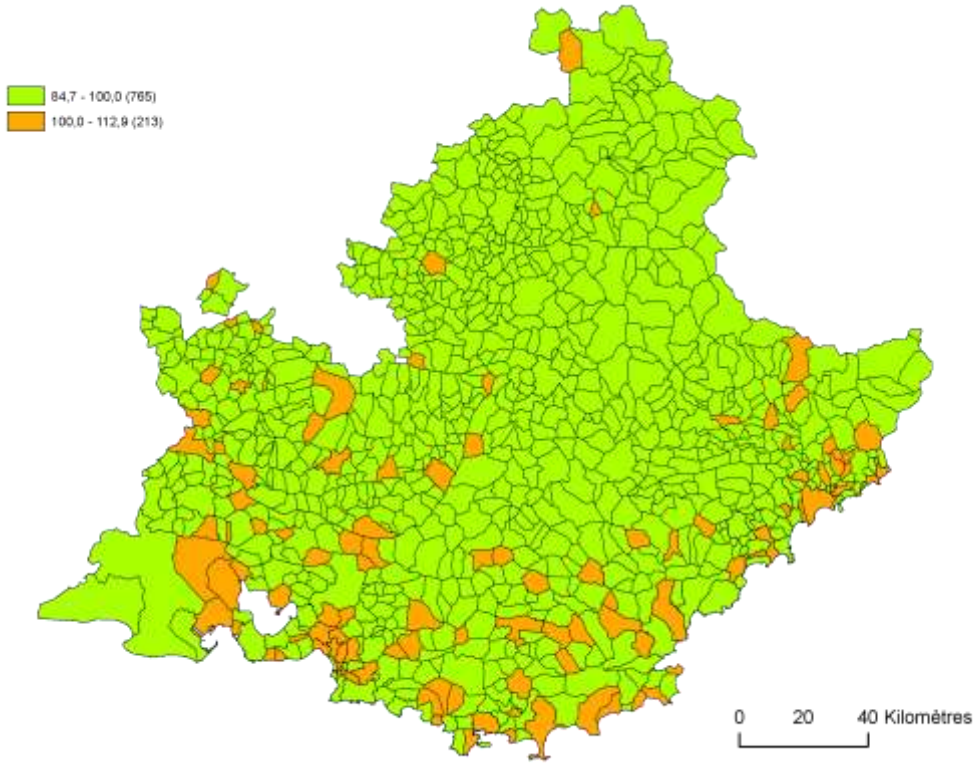
Le test de Potthoff-Whittinghill sur l'ensemble des SMR souligne une hétérogénéité spatiale non statistiquement significative (p = 0,48).

Afin de contrôler l'instabilité des estimations des SMR, le calcul de l'indice de Moran est réalisé sur des données lissées par une méthode non spatiale (modèle poisson lognormal avec effet aléatoire). En se basant sur une structure de contigüité, l'indice de Moran n'est pas significativement différent de zéro (I = 0,02 – p = 0,30). En utilisant une matrice de distance, l'ACS reste faible (I = 0,02) mais devient statistiquement significative (p<0,01).
Au vu de ces résultats, on peut conclure à l'absence totale d'ACS au niveau communal pour cet indicateur.

SMR non lissés



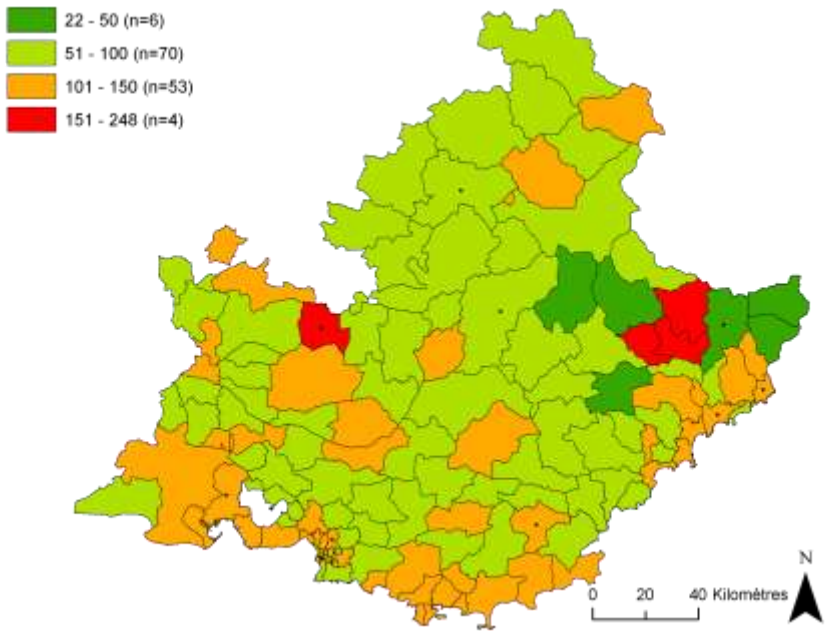
SMR lissés (lissage non spatial)



