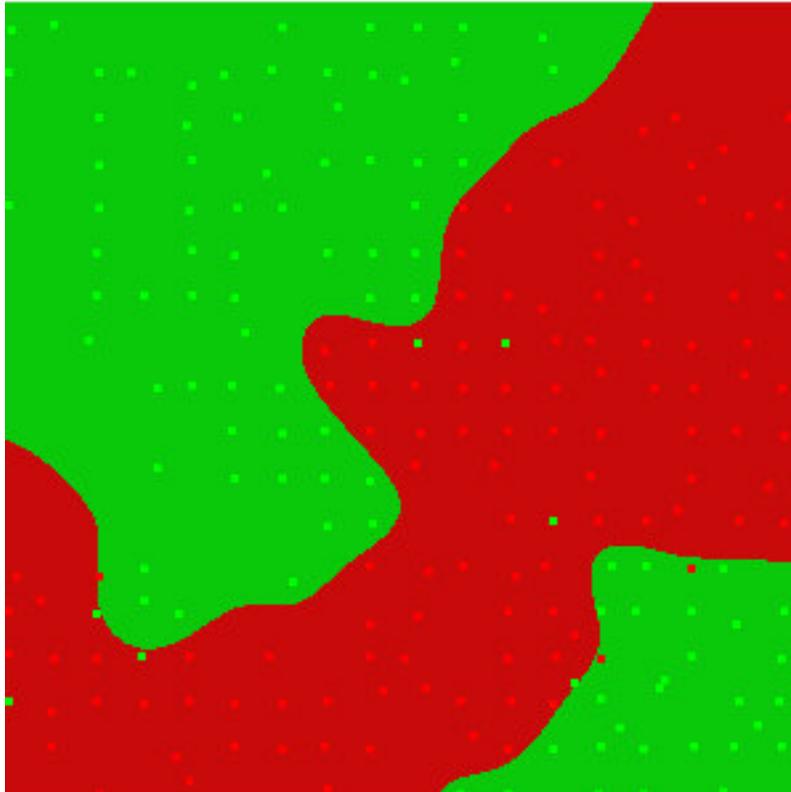


Dossier de M8



Réalisation d'une carte de répartition des risques à partir d'un échantillon de mesures de radioactivité

Table des matières

1. But du projet.....	3
2. Démarche.....	3
1. Recherches initiales.....	3
2. Etude des données.....	3
3. Création du programme.....	5
3. Programmation.....	5
1. Calculs incroyables.....	5
2. Calcul du sigma.....	6
1. Explication du sigma.....	6
2. Déroulement de l'expérience.....	6
3. Interprétation.....	7
4. Traitement des résultats.....	8
1. Frontière et courbes de niveau.....	8
2. Carte finale (avec cotations) :.....	9
3. Commentaires sur la carte.....	9
5. Conclusion.....	10

1. But du projet

Ce projet vise à réaliser une carte la plus exacte possible de la répartition des retombées radioactives dans la région de Tchernobyl. Les données recueillies se présentaient sous la forme d'un fichier texte et ont été validées par M. Canu. Il s'agit de mesures positives ou négatives assorties de deux variables correspondant à leur positionnement en X et Y.

Nous avons essayé par la suite de traiter ces données de manière automatique à l'aide d'un petit logiciel développé par nos soins en Java. Nous détaillerons de manière plus précises les différentes étapes du traitement réalisé par ce programme par la suite.

Ensuite, les cartes sont retouchées à l'aide du logiciel The Gimp afin de les enrichir par des courbes de niveau, des annotations, etc.

Le fichier image réalisé par le programme permet d'avoir une idée assez précise des différents degrés de contamination.

2. Démarche

1. Recherches initiales

Au début du semestre, nous voulions nous intéresser aux risques engendrés par la proximité des centrales nucléaires (habitants et environnement). Nous n'avons malheureusement pas trouvé de données exploitables. En effet, il s'agissait toujours de rapport, mais nous n'avions pas accès aux sources.

