

Traitement des non-défectés dans ND expo

Cas des expositions quotidiennes moyennes calculées à partir de plusieurs échantillons

Jérôme Lavoué

été 2014

Recommandation pratique

Pour déterminer comment saisir dans NDexpo une EQM constituée d'échantillons dont certains sont rapportés $<VMR$.

Calculer EQM_{min} en remplaçant $<VMR$ par 0 dans la formule du calcul d'EQM

Calculer EQM_{max} en remplaçant $<VMR$ par VMR dans la formule du calcul d'EQM

1. Si $EQM_{min}=0$ ou $EQM_{min}<EQM_{max}/2$

Saisir l'EQM comme non détectée dans NDexpo ($<EQM_{max}$)

2. Si $EQM_{min}>EQM_{max}/2$

Saisir la valeur $(EQM_{min}+EQM_{max})/2$ comme détectée dans NDexpo

Explication détaillée

NDexpo est un outil qui vise à permettre de remplacer des valeurs rapportées comme étant inférieure à une certaine concentration dans l'air par des valeurs qui éviteront des biais lors de l'estimation des paramètres d'un profil d'exposition.

Typiquement, on saisira dans NDexpo une série de valeurs jugées représentatives du profil d'exposition d'intérêt. Si l'on s'intéresse à la distribution d'expositions 8-heures, il faut donc saisir dans NDexpo une série de valeurs correspondant chacune à une exposition moyenne quotidienne (EQM). Certaines des valeurs EQM saisies seront des concentrations, et certaines seront du type 'inférieure à'. Si l'on s'intéresse à la distribution d'exposition de courte durée (par exemple profil d'expositions 15 min.), il faudra saisir dans NDexpo une série de valeurs d'exposition de courte durée. Globalement, il faut saisir dans NDexpo (comme on le ferait dans IHSTAT) une série de valeurs qui sont représentatives du profil que l'on cherche à caractériser.

Ce document traite spécifiquement du cas où l'on s'intéresse au profil des expositions 8-heures : l'ensemble des journées d'exposition subies par un travailleur ou groupe d'exposition. Dans ce cas, chaque valeur saisie dans NDexpo (ou IHSTAT) doit être une valeur EQM.

Or il est fréquent de déterminer une EQM au moyen de plusieurs échantillons répartis sur la journée. On détermine alors l'EQM en calculant une moyenne des différents échantillons pondérée par les durées d'échantillonnage respectives, en incluant possiblement des périodes de non exposition.

La question se pose alors : que saisir dans NDexpo lorsque le résultat de l'un des échantillons d'une EQM est rapporté comme inférieur à une certaine concentration.

Considérons l'exemple suivant :

Échantillon	Masse (mg)	Volume (m3)	Concentration (mg/m3)	Durée (min)
1	0.24	0.024	10	120
2	0.24	0.024	10	120
3	1.08	0.036	30	180
4	< 0.06	0.012	<5	60

Pour connaître l'information concernant l'EQM correspondant à ces données nous pouvons calculer une valeur maximale pour l'EQM, en remplaçant <5 par 5 dans le calcul :

$$EQM_{max} = \frac{10*120+10*120+30*180+5*60}{120+120+180+60} = 16.875 \text{ mg/m}^3$$

Nous pouvons également calculer une valeur minimale pour l'EQM en remplaçant <5 par 0 :

$$EQM_{min} = \frac{10*120+10*120+30*180+0*60}{120+120+180+60} = 16.250 \text{ mg/m}^3$$

L'information que nos échantillons fournissent sur l'EQM est donc synthétisée par l'information suivante :

La valeur EQM se situe entre 16.250 et 16.875 mg/m3

Or, NDexpo, ainsi que toutes les approches actuelles pour le traitement des valeurs non détectées, ne considère que la borne supérieure de l'intervalle de possibilité pour EQM : EQM <16.875 mg/m3

Dans notre exemple, nous remarquons que cela représente une perte très importante d'information.

Considérons un jeu de 4 EQM accompagnant l'EQM précédente, à partir desquelles on souhaite estimer le 95^{ème} percentile du profil :

7.932 – 14.331 – 60.060 – 18.930

Notre série complète est donc :

7.932 – 14.331 – 60.060 – 18.930 – [16.250 ; 16.875]

Si l'on saisit cette série dans NDexpo, avec < 16.875 pour l'EQM problématique, NDexpo propose de la remplacer par 10.180 pour des calculs ultérieurs dans. Cet exemple illustre la conséquence de la perte d'information causée par le fait qu'on ne considère pas la limite inférieure pour l'EQM. NDexpo et toutes les autres méthodes supposent que la limite inférieure est 0.

Solutions

Solution idéale

La solution idéale serait une approche où l'on peut saisir une information sous forme d'intervalle : $[EQM_{\min}-EQM_{\max}]$. Comme déjà souligné cela n'existe pas pour l'instant. Cependant, ce type de calcul est possible par approche bayésienne, et sera inclus à terme, sur le site www.expostats.ca.

Solution pragmatique

Le problème vient du fait que les approches actuelles (incluant NDexpo) ne considèrent que des données du type $[0-EQM_{\max}]$.

- A. Lorsqu'on se trouve dans une situation où EQM_{\min} est effectivement 0 (par exemple tous les échantillons sont $<VMR$, ou encore l'EQM ne vient que d'un seul échantillon 8-h qui est $<VMR$), les approches existantes sont adéquates.
- B. Lorsqu'on se trouve dans une situation où EQM_{\min} est non nulle mais proche de 0 par rapport à EQM_{\max} (par exemple $EQM_{\min}=3$ et $EQM_{\max}=16$), les approches existantes sont également probablement adéquates.
- C. Lorsqu'on se trouve dans une situation où EQM_{\min} est éloigné de 0 et proche de EQM_{\max} (par exemple : $EQM_{\min}=12$ et $EQM_{\max}=16$), je pense qu'une solution pragmatique est d'utiliser comme valeur EQM finale le centre de l'intervalle $[EQM_{\min}-EQM_{\max}]$ ($EQM=14$ dans l'exemple $[12-16]$). Cette solution revient à remplacer la valeur $<x$ par $x/2$.

Il ne sera pas toujours évident de juger si une situation particulière est à attribuer au cas B ou au cas C ci-dessus. Voici une recommandation ad-hoc pour faire la distinction :

Si $EQM_{\min} < EQM_{\max}/2$: on rentre $<EQM_{\max}$ dans NDexpo (définition arbitraire du cas B)

Si $EQM_{\min} > EQM_{\max}/2$: on rentre $(EQM_{\min} + EQM_{\max})/2$ (définition arbitraire du cas C)