Société de Calcul Mathématique, S. A.
Algorithmes et Optimisation

Incertitudes sur les émissions atmosphériques calculées par des méthodes de désagrégation

Mise en œuvre sur un exemple

Rapport adressé au

CITEPA

par la

Société de Calcul Mathématique SA

Ce rapport prend en compte les remarques effectuées lors de la réunion du 29/07/2009

En application de notre contrat du 8 avril 2009

Rédaction : Hélène Bickert

Septembre 2009

Siège social et bureaux : 111, Faubourg Saint Honoré, 75008 Paris. Tel. : 01 42 89 10 89. Fax : 01 42 89 10 69
Société Anonyme au capital de 56 200 Euros. RCS : Paris B 399 991 041. SIRET : 399 991 041 00035. APE : 7219Z

Résumé

Les quantités de polluants, les consommations énergétiques sont généralement mesurées sur de larges étendues (tout un pays) ou sur de larges intervalles temporels (à l'année). Or, l'application de certaines directives nécessite de connaître la valeur de ces grandeurs sur des zones géographiques plus fines, ou sur des périodes plus courtes. Lorsque les mesures spécifiques ne sont pas disponibles, on réalise une désagrégation des grandeurs à l'aide d'un indicateur local. La désagrégation peut être spatiale, temporelle, ou les deux.

La question se pose : que deviennent les incertitudes globales, sur le niveau du polluant, lors de cette désagrégation ?

Dans le rapport n°1, nous indiquons une méthode pour construire les lois de probabilité des grandeurs locales obtenues par désagrégation, dans trois situations dépendant de l'état de connaissance des lois de probabilité des grandeurs.

Nous appliquons cette méthode à un exemple de désagrégation spatio-temporelle choisie par le CITEPA : la quantité de solvant contenue dans la peinture utilisée dans la construction de bateaux. La quantité est connue pour l'année 2004 et pour toute la France ; il s'agit de la désagréger :

- Par commune, en fonction des effectifs de la branche “construction de bateaux” ;

- Par jour (ou par heure) de l'année, en fonction des activités journalières (ou horaires) de cette même branche.

- Par commune et par jour (ou par heure) de l'année.

On se place dans la situation 3 définie dans le précédent rapport : la grandeur à désagréger et les indicateurs locaux sont des variables aléatoires de lois connues. Pour chaque grandeur, nous connaissons la valeur moyenne et l'incertitude associée. Nous faisons ensuite l'hypothèse que la variable aléatoire suit une loi uniforme ou normale, dont les paramètres sont définis à partir de la valeur moyenne et l'incertitude.

La méthode tient compte de l'erreur attendue suite à l'utilisation des indicateurs.

Sommaire

I. Problématique

Les quantités de polluants générées par diverses activités (transport, industries, etc.) sont souvent déterminées sur des intervalles temporels larges (année, mois), et de larges étendues (pays tout entier, région, etc.). Or l'application de certaines directives nécessite de connaître ces grandeurs soit sur des zones géographiques plus fines, soit sur des périodes plus courtes.

Comme les informations spécifiques ne sont généralement pas disponibles, il est nécessaire de "désagréger" les données obtenues pour une large zone, ou pour une longue durée : le résultat de la désagrégation est une donnée locale permettant d'estimer les émissions des polluants, pour chaque zone géographique ou chaque période. Cette désagrégation se fait à l'aide d'un indicateur, dont les valeurs locales sont connues. Ce peut être la population, ou le nombre d'employés d'industries de la zone pour une désagrégation spatiale ; des données climatiques pour une désagrégation temporelle.

De plus, la désagrégation peut être de plus ou moins bonne qualité, selon l'indicateur choisi. On sait que le résultat obtenu est plus ou moins proche de la valeur réelle : on s'attend à une certaine erreur.

La question se pose : que deviennent les incertitudes globales lors de cette désagrégation ? Comment prendre en compte la qualité de la désagrégation ?

Les utilisateurs de données dans le domaine environnemental réclament, de plus en plus, des évaluations des incertitudes : le résultat numérique seul ne suffit plus. De plus, comme ces questions sont souvent objet de discussions, voire de polémiques liées aux intérêts économiques, la qualité de la méthode est essentielle.