Par la suite, nous avons fait part de notre inquiétude à notre très estimé professeur de statistiques, qui nous a aimablement procuré des données concernant la quantité de radioactivité aux alentours de Tchernobyl.

Nous avons donc reçu deux fichiers de données (validation et test) qui nous ont permis d'obtenir la carte.

2. Etude des données

Nous disposons de 217 mesures, réparties en deux catégories :

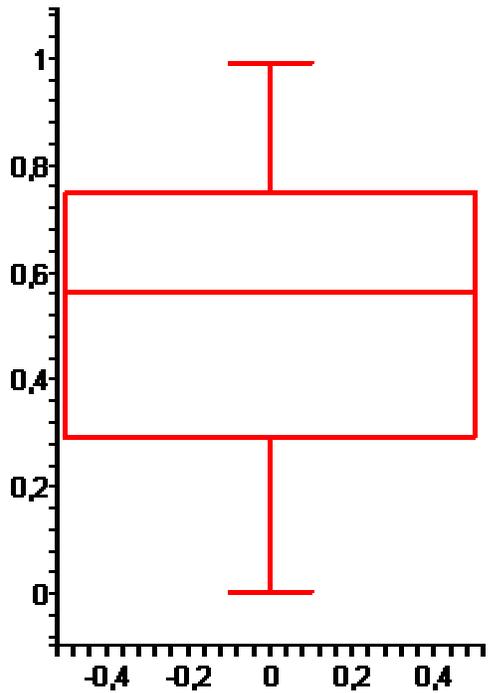
- Mesures positives (112 valeurs)
- Mesures négatives (105 valeurs)

Il faut noter qu'il n'y a pas d'indication sur le degré de la contamination. En effet, les valeurs des mesures sont soit 0 (pas de contamination) ou 1 (contamination détectée).

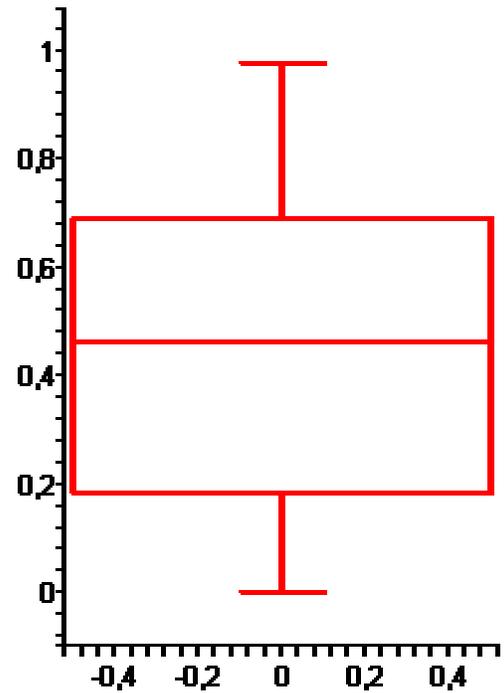
Les positions en X et Y sont normalisées (de 0 à 1) afin de simplifier les calculs. Par contre, ils rendent impossible toute tentative de superposer une carte réelle à la carte des risques.

Nous avons utilisé le logiciel Maple afin d'obtenir les boîtes à moustache illustrant la répartition des données sur l'axe des X et des Y.

Répartition des mesures :



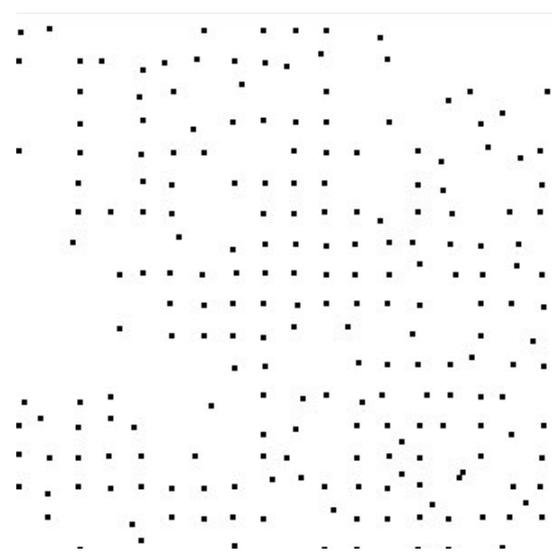
Axe des X



Axe des Y

Nous pouvons observer que la répartition des mesures est plus ou moins homogène, que ce soit sur l'un ou l'autre des axes.

