

Le 1^{er} juillet 2013

Maxime Beauchamp

LCSQA/INERIS – 03.44.55.61.32

maxime.beauchamp@ineris.fr

Laure Malherbe

LCSQA/INERIS – 03.44.55.62.18

laure.malherbe@ineris.fr

Réf. : DRC-13-126737-09252A

OBJET : Cartographie nationale des dépassements de valeurs seuils réglementaires

Cette note a pour objet d'expliciter les derniers travaux réalisés par l'INERIS en matière de cartographie nationale de dépassement de valeur seuil et d'indiquer les développements en cours.

Contexte

Pour tout dépassement de valeur limite ou de valeur cible constaté sur une station de mesure, la législation européenne et française impose de définir les zones touchées par le dépassement et d'estimer les populations exposées.

Les travaux synthétisés ci-après reprennent les éléments théoriques qui ont été développés pour les cartographies de représentativité en concentration et de dépassement de seuil à l'échelle locale dans le rapport LCSQA 2010¹ ainsi que dans la note méthodologique 2011² qui lui fait suite. Ces éléments permettent d'établir des cartes de probabilité de dépassement d'une valeur limite (ex : 40 $\mu\text{g.m}^{-3}$ en moyenne annuelle pour le NO_2 , 50 $\mu\text{g.m}^{-3}$ pour les PM_{10} en moyenne journalière) intégrant l'incertitude des cartes analysées qui combinent les observations aux sites de mesure et les données de modélisation du modèle de chimie-transport CHIMERE. Comme il est illustré dans la première partie, ces cartes de probabilité fournissent une information sur les surfaces en dépassement après application d'un seuil.

¹ Rapport LCSQA : Beauchamp M., Malherbe L., Létinois L., 2011. Application de méthodes géostatistiques pour la détermination de zones de représentativité en concentration et la cartographie des dépassements de seuils, www.lcsqa.org.

² Note méthodologique : Beauchamp M., Malherbe L., Létinois L., 2011. Cartographie du NO_2 à l'échelle locale, Représentativité des stations, Dépassements de seuils. www.lcsqa.org.

Afin de compléter ces travaux, de nouvelles techniques sont mises au point pour évaluer les dépassements de valeurs limites intégrant un cumul d'informations sur plusieurs jours : les résultats pour la valeur limite annuelle PM_{10} fixée à 35 jours de dépassement du seuil journalier dans l'année seront présentés dans une deuxième partie. Des développements portant sur l'évaluation des dépassements de l' AOT_{40} sont également en cours à l'INERIS.

La troisième partie de cette note s'attache à l'étape suivante de l'estimation, à savoir le calcul de la population exposée ; les différentes options de calcul y sont présentées. La spatialisation de la population, que cette évaluation présuppose, n'est pas ici abordée : ce problème est traité dans un document séparé³.

³ Létinois L., 2013. Méthodologie de répartition spatiale de la population. www.lcsqa.org.

1. VALEUR LIMITE D'UNE CONCENTRATION MOYENNE

Dans les premiers travaux publiés par le LCSQA, la zone de dépassement d'une valeur seuil est cartographiée de manière simple, en comparant les valeurs de concentration estimées par krigeage (ex : les concentrations moyennes journalières de PM₁₀) au seuil de concentration considéré (ex : 50 µg/m³). Pour prendre en compte l'incertitude d'estimation, ces valeurs sont augmentées d'un terme déduit empiriquement de l'écart-type de krigeage⁴ ou fixé à une constante⁵.

La nouvelle méthodologie d'évaluation des dépassements de valeur limite repose non plus sur un calcul empirique mais sur le calcul d'une probabilité de dépassement. Il est conseillé de lire attentivement les développements méthodologiques présentés en annexe 2. Le modèle probabiliste utilisé est un modèle gaussien, c'est-à-dire que la loi de l'erreur d'estimation des cartes analysées est considérée comme normale, d'espérance égale à la valeur estimée et de variance égale à l'écart-type de krigeage.

Le principal avantage de la nouvelle méthodologie est de fournir une carte de probabilité de dépassement de la valeur limite. Deux risques statistiques sont utilisés pour construire des zones de dépassement, de non-dépassement, et des zones où le modélisateur ne se prononce pas (zones dites « d'incertitude ») :

- le risque α , qui correspond au risque de dépasser dans la zone de non-dépassement : il doit être faible (5%) ;
- le risque β , qui correspond au risque de ne pas dépasser dans la zone de dépassement : il peut être plus élevé (1/3 conventionnellement).

La figure 1 présente les résultats obtenus pour les cartes analysées de NO₂ entre 2005 et 2009 : seul un domaine centré sur l'Île-de-France est présenté car c'est la région la plus concernée par les dépassements de la valeur limite annuelle (40µg.m⁻³) en pollution de fond.

⁴ Malherbe L., Cárdenas G., 2009. Evaluation des zones touchées par les dépassements de valeurs limites. Note méthodologique, version 1. www.lcsqa.org.

⁵ Groupe de travail national « Zones sensibles », 2010. Méthodologie de définition des zones sensibles, www.lcsqa.org.

