

Note technique

ASSISTANCE EN CARTOGRAPHIE ET TRAITEMENT DE DONNEES

Synthèse des actions 2014

Laure Malherbe (INERIS), Maxime Beauchamp (INERIS),
Laurent Létinois (INERIS), Frédéric Tognet (INERIS)

SYNTHESE

Le LCSQA assure chaque année une assistance technique et méthodologique relative au traitement statistique et géostatistique des données et à l'élaboration de cartographies. Afin de mieux répondre aux besoins des AASQA concernant la cartographie de la pollution en milieu urbain, et en complément des travaux méthodologiques conduits sur ce sujet, cette assistance s'est étendue à l'utilisation des modèles de dispersion.

La présente note synthétise les actions réalisées en 2014 dans ces différents domaines pour le compte des AASQA. Ces actions ont pris différentes formes : échanges téléphoniques, déplacement sur site, sessions de formation. Parmi les tâches effectuées en 2014, on relèvera plus particulièrement la mise au point d'une nouvelle formation sur l'application de la géostatistique et la production de cartographies avec le logiciel R et un appui technique en modélisation auprès de Qualitair Corse.

1. INTRODUCTION

La planification de campagnes de mesure, l'exploitation des données de surveillance, qu'il s'agisse de mesures fixes, de mesures temporaires ou de résultats de modélisation, et l'élaboration de cartographies supposent la mise en œuvre de techniques statistiques ou géostatistiques. Chaque année, le LCSQA propose son appui aux AASQA dans l'application des méthodes existantes et l'interprétation des résultats obtenus. Afin de prendre en compte l'évolution des moyens et des pratiques et le développement de la modélisation à l'échelle urbaine, cette assistance s'étend désormais à l'utilisation des modèles de dispersion. La présente note synthétise les actions conduites en 2014 dans ces différents domaines.

2. TRAITEMENT STATISTIQUE DES DONNEES

2.1 Définition de plans d'échantillonnage et reconstitution de données

Air Normand a rencontré des difficultés dans l'utilisation de l'outil de planification de l'échantillonnage. Un bug survenu lors de modifications sur le serveur informatique du LCSQA a été identifié et corrigé (il s'agissait d'une mauvaise prise en compte des strates dans le cas de plans d'échantillonnage stratifiés).

Afin de pallier temporairement ce problème, le LCSQA a envoyé à Air Normand un mode opératoire lui permettant de lancer le programme de planification en dehors de l'interface. Ce mode d'emploi est joint en annexe.

Le LCSQA prévoit pour 2015 une vérification complète de l'outil de planification de l'échantillonnage (<http://www.lcsqa.org/page/echantillonnage>) et de reconstitution de données (<http://www.lcsqa.org/page/reconstitution>).

2.2 Formation en statistique

2.2.1 Session de base

Une nouvelle session de formation en statistique avec le logiciel R a été organisée les 3 et 4 décembre 2014. Animée par le LCSQA, en collaboration avec Frédéric Lavancier, de l'université de Nantes, elle a accueilli quinze participants des AASQA (Air Breizh, Air COM, Air PACA, Airparif, Air Rhône-Alpes, Atmo Nord-Pas-de-Calais, Atmo Poitou-Charentes, ORA) et du LCSQA (INERIS). Il s'agissait de la septième édition de la formation conçue en 2010 par le LCSQA avec le concours de F. Lavancier et de la société Stat-Consult¹.

¹ Malherbe L., 2010. Assistance à l'exploitation de données de campagnes et à la réalisation de cartographies Organisation d'une formation en statistique. Rapport LCSQA, ref. DRC-10-111609-05224A,

Cette formation s'adresse aux acteurs de la surveillance qui souhaitent développer leurs compétences en analyse et en traitement de données. Alternant points théoriques et cas pratiques avec le logiciel R, elle aborde les sujets suivants :

- *Analyses graphiques* : univariées, bivariées, multivariées (ACP, classification),
- *Tests statistiques et modélisation* : significativité, tests statistiques, comparaison de moyennes, généralisation : analyse de variance (ANOVA), régression linéaire,
- *Aspects temporels* : séries temporelles et dépendance temporelle, analyse des cycles et de la tendance, précautions à prendre en présence de dépendance temporelle.

Une partie de la formation est réservée aux échanges et le cas échéant, à l'étude des jeux de données apportés par les participants.