Répartition des mesures sur la carte. (Tous types confondus) :



Les mesures étant réalisées de manière quasi-homogène, nous pouvons poursuivre nos calculs.

3. Création du programme

En premier lieu, nous n'avons pas su comment utiliser ces données. Puis nous avons commencé à raisonner sur un système simple : Chaque pixel de la carte reçoit pour valeur la somme de toutes les contributions des mesures, pondérées par la distance entre la mesure et le pixel. Le principal problème a donc été de trouver un système de pondération qui permettait d'avoir des valeurs cohérentes. Pour tester le programme, nous avons utilisé un système de pondération simple : on ajoutait à la valeur de chaque pixel l'inverse de sa distance à la mesure. Mais ce système n'était pas optimal. Nous nous sommes donc tourné vers un moyen plus à même de modéliser la réalité et avons utilisé les courbes gaussiennes.

Le programme est réalisé en Java.

3. Programmation

1. Calculs incroyables

Soit pour A un point de la carte, et M_x l'ensemble des n mesures

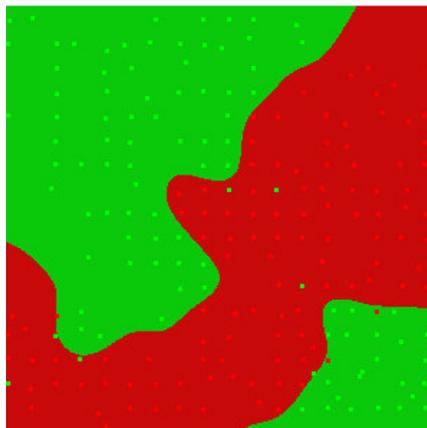
$$poids(A) = \sum_{p=1}^n e^{\frac{-distance(A, M_p)^2}{2\sigma}}$$

Pour obtenir cette formule, nous nous sommes inspirés de la formule de la loi normale. Le sigma nous donne donc une idée de la zone d'influence de chaque mesure. Programmation

Chaque pixel de la carte est ensuite affiché en rouge, si sa valeur est inférieure à un certain seuil, sinon, il sera affiché en vert. C'est ainsi que j'ai réalisé la carte suivante, qui permet d'avoir une idée de probable radioactivité de certaines zones :

J'ai utilisé pour cela un sigma de 250 et une carte de 300 pixels x 300 pixels.