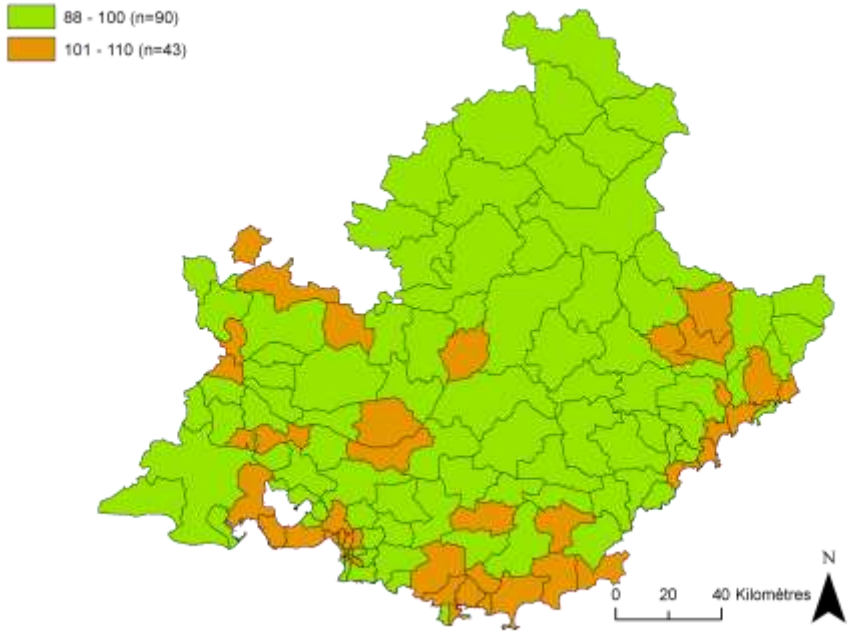
Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 1. SMR* (base 100) de l'ALD 30 C34 en région Paca au 31 décembre 2011 à l'échelle des communes

SMR non lissés



SMR lissés (lissage non spatial)



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
Les communes marquées par une étoile ont un SMR statistiquement plus élevé ou plus faible que la valeur de référence (100). Test du khi-deux, seuil de significativité 5 %.
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 2. SMR* (base 100) de l'ALD 30 C34 en région Paca au 31 décembre 2011 à l'échelle des espaces de santé de proximité

Annexe 6. ALD 30 C18 - Tumeurs malignes du côlon

Le taux de prévalence brut de l'ALD 30 C18 est de 266 pour 100 000 habitants en région Paca.

- Lissage

Tableau 9. Synthèse des résultats des modèles de lissage

	SMR		
	Min-Max	**Intervalle interquartile (Q3-Q1)**	**AIC**
Données non lissées	0-887	124	
Lissage non spatial	89-121	1	1 866
Lissage spatial BYM	85-119	5	1 895
Lissage spatial GLMM	85-129	2	1 864

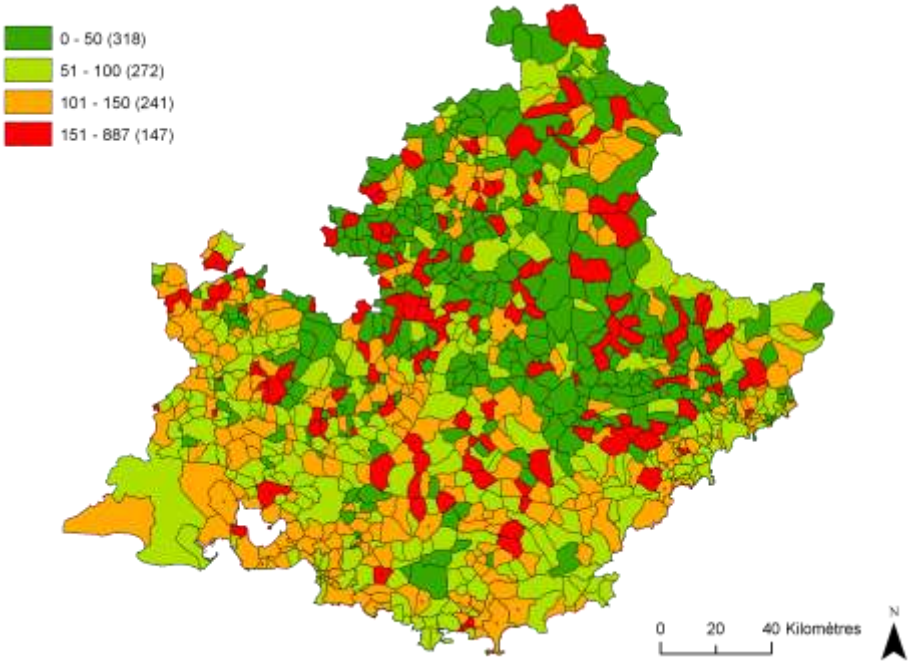
NA : non applicable pour les données non lissées.