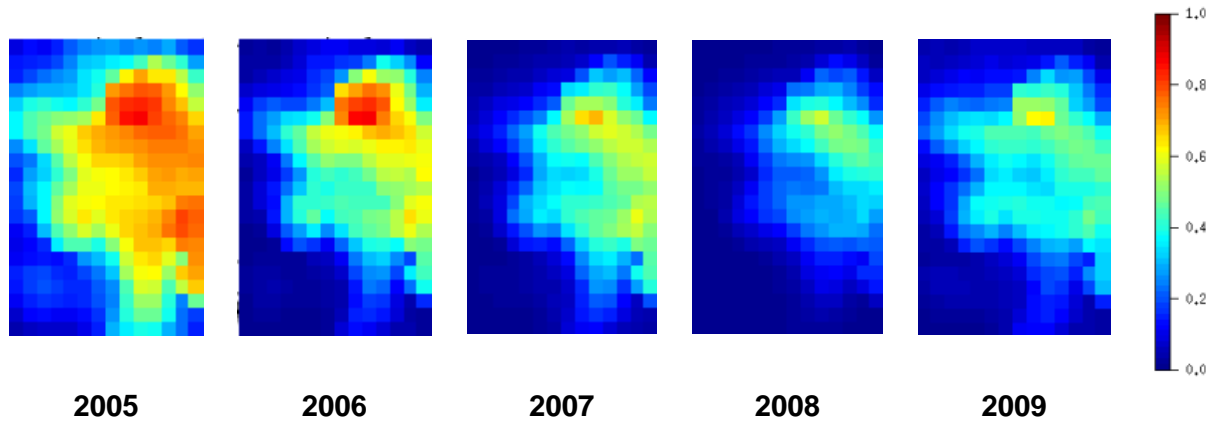


Figure 1 : Probabilité de dépassement de la valeur limite annuelle NO₂ sur l’Ile-de-France entre 2005 et 2009

La figure 2 présente la probabilité que la moyenne journalière de PM₁₀ dépasse la valeur limite de 50µg.m⁻³ le 27 avril 2007 et les zones de dépassement, de non-dépassement et d’incertitude qui en sont déduites.

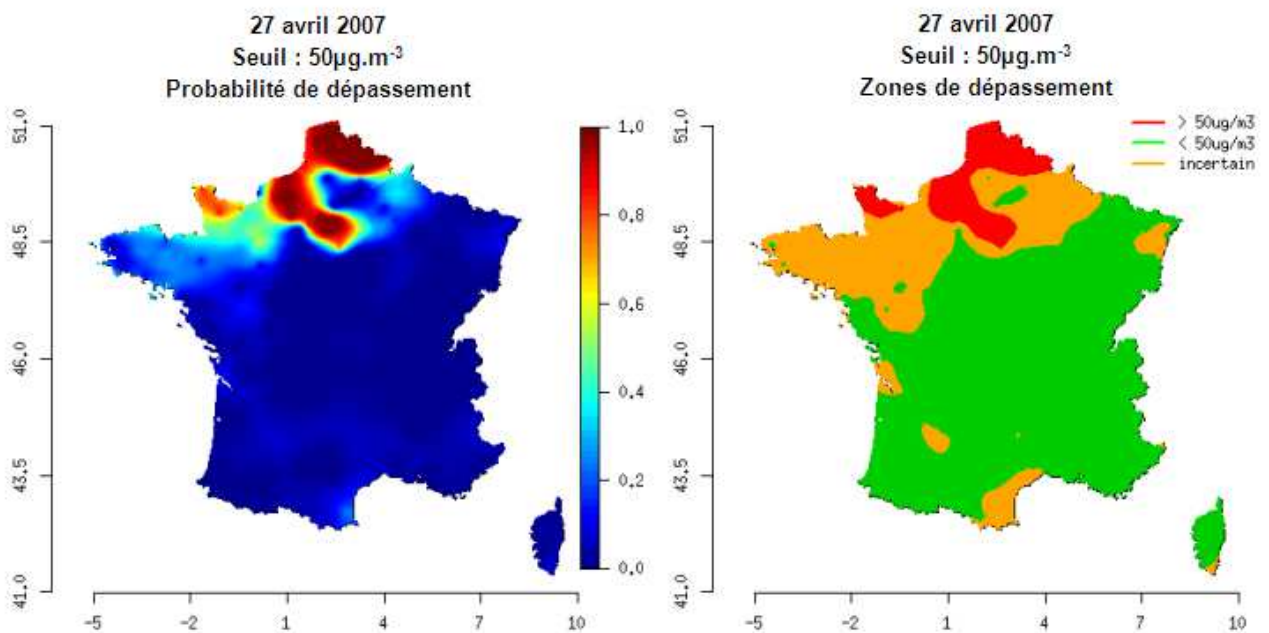


Figure 2 : Probabilité de dépassement de la valeur limite journalière PM₁₀ sur la France le 27 avril 2007 et zones de dépassement associées

2. NOMBRE DE JOURS DE DEPASSEMENTS DE LA VALEUR LIMITE D'UNE CONCENTRATION MOYENNE : APPLICATION AUX PM₁₀

Une simple addition des zones de dépassement et d'incertitude sur toute l'année surestime très fortement le nombre de dépassements de la valeur limite journalière PM₁₀ puisque les zones d'incertitude tendent à être grandes. De même, additionner uniquement les zones de dépassement sous-estime ce nombre, même s'il semble plus réaliste (mettre en relation les figure 3 et 4 en 2007).