II. Mise en œuvre sur un exemple

L'objectif est de désagréger la quantité de solvant contenue dans la peinture utilisée dans la construction de bateaux. La quantité est connue pour l'année 2004 et pour toute la France ; il s'agit de la désagréger :

- Par commune, en fonction des effectifs du secteur d'activité “construction de bateaux” ;

- Par jour (ou par heure) de l'année, en fonction des activités journalières (ou horaires) de ce même secteur d'activité.

- Par commune et par jour (ou par heure) de l'année.

La quantité de solvant, notée Q est connue sous la forme d'une valeur moyenne et d'une incertitude associée. Pour chaque indicateur, nous connaissons la valeur moyenne, l'incertitude associée, ainsi que l'erreur attendue en utilisant l'indicateur pour réaliser la désagrégation.

Lorsqu'on cherche à désagréger une grandeur, trois situations peuvent se présenter, selon la connaissance des grandeurs et de leurs lois :

- Situation 1 : la grandeur à désagréger est connue sans incertitude et on n'a pas d'informations sur les lois des indicateurs locaux ;

- Situation 2 : la grandeur à désagréger est une variable aléatoire de loi connue et on n'a pas d'information sur les lois des indicateurs locaux ;

- Situation 3 : la grandeur à désagréger et les indicateurs locaux sont des variables aléatoires de lois connues.

Le précédent rapport donne la méthodologie à mettre en œuvre pour réaliser une désagrégation dans les trois cas. Dans cet exemple, nous nous plaçons dans la situation 3 : la grandeur à désagréger et les indicateurs locaux sont des variables aléatoires de lois connues. Nous devons donc introduire des hypothèses sur ces lois.

L'objectif est d'évaluer l'incertitude sur la valeur de chaque quantité désagrégée ; elle est donnée par sa loi de probabilité.

Trois types de désagrégation sont réalisables :

- Désagrégations simples : par commune (désagrégation spatiale) ou par jour de l'année (désagrégation temporelle) ;

- Désagrégations doubles : par commune puis par jour de l'année (désagrégation spatio-journalière) ou par jour puis par heure de l'année (désagrégation horaire) ;

- Désagrégation triple : par commune, par jour puis par heure de l'année (désagrégation spatio-horaire).

1. Désagrégations simples

Nous noterons :

- Q la grandeur globale, qu'il s'agit de désagréger ;
- K le nombre de « zones » (communes ou jours) en lesquels on veut désagréger la grandeur ;
- Z_k, $k = 1, \dots, K$ l'indicateur local correspondant à la $k$ème zone de désagrégation ;
- $Z = \sum_{k=1}^{K} Z_k$ la somme des indicateurs locaux.

La désagrégation de la grandeur globale Q, réalisée au moyen des indicateurs locaux Z_k est donnée par la formule suivante :

$$Q_k = Q \times \frac{Z_k}{Z} \quad (1)$$

Cette formule a un sens bien clair lorsque Q, Z et Z_k sont des valeurs précises, ce qui est rarement le cas dans la pratique. On les considère donc comme des variables aléatoires ; de ce fait, les désagrégations Q_k seront elles-mêmes des variables aléatoires, dont on cherche la loi.

Lois de probabilité des grandeurs

Comme on l'a vu, les lois ne sont pas connues : pour chaque grandeur, nous ne connaissons que la valeur moyenne et l'incertitude associée. Nous faisons l'hypothèse que la variable aléatoire suit une loi uniforme ou normale, dont les paramètres sont définis à partir de la valeur moyenne et l'incertitude. Notons que ces choix sont tout à fait arbitraires.

Considérons un indicateur Z_k. Notons m_k la valeur moyenne et inc_k l'incertitude associée. Elle est donnée sous la forme d'un pourcentage.

Dans le cas d'une loi uniforme, les bornes de la densité de probabilité sont les suivantes : $]m_k - inc_k \times m_k ; m_k + inc_k \times m_k[$. Ceci suffit à définir la loi de Z_k :

$$f_k(x) = \begin{cases} \dfrac{1}{2inc_k \times m_k}, & x \in]m_k - inc_k \times m_k ; m_k + inc_k \times m_k[\\ 0, & sinon \end{cases}$$

Pour utiliser une loi normale, il faut définir son espérance μ_k et son écart-type σ_k. Nous prenons $\mu_k = m_k$ pour l'espérance. Nous choisissons de définir l'écart-type σ_k comme suit :

$$\sigma_k = inc_k \times m_k/2$$

En effet, 95% de la densité de probabilité d'une gaussienne est contenu dans l'intervalle $[\mu_k - 2\sigma_k; \mu_k + 2\sigma_k]$.