En se basant sur le critère d'AIC, le modèle de lissage spatial GLMM est retenu comme « meilleur » modèle (tableau 1). Il apparait cependant que l'échelle communale ne soit pas adaptée pour représenter cet indicateur sanitaire. Pour un grand nombre de communes, le poids du lissage est trop important (figure 1). Afin de donner plus de poids aux données, il apparait nécessaire d'augmenter l'échelle d'analyse. Une démarche au niveau des espaces de santé de proximité (ESP) est présentée en figure 2.

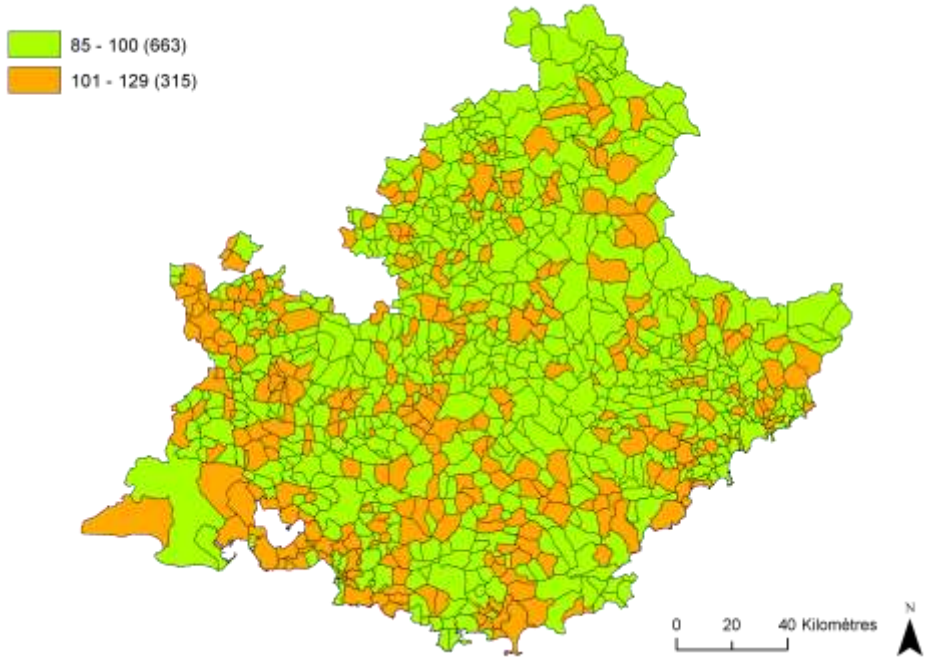
- Tendance globale au clustering

Le test de Potthoff-Whittinghill sur l'ensemble des SMR souligne une hétérogénéité spatiale non statistiquement significative ($p = 0,48$).
Afin de contrôler l'instabilité des estimations des SMR, le calcul de l'indice de Moran est réalisé sur des données lissées par une méthode non spatiale (modèle poisson lognormal avec effet aléatoire). En se basant sur une structure de contigüité, l'indice de Moran est faible mais significativement différent de zéro ($I = 0,06$ – $p<0,01$). En changeant la structure de voisinage par une matrice de distance, l'indice de Moran reste faible ($I = 0,01$) et n'est plus statistiquement significatif ($p = 0,13$).
Au final, avec des valeurs de l'indice de Moran proches de zéro, on ne peut conclure à la présence d'ACS au niveau communal pour cet indicateur.