2.2.2 Session de perfectionnement

Une journée de travaux dirigés a été proposée aux personnes ayant déjà suivi la session de base (en 2014 ou antérieurement) ou possédant une expérience dans le traitement des données avec R. Organisée le 5 décembre 2014, elle a rassemblé cinq participants issus d'Airparif, Air Rhône-Alpes, Atmo Poitou-Charentes et ORA.

L'objectif était d'offrir aux stagiaires la possibilité de mettre en pratique et approfondir les connaissances acquises lors de la session de base et de les guider dans l'exploitation de leurs jeux de données.

3. CARTOGRAPHIE

L'assistance en cartographie inclut l'application de méthodes géostatistiques, l'utilisation d'outils de représentation cartographique et la mise en œuvre de modèles de dispersion.

3.1 Formation en cartographie avec R

Une formation théorique en géostatistique appliquée à la qualité de l'air a été organisée en 2009. Des échanges ultérieurs avec les AASQA ont montré le besoin d'une formation plus pratique, orientée sur l'utilisation du logiciel R.

Depuis quelques années, l'INERIS a de plus en plus recours à ce logiciel pour construire des cartes de pollution atmosphérique. Cet outil, libre et gratuit, prend une importance croissante dans la communauté scientifique et les librairies qu'il propose en matière de géostatistique et de cartographie fournissent des solutions intéressantes pour les acteurs de la qualité de l'air.

Une nouvelle formation a ainsi été élaborée. Elle permet aux AASQA qui souhaitent développer leurs compétences en R, d'apprendre à construire des cartographies par des méthodes géostatistiques et à visualiser les résultats obtenus.

<http://www.lcsqa.org/rapport/2010/ineris/assistance-exploitation-donnees-campagnes-realisation-cartographies-12>

Deux sessions ont eu lieu en 2014. Onze participants (Air Rhône-Alpes, ASPA, Atmo Champagne-Ardenne, Atmo Poitou-Charentes, Atmosf'Air, Lig'Air, INERIS) étaient présents à la session des 14 et 15 avril 2014 et neuf participants (Air Breizh, Air PACA, Atmo Auvergne, Atmo Picardie, Qualitair Corse, INERIS, Mines Douai) à la session des 30 et 31 octobre 2014.

Pour une meilleure appropriation des méthodes, cette formation requiert un minimum d'expérience en R ou en programmation. Son déroulé suit l'organisation usuelle d'une étude géostatistique (analyse exploratoire, étude variographique, estimation par krigeage monovariable ou multivariable). Chaque point théorique nécessaire à la compréhension de la démarche est immédiatement suivi d'exercices sur des données de qualité de l'air. La dernière partie fournit des éléments utiles pour améliorer l'aspect des cartes.

3.2 Assistance en modélisation

Au mois de septembre 2014, le LCSQA est intervenu auprès de Qualitair Corse. Il s'agissait d'assister l'équipe de modélisation de cette AASQA dans le développement d'une chaîne de calcul. Les différentes composantes de cette chaîne, du calcul des émissions à la modélisation de la dispersion à l'échelle d'un hyper-centre, ont été examinées. A la fin des deux jours, une première série de simulations a été réalisée avec ADMS-Urban sur le centre-ville d'Ajaccio ainsi qu'une cartographie de ces premiers résultats de simulation. Ce travail effectué à partir de données partielles a permis d'identifier et d'ordonner selon leur priorité les tâches nécessaires pour obtenir des résultats exploitables.

4. CONCLUSION

Comme les années précédentes, le soutien apporté en 2014 concernait aussi bien l'analyse statistique des données que l'élaboration de cartographies. La large audience lors des formations organisées sur ces sujets confirme l'intérêt croissant des AASQA pour ces types de traitement et pour le logiciel R. Celui-ci offre en effet une grande variété de fonctions statistiques et géostatistiques et se révèle un outil d'exploitation efficace. Compte tenu des avis positifs formulés par les participants, ces actions de formation se poursuivront en 2015.

A la suite du travail conduit en 2013 auprès de MADININAIR, l'assistance a été également étendue à la modélisation de la dispersion à petite échelle, qui est de plus en plus employée à des fins de surveillance et de planification.

Notons que certaines actions d'assistance sont plus spécifiquement incluses dans les études LCSQA qui leur correspondent. C'est le cas de l'aide apportée sur :

- la mise en œuvre du modèle CHIMERE, qui s'inscrit dans l'étude LCSQA relative à PREV'AIR et CHIMERE,
- l'utilisation des données de population, qui s'inscrit dans l'étude LCSQA sur la spatialisation de la population,
- l'interpolation des sorties de modèles, qui est prévue dans l'étude LCSQA concernant la cartographie des populations exposées.