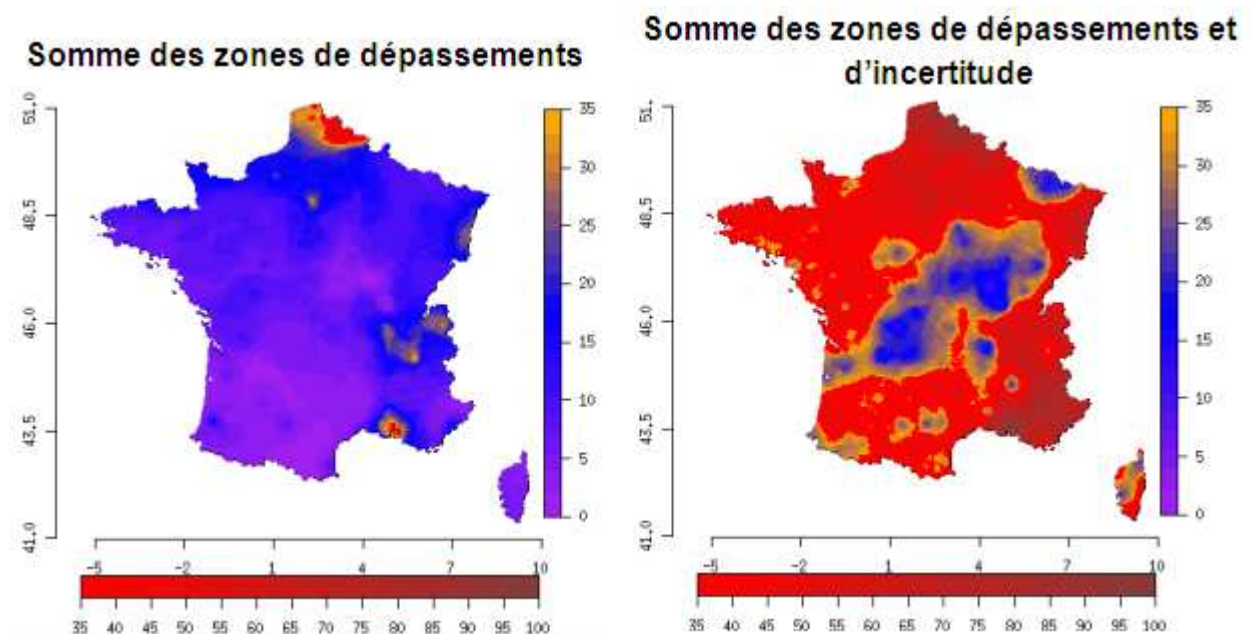


Figure 3 : Nombre de dépassements en 2007 obtenu en sommant a) les zones de dépassement journalières (carte de gauche) b) les zones de dépassement journalières augmentées des zones d'incertitude (carte de droite)

En conséquence, on cherche à développer une méthodologie exploitant les 365 cartes de probabilité journalière afin de proposer un modèle probabiliste capable d'estimer la probabilité de dépasser le seuil annuel des 35 jours de dépassement. Cette approche est présentée en annexe A.2.2.

De nombreux dépassements ont pu être observés aux stations de mesure en 2007, 2009 et 2011. L'application de la méthodologie permet de produire des cartes décrivant les zones où la probabilité de dépasser la valeur limite est élevée.