SMR non lissés



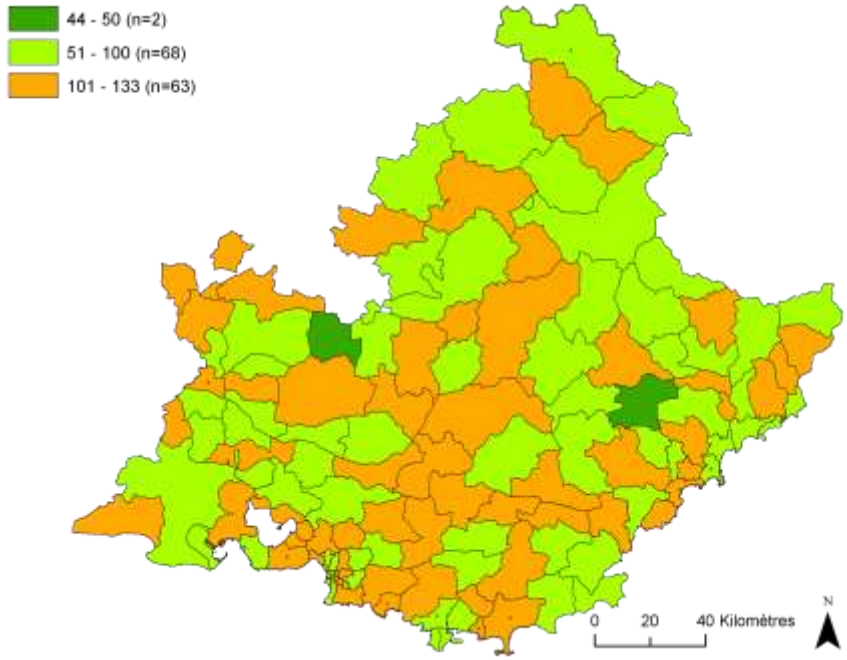
SMR lissés (lissage spatial GLMM)



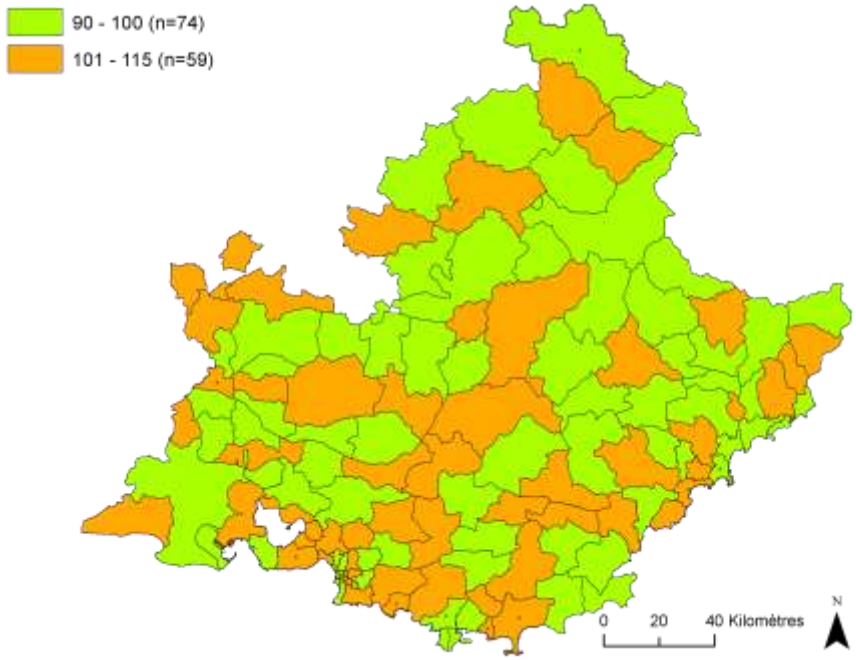
Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 1. SMR* (base 100) de l'ALD 30 C18en région Paca au 31 décembre 2011 à l'échelle des communes

SMR non lissés



SMR lissés (lissage non spatial)



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
Les communes marquées par une étoile ont un SMR statistiquement plus élevé ou plus faible que la valeur de référence (100). Test du khi-deux, seuil de significativité 5 %.
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 2. SMR* (base 100) de l'ALD 30 C18 en région Paca au 31 décembre 2011 à l'échelle des espaces de santé de proximité

Annexe 7. ALD 30 C67 - Tumeurs malignes de la vessie

Le taux de prévalence brut de l'ALD 30 C67 est de 181 pour 100 000 habitants en région Paca.