ANNEXE I

Mode opératoire pour l'utilisation du programme de planification de l'échantillonnage

- 1) Entrer les paramètres choisis dans l'interface (<http://www.lcsqa.org/page/echantillonnage>) et lancer le calcul. La notice d'utilisation est disponible à l'adresse : <http://www.lcsqa.org/outil/missions-diverses/strategie-surveillance/planification-temporelle-echantillonnage>
- 2) Enregistrer l'archive .zip des résultats. Le dossier zip contient deux sous-dossiers : *reception* et *resultats*.

Le dézipper et placer les sous-dossiers *reception* et *resultats* dans un répertoire :

Par exemple :

C:/tempR/Pech_donnees/reception

C:/tempR/Pech_donnees/resultats

Si les paramètres saisis n'ont pas été correctement pris en compte, le programme de planification peut être lancé directement dans R selon la procédure décrite ci-après.

(NB : Ces situations d'erreur sont exceptionnelles. Un contrôle complet de l'interface sera réalisé en 2015 afin de s'assurer qu'elle fonctionne correctement.)

1) Dans le sous-dossier *reception*

Ouvrir le fichier *psstrate_.....csv*

Ajouter à la main la ou les strates manquantes.

Exemple, on a défini quatre strates mais seules les trois premières ont été prises en compte (le fichier ne comprend que trois lignes):

Dates;numStrate

01/01/2004 00:00;1

01/04/2004 00:00;2

01/07/2004 00:00;3

01/10/2004 00:00;4

En gras en rouge, la 4^e strate manquante est ajoutée manuellement dans le fichier .csv.

2) Dans le sous-dossier *resultats*

Ouvrir le fichier *[nom_utilisateur]_parametre.R*

Modifier le chemin d'accès des fichiers *Reference* et *PsStrate*. Exemple :

```
fichierReference ← "C:/wamp/www/test_r/reception/reference_53383938a225c.csv"
fichierPsStrate ← "C:/wamp/www/test_r/reception/psstrate_53383938a225c.csv"
```

```
fichierReference                                     <-
"C:/tempR/Pech_donnees/reception/reference_53383938a225c.csv"
fichierPsStrate                                     <- "C:/tempR/Pech_donnees
/reception/psstrate_53383938a225c.csv"
```

Ouvrir le fichier [nom_utilisateur]_script.R

Modifier les chemins d'accès selon l'exemple suivant :

Rechercher : wamp/www/test_r

Remplacer tout par : **tempR/Pech_donnees**

3) Ouvrir R

Ouvrir le fichier **C:/ tempR/Pech_donnees /resultat/[nom_utilisateur]_parametre.R**

Faire Ctrl A et Ctrl C : le contenu du script est sélectionné et copié.

Faire Ctrl V dans R

Par la suite, si l'on veut tester différents plans d'échantillonnage, il suffit de modifier les strates dans le fichier reception/psstrate_.....csv et de modifier les paramètres dans le fichier resultat/[nom_utilisateur]_parametre.R puis de relancer R comme indiqué ci-dessus.

Les résultats s'écrivent dans le fichier Pech_donnees/resultat/[nom_utilisateur]_resultat_choix..csv.

Contenu du fichier de paramètres :

```
fichierReference <- "C:/ tempR/Pech_donnees /reception/reference_53383938a225c.csv"
```

```
fichierPsStrate <- "C:/ tempR/Pech_donnees /reception/psstrate_53383938a225c.csv"
```

```
tg <- 7
```

choix <- 1 # 1 si on veut calculer le nombre de grappes par strate pour une incertitude donnée, 2 si on veut calculer l'incertitude pour un nombre donné de grappes par strate.

incRel <- "20 %" # l'incertitude imposée si choix = 1. Si on impose une incertitude dans l'unité propre du polluant, exemple 10 µg/m3, écrire "10 up". Si choix = 2, ce champ peut rester vide ("") ou avec un nombre, cela n'a pas d'importance.

mh <- c(5,4,4,4) # le nombre de grappes dans chaque strate (dans l'ordre des strates) si choix = 2. Si choix=1, ce champ peut rester vide (mh <- c()) ou avec des nombres, cela n'a pas d'importance.

interet <- "N2_BRT" #variable du fichier de référence utilisée pour définir le plan.

numero <- "malherbe" #nom d'utilisateur, à laisser tel quel.