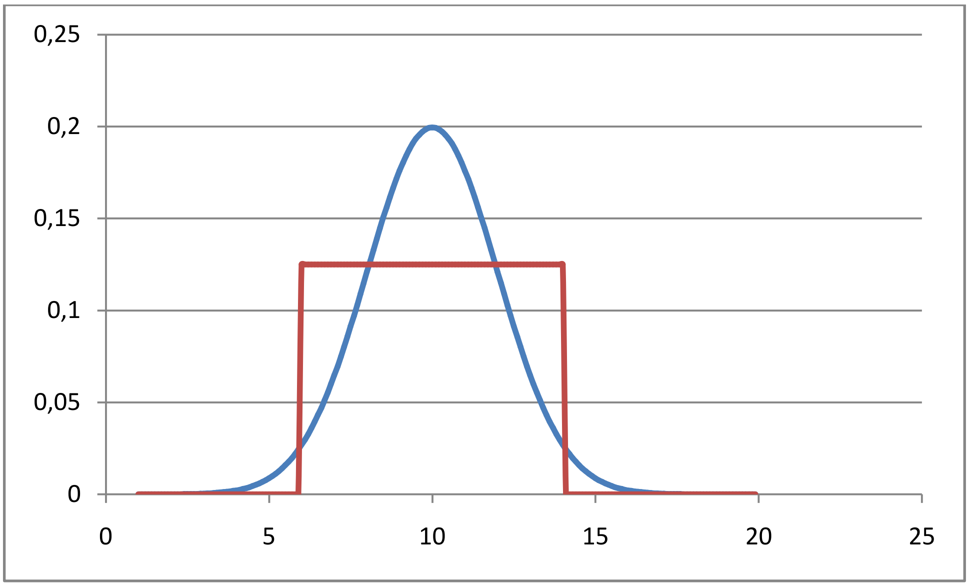


Figure 1 : Exemple de superposition d'une loi normale m=10, sigma=2 et d'une loi uniforme

Nous pouvons ainsi construire les lois de probabilité de Q et de chaque indicateur local Z_k. Il nous faut à présent déterminer celle de la somme des indicateurs Z.

Lorsque les indicateurs locaux Z_k suivent des lois normales, leur somme est également une variable gaussienne, dont les paramètres se déterminent comme suit :

$$m = \sum_k m_k$$

$$\sigma^2 = \sum_k {\sigma_k}^2$$

Ceci est vrai car il s'agit de lois gaussiennes ; ce n'est pas le cas pour des lois uniformes. Dans ce cas, nous déterminons la loi de probabilité de Z par une simulation de type Monte-Carlo : ceci consiste à réaliser un certain nombre de tirages aléatoires de chaque Z_k. A chaque run, nous calculons la somme des variables aléatoires. Nous construisons ensuite la loi de probabilité de Z en prenant l'histogramme de l'échantillon de Z ainsi obtenu. Dans ce cas, la loi de probabilité de Z est discrète.

Prise en compte de l'erreur attendue

On se place dans le cas où les indicateurs locaux Z_k sont des variables gaussiennes.

A ce stade, les écart-types σ_k ne représentent que l'incertitude de chaque indicateur, c'est-à-dire le fait que la grandeur est bien connue ou non.

Outre les incertitudes propres à l'indicateur, nous introduisons la notion de justesse d'un indicateur. Lorsqu'on utilise un indicateur pour réaliser une désagrégation, on sait que le résultat obtenu sera plus ou moins proche de la valeur réelle : on s'attend à une certaine erreur.

Mettons que l'erreur attendue est de 20 %. Cela signifie que la valeur réelle est estimée être contenue dans un intervalle de ± 20 % autour de la valeur calculée par désagrégation. Cette erreur est une donnée relative, que nous notons e_k.