- Lissage

Tableau 1. Synthèse des résultats des modèles de lissage

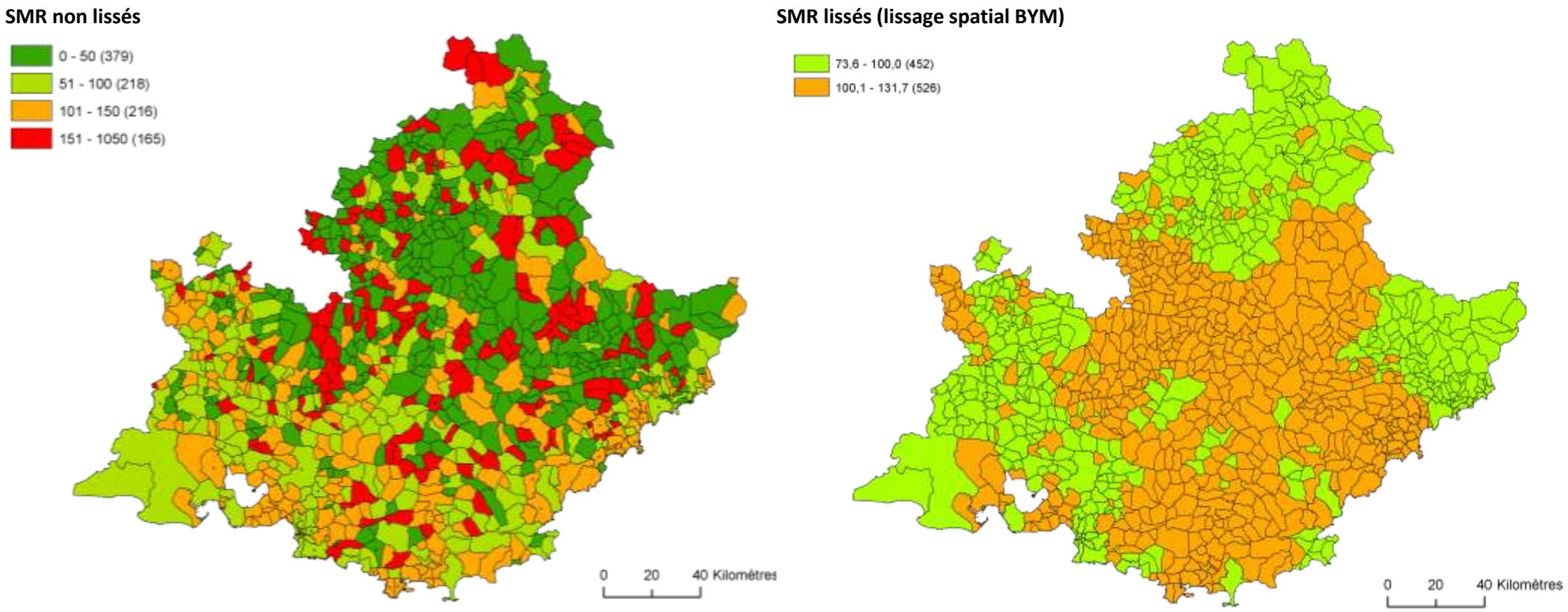
	SMR		
	Min-Max	**Intervalle interquartile (Q3-Q1)**	**AIC**
Données non lissées	0-1047	128	
Lissage non spatial	80-135	4	2317
Lissage spatial BYM	74-132	10	2300
Lissage spatial GLMM	78-141	5	2311

NA : non applicable pour les données non lissées.

En se basant sur le critère d'AIC, le modèle de lissage spatial BYM est retenu comme « meilleur » modèle (tableau 1). Il apparait cependant que l'échelle communale ne soit pas adaptée pour représenter cet indicateur sanitaire. Pour un grand nombre de communes, le poids du lissage est trop important (figure 1). Afin de donner plus de poids aux données, il apparait nécessaire d'augmenter l'échelle d'analyse. Une démarche au niveau des espaces de santé de proximité (ESP) est présentée en figure 2.

- Tendance globale au clustering

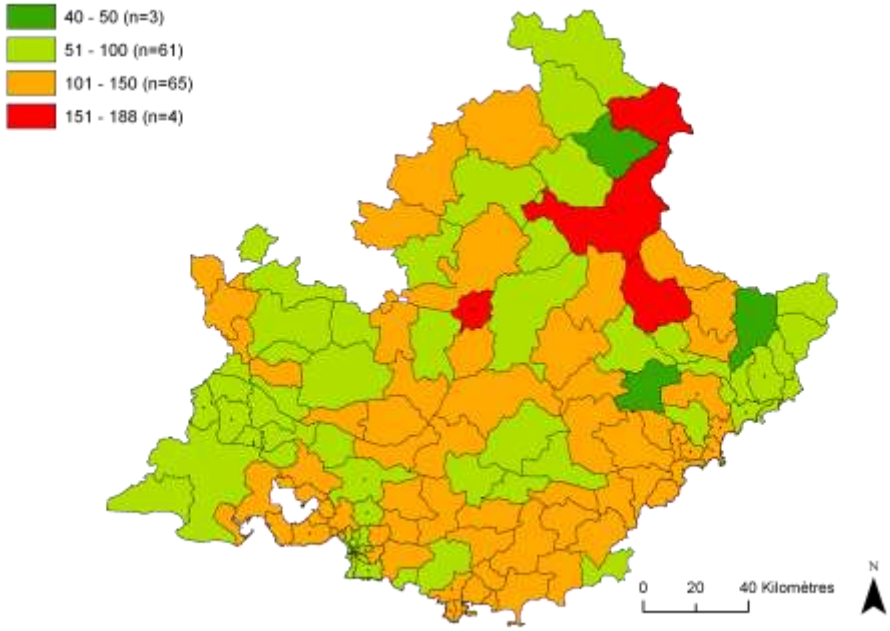
Le test de Potthoff-Whittinghill sur l'ensemble des SMR souligne une hétérogénéité spatiale non statistiquement significative ($p = 0,44$).
Afin de contrôler l'instabilité des estimations des SMR, le calcul de l'indice de Moran est réalisé sur des données lissées par une méthode non spatiale (modèle poisson lognormal avec effet aléatoire). En se basant sur une structure de contigüité, l'indice de Moran est faible mais significativement différent de zéro ($I = 0,09$ – $p<0,01$). En changeant la structure de voisinage par une matrice de distance, l'indice de Moran reste faible ($I = 0,12$ – $p<0,01$).
Au final, on peut considérer l'ACS comme négligeable pour les tumeurs de la vessie.