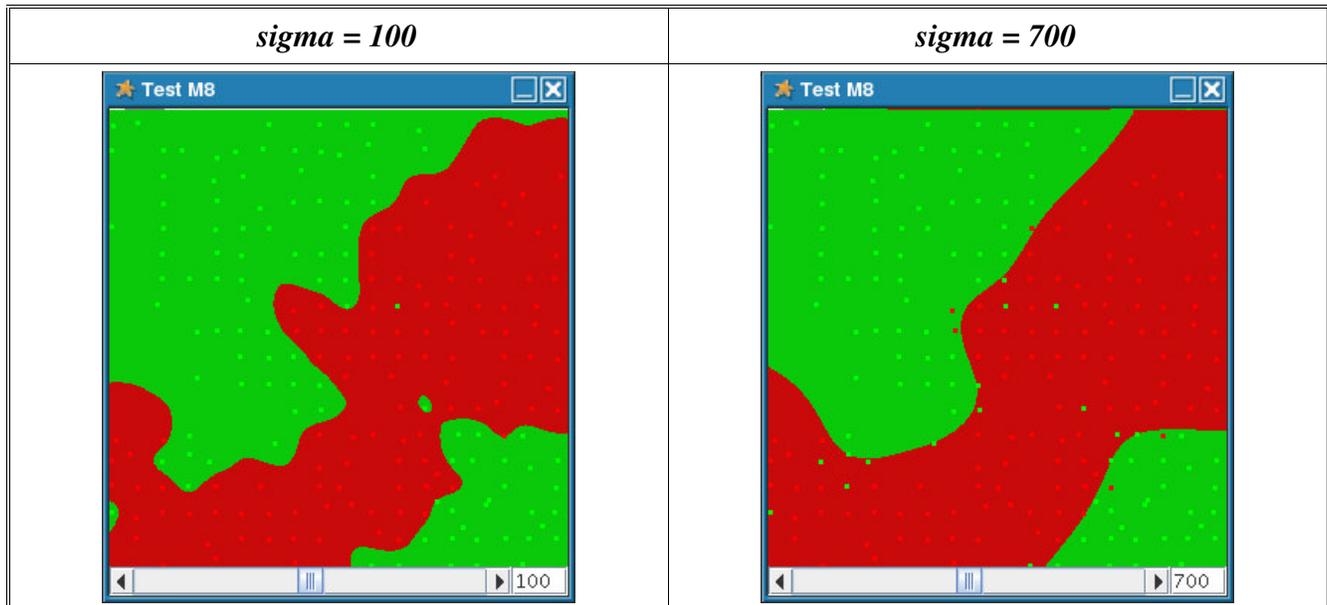
Cette dernière information est importante car la distance entre un point et la mesure est calculée en pixels et non dans l'unité de mesure utilisée dans le fichier texte. On observe 10 points aberrants qui ne correspondent pas à leur zone respective.



2. Calcul du sigma

1. Explication du sigma

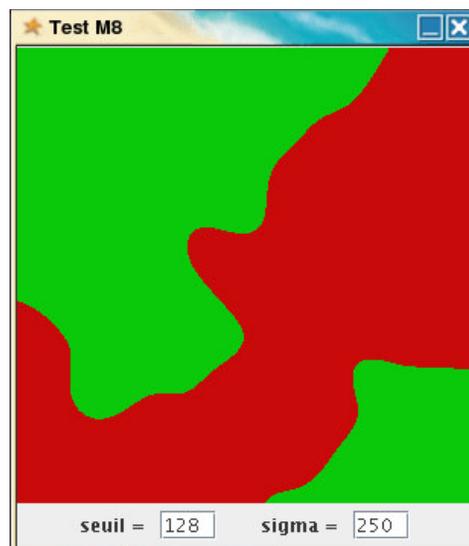
La valeur de σ correspond à la zone d'influence de chaque mesure. On peut voir très facilement qu'une valeur de sigma élevée a tendance à "lisser" les frontières tandis qu'un sigma bas a tendance à faire coller les frontières aux mesures.



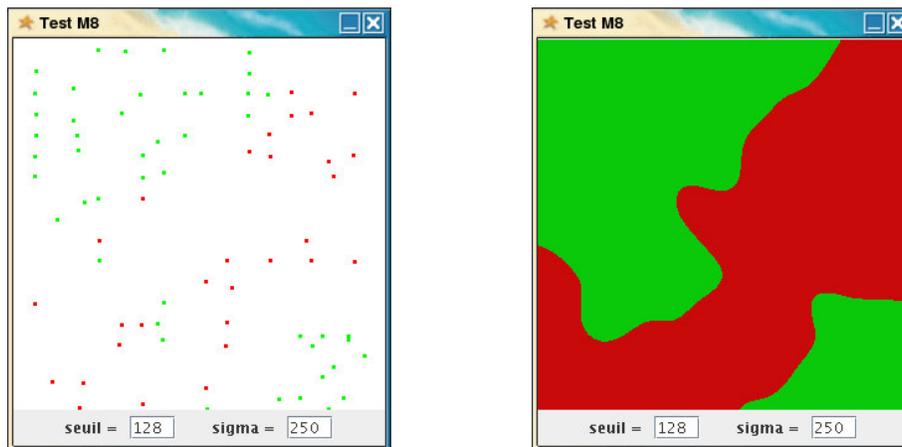
Pour obtenir un sigma optimal, on réalise une série d'expériences à partir des données d'apprentissage et de test en prenant des valeurs différentes de sigma. On retient ensuite le sigma le plus approprié.