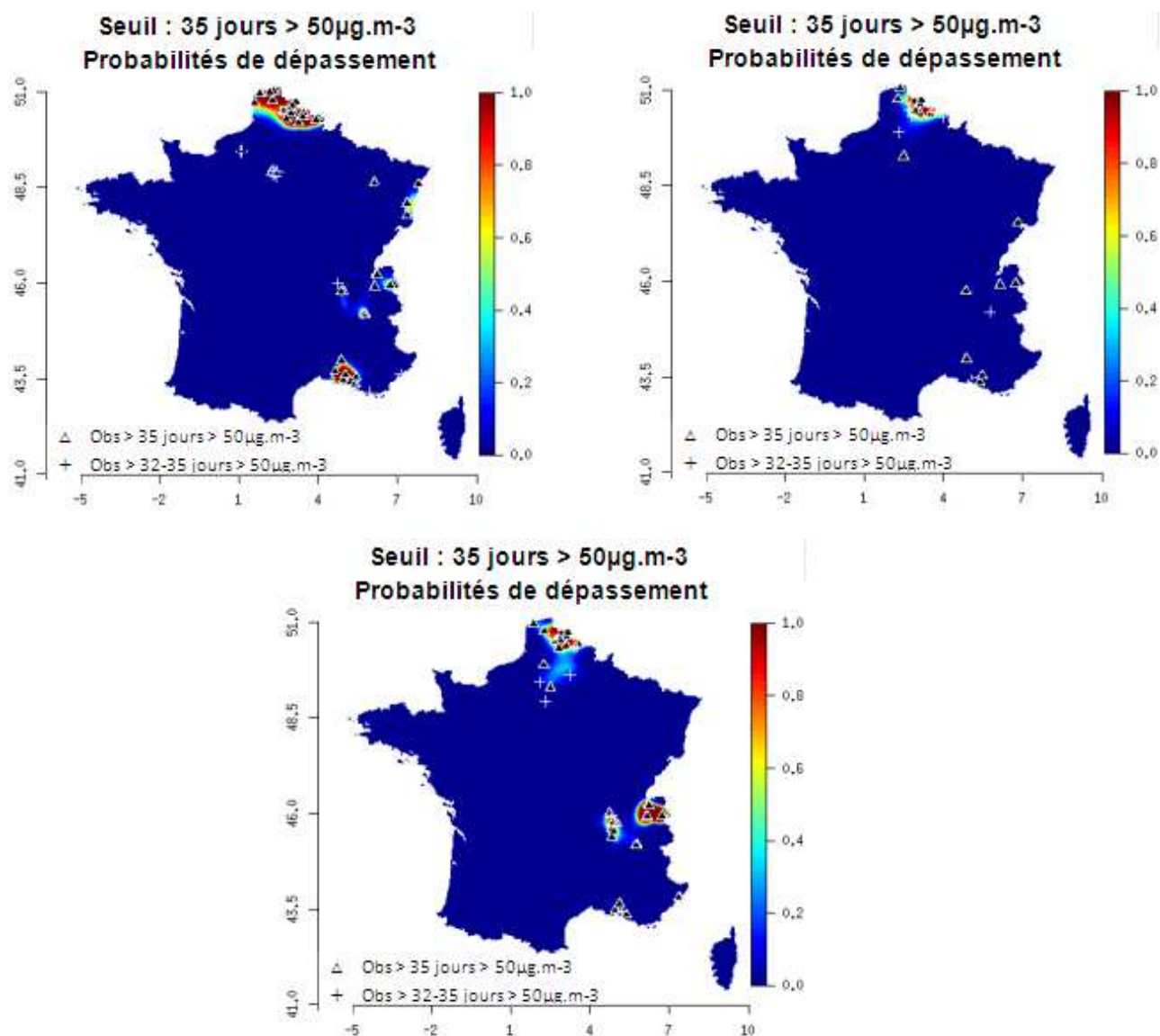


Figure 4 : Probabilités de dépassement de la valeur limite annuelle fixée à 35 journées. Année 2007, 2009 et 2011.

La figure 4 montre qu'en 2007 et 2011, la plupart des sites de mesure en dépassement sont situés dans une zone où la probabilité de dépasser la valeur limite est forte.

En 2009, certains sites situés en Rhône-Alpes ou dans la région de Marseille présentent de faibles probabilités en dépit de dépassements observés aux stations. Ces probabilités peuvent s'expliquer par le faible nombre de stations en dépassement dans cette zone qui comprend en revanche une forte densité de mesure des PM₁₀ avec un nombre de dépassements inférieur à 30.

Ces dépassements sont donc très localisés dans une zone proche de la station de mesure et il est difficile de les prendre en compte en l'absence d'informations additionnelles. Ce constat, qui a été confirmé par des résultats de validation croisée sur ce type de stations, ne remet pas en cause cette nouvelle approche, les mauvais résultats sur ce type de cas étant liés aux mauvaises

estimations des probabilités journalières et non au modèle de cumul de ces probabilités (voir A.2.2).

La figure 5 montre la carte de probabilité en 2009 après un seuillage à 5% et 1% pour définir la zone de non-dépassement. Contrairement aux cartes de probabilité journalières, la zone d'incertitude est très réduite et peut raisonnablement être incluse dans la zone de dépassement si besoin. Des seuils très faibles permettent de récupérer des zones peu étendues à proximité des stations de mesure présentant des dépassements très localisés.

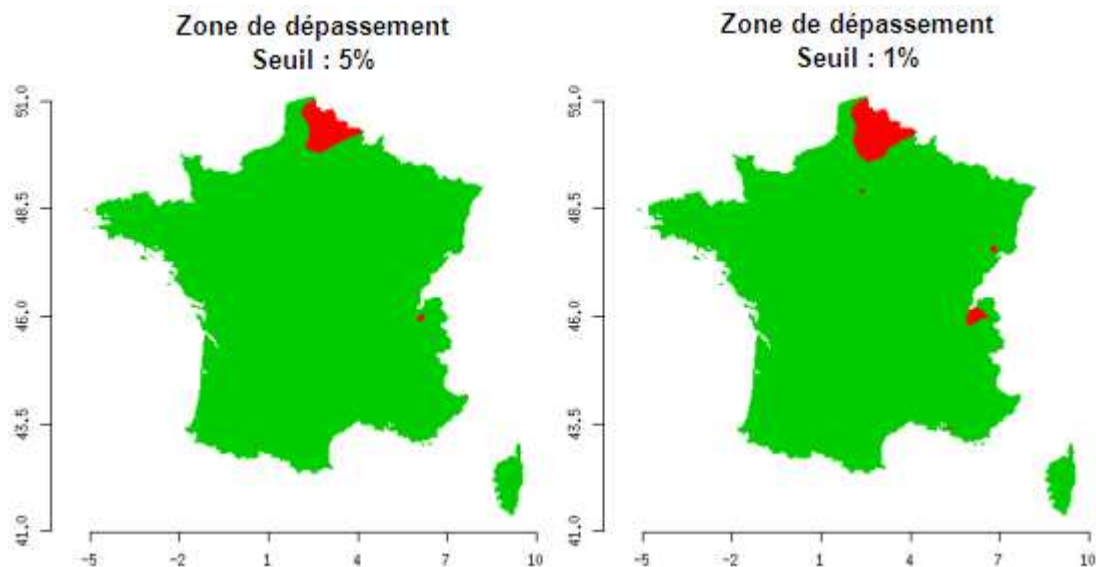


Figure 5 : Zone de dépassement (en rouge) de la valeur limite en 2009 après application d'un seuil de 5% (carte de gauche) et de 1% (carte de droite) à la carte de probabilité de la figure 4. Le seuil de probabilité 5% ou 1% est le risque que la valeur limite soit dépassée dans la zone verte.