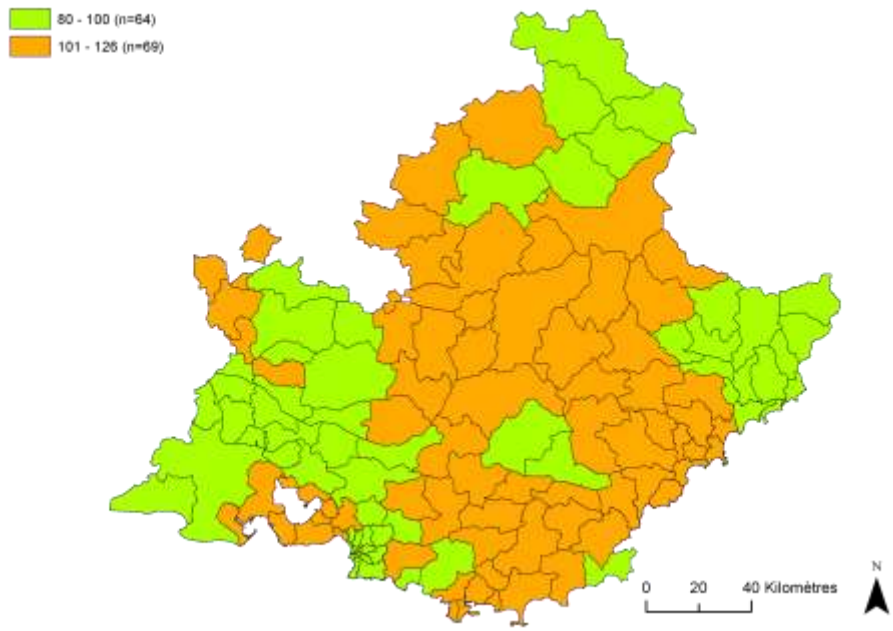
Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF.
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 1. SMR* (base 100) de l'ALD 30 C67en région Paca au 31 décembre 2011 à l'échelle des communes

SMR non lissés



SMR lissés (lissage spatial BYM)



Projection : Lambert 93
Source : services médicaux des régimes de sécurité sociale (RG, RSI, MSA, CNMSS, SNCF).
Exploitation : ORS Paca
Les communes marquées par une étoile ont un SMR statistiquement plus élevé ou plus faible que la valeur de référence (100). Test du khi-deux, seuil de significativité 5 %.
** : Standardisation selon l'âge (0-29/30-49/50-64/65-79/80+) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 2. SMR* (base 100) de l'ALD 30 C67 en région Paca au 31 décembre 2011 à l'échelle des espaces de santé de proximité

Annexe 8. Mortalité prématurée par tumeurs malignes chez les moins de 65 ans

Le taux de prévalence brut de mortalité prématurée par cancers est de 82 pour 100 000 habitants en région Paca.

- Lissage

Tableau 1. Synthèse des résultats des modèles de lissage

	SMR		
	Min-Max	Intervalle interquartile (Q3-Q1)	AIC
Données non lissées	0-809	92	NA
Lissage non spatial	55-139	5	1 817
Lissage spatial BYM	42-156	13	1 800
Lissage spatial GLMM	49-144	7	1 822