2. Déroulement de l'expérience

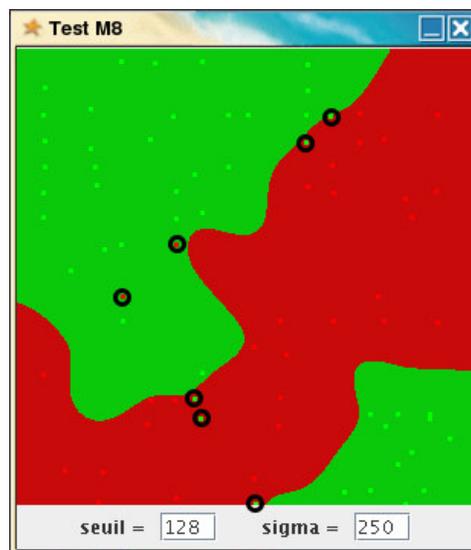
- On choisit un sigma (ici 250)
- On réalise une carte à partir des données du fichier d'Apprentissage



- Cette carte sera superposée sur les données du fichier de Test



- On compte les points mal placés (ici les point cerclés de noir)



3. Interprétation

Le sigma ayant donné lieu au plus petit nombre de points mal placés sera retenu. Malheureusement, nous n'avons pas observé de différences dans le nombre de points mal placés, et ce quelque soit la valeur du sigma (entre 100 et 700). Nous avons alors fixé arbitrairement le sigma à 250, car cela nous semblait être un bon compromis entre les deux extrêmes.