3. EVALUER LA POPULATION EXPOSEE

En ce qui concerne la population exposée aux dépassements, deux approches pour estimer ce chiffre peuvent être considérées.

Premièrement, il est possible d'évaluer la quantité $B(s)$, population exposée en fonction du seuil s , par la formule $B(s) = \sum_i b_i \cdot I_{Z(x_i) \geq s}$ où :

- $Z(x_i)$ est l'estimation de la concentration fournie par krigeage si l'on s'intéresse aux dépassements de concentration, ou du nombre de dépassements si l'on s'intéresse aux dépassements cumulés sur l'année ;
- $I_{Z(x_i) \geq s}$ est l'indicatrice de dépassement du seuil s dans la maille i (pour les PM_{10} ; $s = 50 \mu\text{g}/\text{m}^3$ dans le premier cas, $s = 35$ dans le second) ;
- b_i est la population de la maille i .

En seuillant des cartes de probabilité selon un risque statistique δ (risque de dépasser dans la zone non dépassement), et en croisant les mailles en dépassement avec les données de population spatialisées, on peut estimer $B(s)$ comme suit :

$$B_{\delta}^*(s) = \sum_i b_i \cdot I_{\mathbb{P}[Z(x_i) \geq s] \geq \delta} \quad (1)$$

L'effet du risque δ choisi est illustré par la figure 6. En général, on préférera un risque faible.

L'autre option est d'estimer simplement l'espérance de la population exposée en sommant sur toutes les mailles le produit de la population dans la maille et la probabilité de dépassement qui lui est associée :

$$\mathbb{E}[B(s)] = \sum_i b_i \cdot \mathbb{P}[Z(x_i) \geq s] \quad (2)$$

La première approche fournit une estimation de la population exposée qui tend à être surestimée, mais elle assure la cohérence géographique entre la surface déclarée en dépassement et la population exposée aux dépassements. La deuxième approche conduit à un chiffre plus faible (figure 7), mais ce dernier ne correspond à aucune zone délimitée de dépassement.

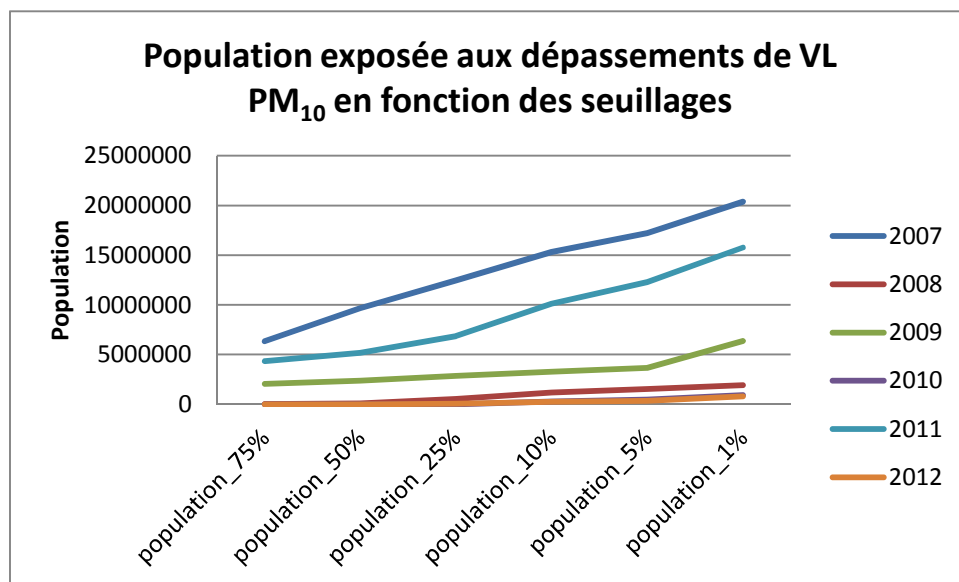
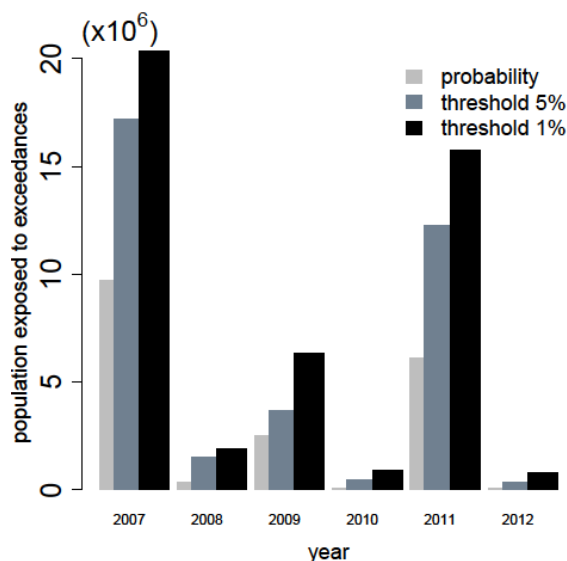


Figure 6 : Population exposée aux dépassements de la valeur limite annuelle PM_{10} . Méthode utilisée : calcul probabiliste. Evolution de la population exposée en fonction du risque δ choisi.



population (x10 ⁶)	2007	2008	2009	2010	2011	2012
$\mathbb{E}[B(s)]$	9.73	0.37	2,51	0.07	6,09	0.08
$B_{5\%}^*(s)$	17.21	1.54	3,65	0.47	12,28	0.36
$B_{1\%}^*(s)$	20.38	1.93	6,36	0.91	15,76	0.81

Figure 7 : Population exposée aux dépassements de la valeur limite PM_{10} , i.e. à plus de 35 dépassements du seuil journalier de $50 \mu\text{g}/\text{m}^3$.