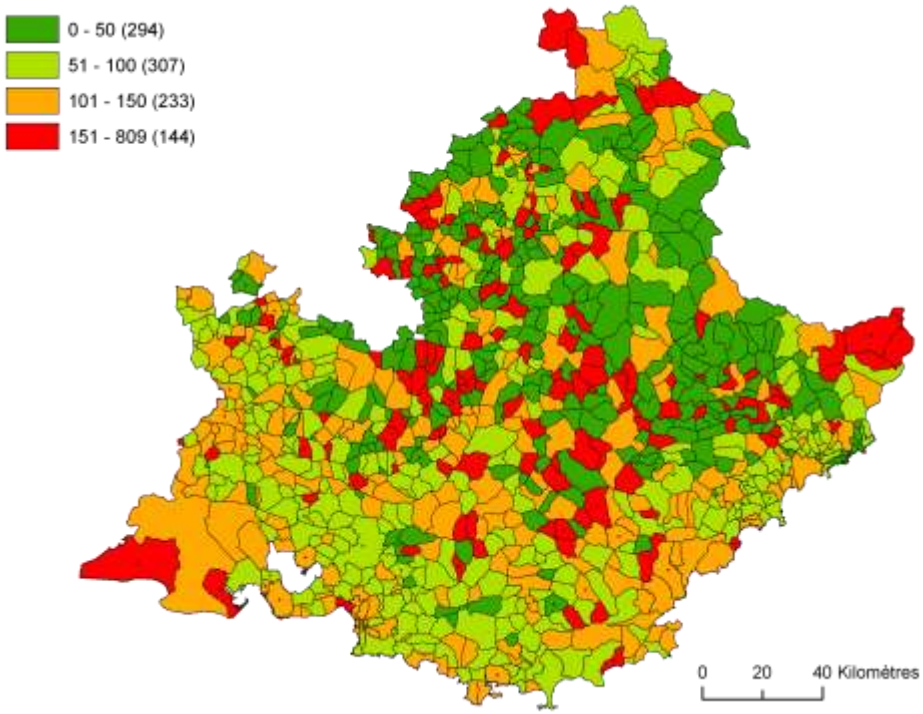
NA : non applicable pour les données non lissées.

En se basant sur le critère d'AIC, le modèle de lissage spatial BYM est retenu comme « meilleur » modèle (tableau 1). Il apparait cependant que l'échelle communale ne soit pas adaptée pour représenter cet indicateur sanitaire. Pour un grand nombre de communes, le poids du lissage est trop important (figure 1). Afin de donner plus de poids aux données, il apparait nécessaire d'augmenter l'échelle d'analyse. Une démarche au niveau des espaces de santé de proximité (ESP) est présentée en figure2.

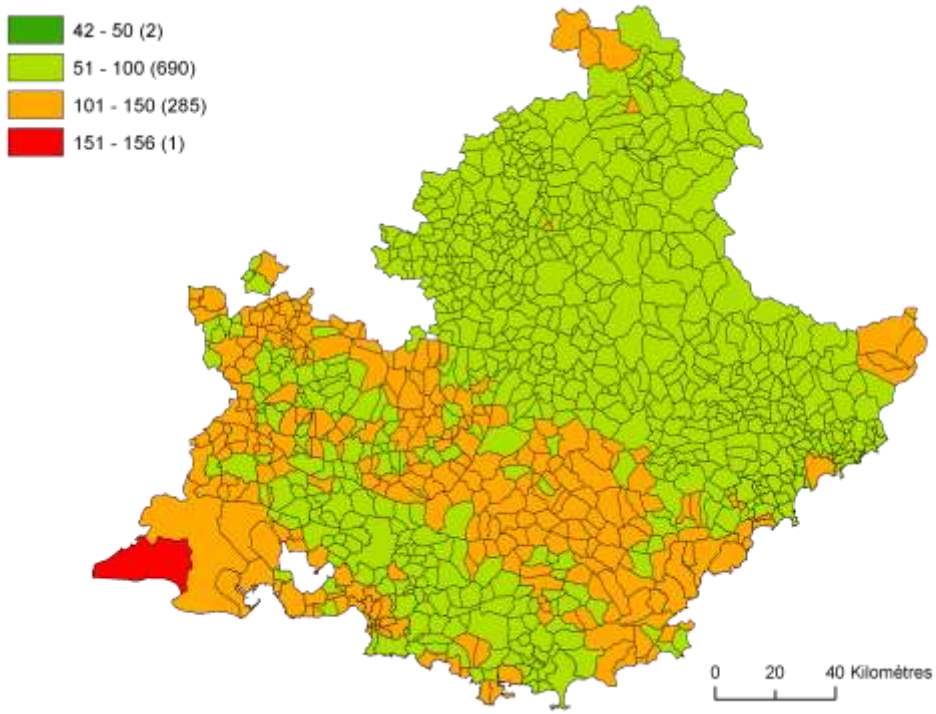
- Tendance globale au clustering

Le test de Potthoff-Whittinghill sur l'ensemble des SMR souligne une hétérogénéité spatiale non statistiquement significative ($p = 0,42$).Afin de contrôler l'instabilité des estimations des SMR, le calcul de l'indice de Moran est réalisé sur des données lissées par une méthode non spatiale (modèle poisson lognormal avec effet aléatoire). En se basant sur une structure de contigüité, l'indice de Moran est faible mais significativement différent de zéro ($I = 0,13$ – $p<0,01$). En changeant la structure de voisinage par une matrice de distance, l'indice de Moran reste faible ($I = 0,15$ – $p<0,01$).
Au vu de ces résultats, la tendance globale au clustering est très faible pour la mortalité prématurée par cancer.