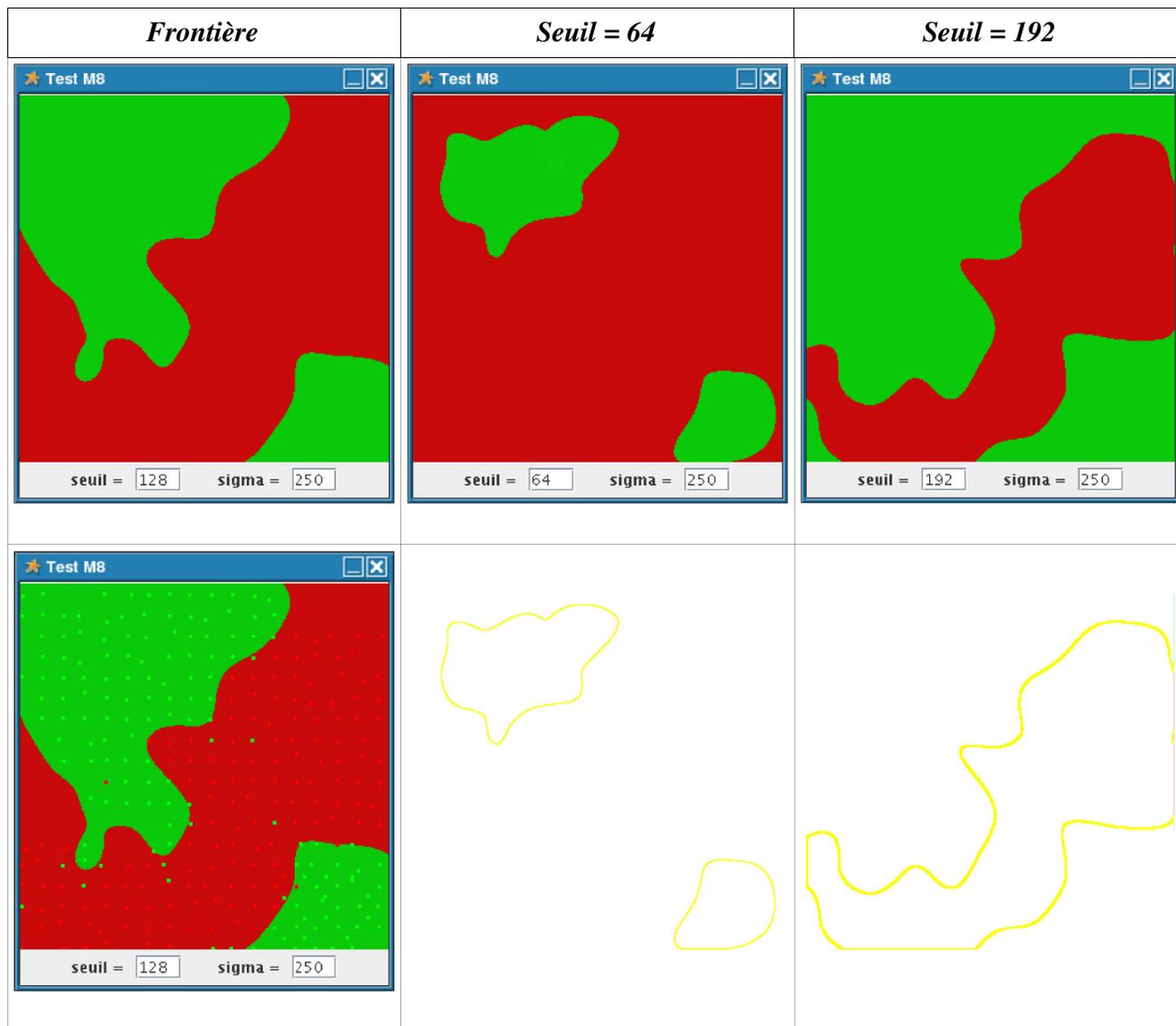
Il est dommage de ne pas avoir eu accès aux 800 valeurs du fichier original, car cela nous aurait permis d'obtenir une valeur de sigma plus précise.