Méthodes utilisées : calcul de l'espérance ; estimation par seuillage à 5% et 1% de la carte de probabilité de dépassement

4. DEVELOPPEMENTS

Les études conduites récemment par l'INERIS sur la question des dépassements portent sur :

1. l'estimation spatiale d'une probabilité de dépassement d'une valeur moyenne (annuelle, journalière, horaire) de concentration ;
2. l'estimation spatiale d'une probabilité de dépassement d'un nombre cumulé de dépassements horaires ou journaliers.

Des méthodologies ont été récemment mises au point, évaluées et appliquées à l'échelle nationale, avec le souci de développer un cadre probabiliste rigoureux qui soit compatible avec des contraintes opérationnelles (souplesse d'utilisation, temps de calcul). Le produit de ces travaux, et en particulier la nouvelle approche relative

aux dépassements cumulés, ont fait l'objet de plusieurs communications⁶. Des travaux d'amélioration sur le point 1 ont été engagés afin d'estimer plus précisément les probabilités de dépassement de valeurs moyennes. Les méthodes de la géostatistique non linéaire seront notamment explorées.

La question des populations exposées nécessite des investigations supplémentaires et des discussions méthodologiques avec le MEDDE et les AASQA comme il est prévu au second semestre 2013 (séminaire technique LCSQA du 27 septembre).

6 Beauchamp M., Malherbe L., Fouquet C. (de), 2013. A pragmatic approach to estimate probabilities of exceeding limit values in air quality: application to PM10 and O3. 23rd Annual Conference of The International Environmetrics Society, Anchorage, Alaska, USA, June 10-14, 2013.

Malherbe L., Beauchamp M., Fouquet C. (de), Létinois L., Ung, A., 2011. Estimation of the areas of air quality limit value exceedances on national and local scales. A geostatistical approach. European Regional TIES Conference: Spatial Data Methods for Environmental and Ecological Processes - 2nd Edition - Foggia, Baia delle Zagare (Italie), 1er-2 septembre 2011.

LISTE DES ANNEXES

Repère	Désignation	Nombre de pages
Annexe 1	Eléments théoriques	4

ANNEXE 1: ELEMENTSTHEORIQUES

A.2.1 Valeur limite d'une concentration moyenne

On note $Z_i(x)$ la concentration de la maille centrée en x le jour i , $Z_i^*(x)$ son estimée par krigeage et $\sigma_{ki}(x)$ l'écart-type de krigeage associé.

On considère l'erreur d'estimation $\varepsilon_i(x)$ qui suit de manière conventionnelle une loi normale de moyenne nulle et d'écart-type égal à $\sigma_{ki}(x)$: $\varepsilon_i(x) = \sigma_{ki}(x).T$ où $T \sim \mathcal{N}(0,1)$.

VL est la valeur limite considérée : VL = 50 μ g/m³ pour les PM₁₀ en moyennes journalières.

On ne considère pas ici la corrélation entre l'estimation par krigeage et son erreur. Alors, évaluer si la maille centrée en x est en dépassement peut-être mis en équation sous cette forme :

$$Z_i(x) > VL \Leftrightarrow Z_i^*(x) + \sigma_{ki}(x).T > VL \Leftrightarrow T > \frac{VL - Z_i^*(x)}{\sigma_{ki}(x)} \quad (1)$$

Dans la méthodologie proposée, les zones en dépassement et en non-dépassement sont définies à partir de l'équation (1), en considérant un seuil α de non-détection, qui est le risque statistique qu'une maille dans la zone de non-dépassement dépasse la valeur limite, et un seuil β d'erreur de détection, qui est le risque qu'une maille déclarée en dépassement ne dépasse pas la valeur limite.

Si la priorité est que le nombre de mailles situées par erreur dans la zone de dépassement soit le plus petit possible, alors α sera fixé à une valeur faible tandis que β pourra être plus fort. Cori propose que α soit fixé à 5% et β à 34% pour que le risque de non-détection soit modéré.

On obtient les définitions suivantes :

- **zone de non-dépassement** : $x / \mathbb{P}[Z_i(x) > VL] < \alpha$

$$\begin{aligned} \mathbb{P}\left[T > \frac{VL - Z_i^*(x)}{\sigma_{ki}(x)}\right] < \alpha &\Leftrightarrow Z_i^*(x) < VL - q_{1-\alpha} * \sigma_{ki}(x) \\ &\Leftrightarrow Z_i^*(x) < VL - 1.65 * \sigma_{ki}(x) \text{ avec } \alpha = 5\% \end{aligned} \quad (2)$$

- **zone de dépassement** : $x / \mathbb{P}[Z_i(x) \leq VL] < \beta$

$$\begin{aligned} \mathbb{P}\left[T \leq \frac{VL - Z_i^*(x)}{\sigma_{ki}(x)}\right] < \beta &\Leftrightarrow Z_i^*(x) > VL - q_{\beta} * \sigma_{ki}(x) \\ &\Leftrightarrow Z_i^*(x) > VL + 0.41 * \sigma_{ki}(x) \text{ avec } \beta = 34\% \end{aligned} \quad (3)$$

$q_{1-\alpha}$ and q_{β} sont les $(1-\alpha)$ et β -quantiles de la loi normale centrée réduite

Les mailles ne respectant aucun de ces deux critères sont placées dans une zone d'incertitude.