SMR non lissés



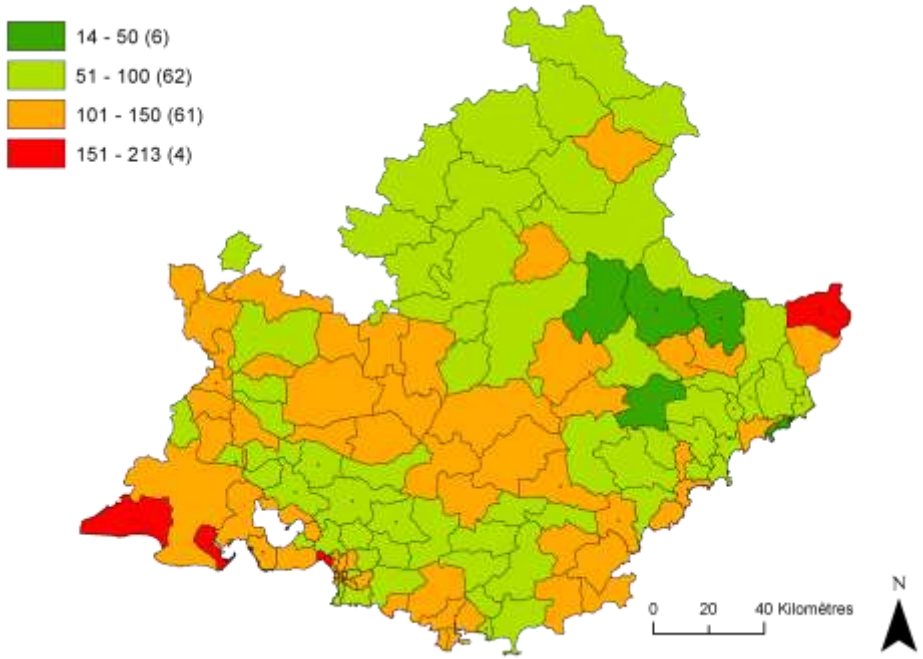
SMR lissés (lissage non spatial)



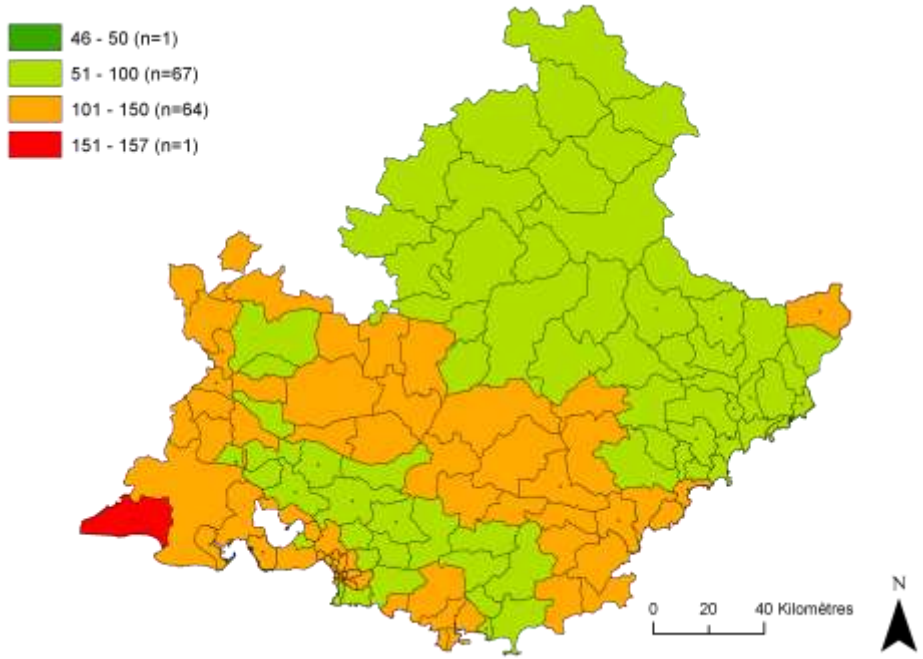
Projection : Lambert 93
Source : CépiDC
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 1. SMR* (base 100) de mortalité prématurée par cancers en région Paca entre 2006 et 2011 à l'échelle des communes

SMR non lissés



SMR lissés (lissage spatial BYM)



Projection : Lambert 93
Source : CépiDC
Exploitation : ORS Paca
** : Standardisation selon l'âge (0-29/30-49/50-64) et le sexe. La prévalence de référence est calculée sur l'ensemble de la population de la région Paca*

Figure 2. SMR* (base 100) de mortalité prématurée par cancers en région Paca au 31 décembre 2011 à l'échelle des espaces de santé de proximité