4. Traitement des résultats

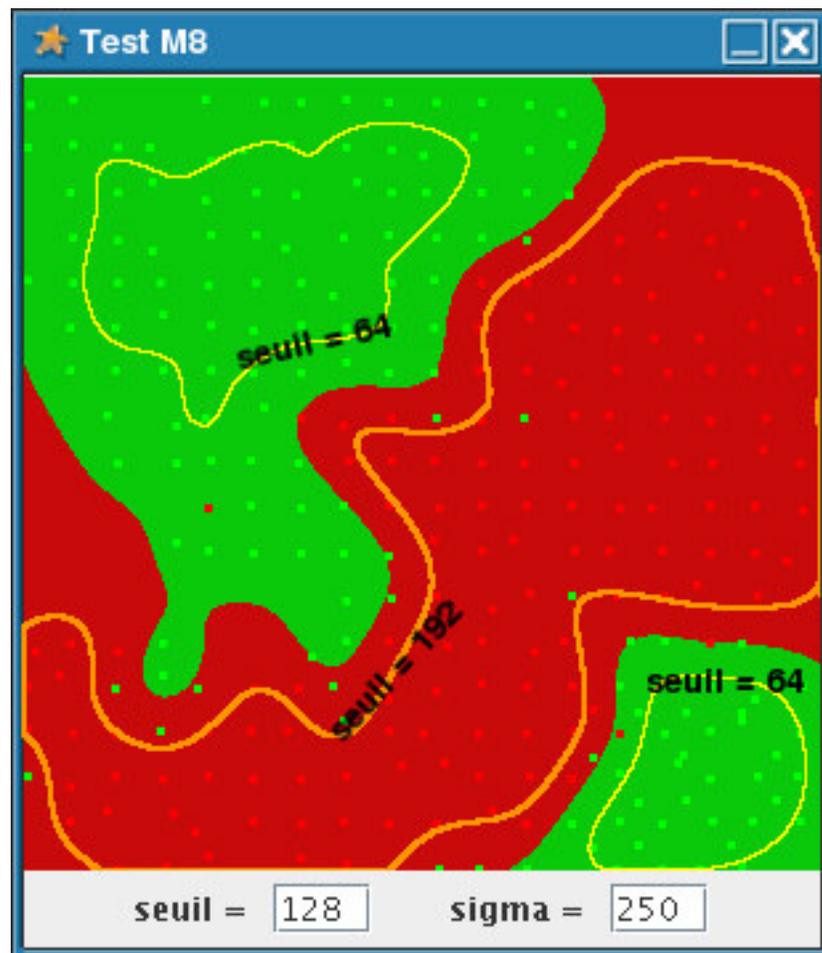
1. Frontière et courbes de niveau

Le programme permet de réaliser l'affichage de la frontière, mais il n'a pas été programmé pour afficher les courbes de niveau. Pour cela, nous réalisons des cartes avec différents seuils pour un sigma donné. Ensuite, nous les retraits avec un logiciel de traitement d'image (The Gimp) qui fera apparaître le contour. L'image finale est obtenue en superposant toutes les courbes obtenues.

Nous avons décidé de prendre toutes les valeurs données afin d'augmenter la précision de la carte. Les données d'apprentissage et les données de test ont donc été fusionnées.



2. Carte finale (avec cotations) :



3. Commentaires sur la carte

Nous avons rajouté les lignes de niveau correspondant à certaines probabilité de radioactivité :

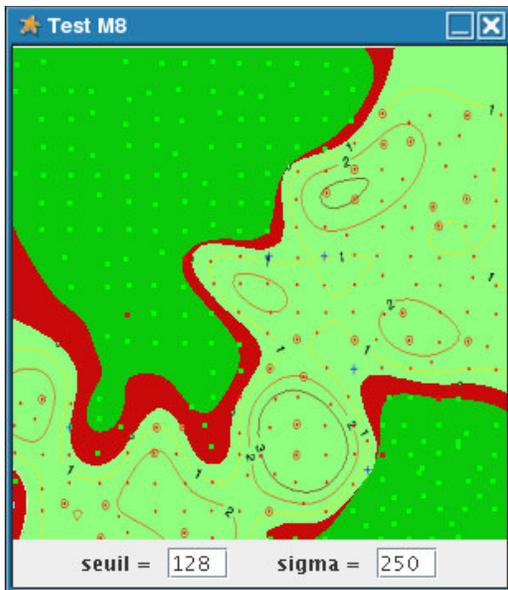
- jaunes pour un seuil de 64 correspondant à une probabilité d'infection de 25%
- la frontière correspondant à une probabilité d'infection de 50%
- orange pour un seuil de 192 correspondant à une probabilité d'infection de 75%

Par probabilité d'infection nous entendons la probabilité d'obtenir une mesure positive à cet endroit.

5. Conclusion

Ce projet nous a permis de découvrir une facette des statistiques que nous n'imaginions pas. En effet, nous ne pensions pas que les statistiques pouvaient aider à résoudre ce genre de problèmes.

Nous avons eu avec les données une carte-type réalisée par M. Canu que nous avons superposée avec la notre :



Nous pouvons observer que la notre est assez proche de la carte de M. Canu. La valeur du sigma que nous avons choisie est certainement plus élevée que la sienne. En effet, notre carte est plus “lisse”, elle ne colle pas autant aux points de la carte.

Ce travail nous a donné un aperçu de la démarche utilisée pour réaliser des cartes de risques. Nous sommes satisfaits de ces résultats que nous avons pu obtenir avec ces calculs pourtant relativement simples.