A.2.2 Nombre de jours de dépassements de la valeur limite d'une concentration moyenne : application aux PM₁₀

Le nombre jours en dépassement est une notion qui apparaît dans la réglementation européenne et des valeurs limites sont définies : VL = 35 jours > 50µg/m³ à ne pas dépasser pour les PM₁₀. Une première réponse peut être apportée par une carte de probabilité de dépassement produite à partir des 365 probabilités journalières obtenues avec la méthodologie décrite dans la section précédente.

On introduit les notations suivantes :

- (i) p_i la probabilité de dépasser la valeur limite journalière le jour i (en x) :
 $p_i = \mathbb{P}[Z_i > 50]$
 (ii) N le nombre de jours en dépassement

On cherche à évaluer la probabilité que N soit égal à k . On note \mathcal{P}_k l'ensemble des $\binom{365}{k}$ combinaisons possibles d'évènements pour lesquelles le nombre de jours en dépassement vaut k . On considère un évènement E_j de \mathcal{P}_k , $j \in [1, |\mathcal{P}_k|]$, on note $\mathcal{D}_{E_j}^+$ l'ensemble des jours qui dépassent la valeur limite journalière, $\mathcal{D}_{E_j}^-$ l'ensemble des jours qui ne la dépassent pas. Le caractère disjoint de ces évènements permet d'écrire :

$$\mathbb{P}[N = k] = \mathbb{P}\left(\bigcup_{j \in [1, |\mathcal{P}_k|]} E_j\right) = \sum_{j \in [1, |\mathcal{P}_k|]} \mathbb{P}[E_j] \quad (4)$$

On simplifie le problème en considérant les dépassements comme indépendants d'un jour à l'autre. Alors, l'équation (4) peut s'écrire :

$$\mathbb{P}[N = k] = \sum_{j \in [1, |\mathcal{P}_k|]} \mathbb{P}[E_j] = \sum_{j \in [1, |\mathcal{P}_k|]} \left(\prod_{i \in \mathcal{D}_{E_j}^+} p_i \prod_{i \in \mathcal{D}_{E_j}^-} (1 - p_i) \right) \quad (5)$$

Si l'on connaît la probabilité que le nombre de jours en dépassement soit égal à k , il est alors facile de satisfaire aux exigences de la réglementation européenne, et d'évaluer la probabilité que le nombre de jours en dépassement dépasse k (puisque toutes les combinaisons d'évènements sont disjointes, les probabilités sont ajoutées)

$$\mathbb{P}[N > k] = \sum_{k=VL+1}^{365} \sum_{j \in [1, |\mathcal{P}_k|]} \left(\prod_{i \in \mathcal{D}_{E_j}^+} p_i \prod_{i \in \mathcal{D}_{E_j}^-} (1 - p_i) \right) \quad (6)$$

Le calcul de cette probabilité, qui correspond à la fonction de répartition de la variable « nombre de jours en dépassement », est impossible pour k grand car le nombre de toutes les combinaisons possibles de jours en dépassement devient très important. Il faut alors modéliser la distribution de cette variable par une loi de probabilité discrète.

Le problème du calcul d'une probabilité associée au nombre d'apparition d'un évènement rappelle une des utilisations possibles de la distribution de Poisson mais l'hypothèse sous-jacente de l'égalité de l'espérance et de la variance n'est pas satisfaite dans notre cas, puisque pour la plupart des mailles, le nombre de journées avec une forte (resp. faible) probabilité de dépasser la valeur limite journalière est beaucoup plus important que le nombre de jours avec une faible (resp. forte) probabilité de la dépasser.

Il a été décrit dans la littérature différentes façons de prendre en compte ce phénomène de sous-dispersion.

Pour sa facilité d'utilisation, notre choix se portera sur les lois de Poisson translatées (Mizere et al, 2006).

Considérons Y une variable de Poisson translatée :

$$Y = k + X \quad (7)$$

où X est Poissonienne avec une espérance μ , et k , un paramètre de translation, qui peut être vu comme le nombre minimum de dépassements.

Alors la distribution est défini comme suit :

$$Y \sim \mathcal{P}_t(k, \mu) \Leftrightarrow \begin{cases} \mathbb{E}[Y] = k + \mathbb{E}[X] = k + \mu \\ \mathbb{V}[Y] = \mathbb{V}[X] = \mathbb{E}[X] = \mu \\ \mathbb{P}[Y = y] = p(y; k; \mu) = \frac{\mu^{y-k}}{(y-k)!} e^{-\mu} \text{ avec } y = k, k + 1, k + 2, \dots \end{cases} \quad (8)$$

k est le paramètre de translation : $k = \mathbb{E}[Y] - \mathbb{V}[Y]$

μ est le paramètre de centrage : $\mu = \mathbb{E}[Y] - k$

Précédemment, $\mathbb{E}[Y]$ et $\mathbb{V}[Y]$ ont été classiquement estimé : $\mathbb{E}[Y] = \sum_{i=1}^{365} p_i$ et $\mathbb{V}[Y] = \sum_{i=1}^{365} p_i(1 - p_i)$

Vérifier la pertinence du modèle apparaît difficile puisque le calcul de (5) n'est pas possible pour k grand. Cependant, pour certains cas où le nombre de jours avec une probabilité nulle de dépasser la valeur limite journalière est important (on note N_0 ce nombre), il est intéressant de considérer l'ensemble \mathcal{P}_k^+ , sous-ensemble de \mathcal{P}_k , dont le nombre d'éléments est égal à $\binom{365 - N_0}{k}$, et pour lequel il est possible de valider l'ajustement de la formule (5) par la distribution de Poisson translatée.

On observe les ajustements pour toutes les stations vérifiant $20 < 365 - N_0 < 25$ entre 2007 et 2011. Ci-après les graphiques superposant les fonctions de masse et de répartition empiriques à leurs ajustements :

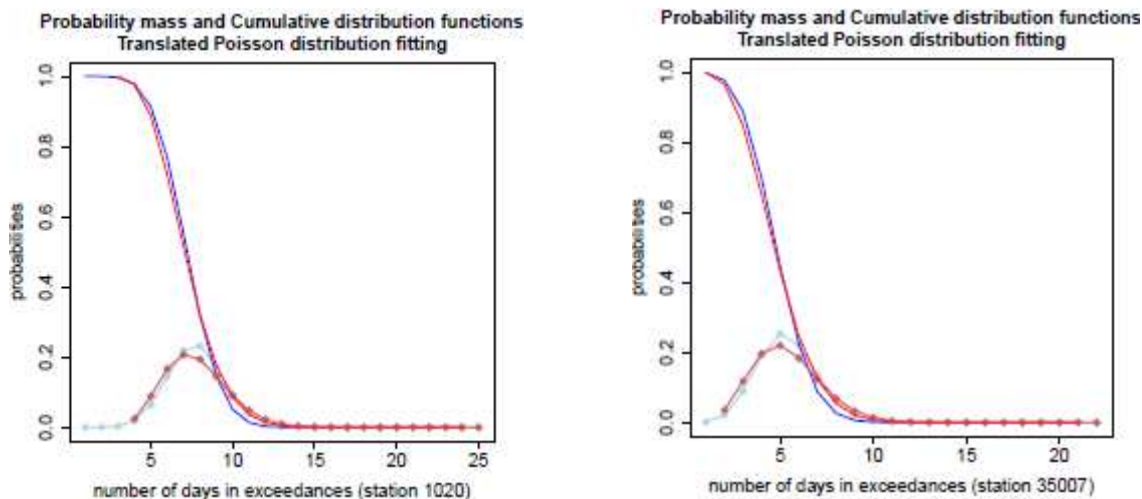


Figure 1: Fonction de masse et fonction de répartition empirique. Ajustement avec une distribution de Poisson translatée. Station 1020 et 35007 en 2009

Le terme d'erreur absolue est également calculé en chaque k et pour chaque station, ce qui donne un aperçu des erreurs commises par le modèle.

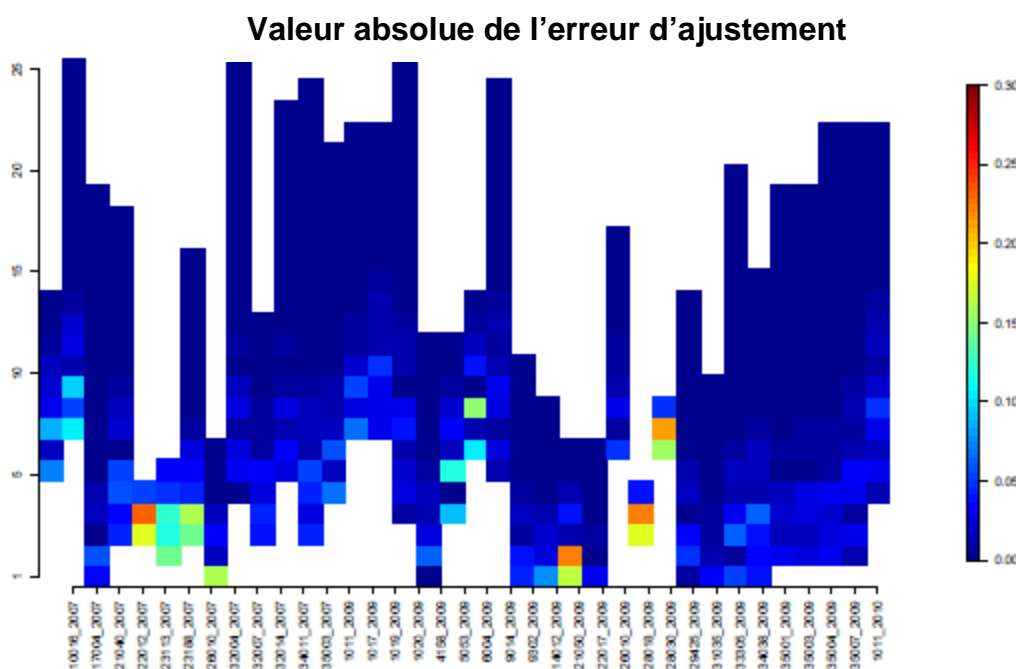


Figure 8 : Valeur absolue de l'erreur d'ajustement de la fonction de masse par une loi de Poisson translatée

De manière générale, la loi de Poisson translatée s'ajuste bien aux fonctions de masses associées à nos variables aléatoires, avec des erreurs la plupart du temps inférieures à 0.05 en valeur absolue. Les problèmes d'ajustement apparaissent surtout dans le cas des variables présentant un très faible nombre de dépassements (<10), ce qui n'aura pas de répercussion sur nos cartes finales puisque la valeur limite est fixée à 35 jours de dépassement.