Remarque : cette notation sous-entend que l'erreur attendue peut varier d'une zone à l'autre. En effet, il se peut très bien que la désagrégation par l'indicateur soit relativement bonne sur certaines zones et mauvaise sur d'autres.

On a la relation suivante :

$$(1 - e_k) \times Q \frac{Z_k}{Z} \leq Q_k \leq (1 + e_k) \times Q \frac{Z_k}{Z}$$

Soit, pour une erreur de 20% :

$$0.8 \times Q \frac{Z_k}{Z} \leq Q_k \leq 1.2 \times Q \frac{Z_k}{Z}$$

Introduisons Z'_k les lois normales définies par :

$$m'_k = m_k$$

$$\sigma'_k = e_k + \sigma_k$$

Et $Z' = \sum_k Z'_k$.

Les lois des Z'_k et de Z' sont une dilatation des lois de Z_k et de Z, prenant en compte l'erreur attendue : on "grossit" l'écart-type des lois de manière à rendre compte de l'erreur due à la désagrégation.

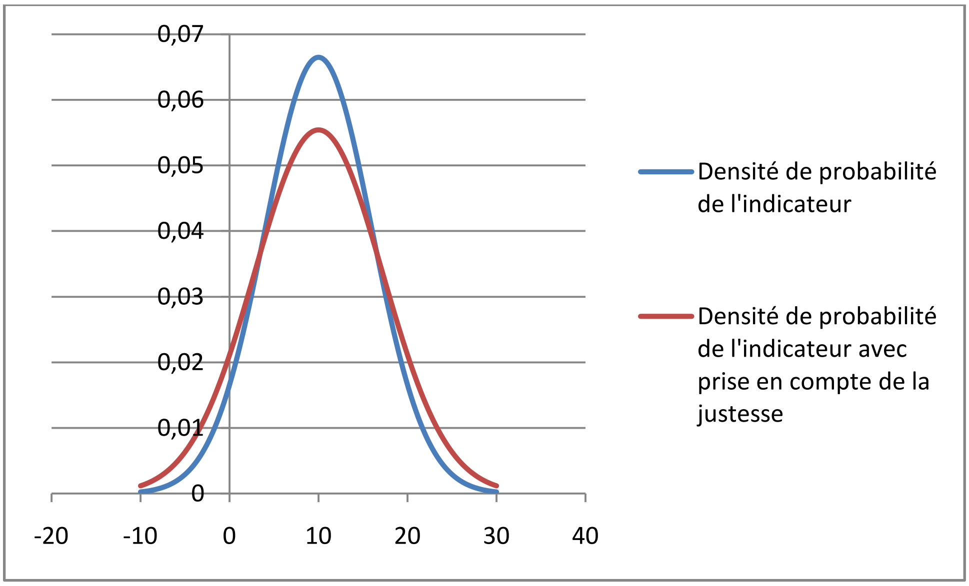


Figure 2 : Exemple de prise en compte de la justesse d'un indicateur

On réalise alors la désagrégation de Q à l'aide l'indicateur Z'_k :

$$Q_k = Q\frac{Z'_k}{Z'} \quad (4)$$

On a bien :

$$\sum_{k=1}^{K} Q_k = \sum_{k=1}^{K} Q\frac{Z'_k}{Z'} = Q$$

La prise en compte de l'erreur se fait suivant le même principe dans le cas de lois uniformes.

Désagrégation

Il s'agit de déterminer la loi de probabilité de Q_k, pour tout k.

Le calcul explicite de cette loi est très lourd : il faut réaliser le calcul de deux produits de convolution. C'est pourquoi nous utilisons une simulation de Monte-Carlo : on réalise N tirages aléatoires de Z'_k, Z' et Q en fonction de leurs lois, et on détermine la loi de probabilité de Q_k par la formule (4). La loi est obtenue en construisant l'histogramme des résultats des N tirages. Il s'agit d'une loi discrète.

Or, le CITEPA n'a besoin de connaître que l'espérance de Q_k et son incertitude.

L'espérance de la loi se calcule simplement :

$$m_{Q_k} = \sum_i qk_i p_i$$

où les qk_i représentent les valeurs prises par Q_k et p_i les probabilités associées.

Pour caractériser l'incertitude, nous utilisons les quantiles à 5% et 95% : la probabilité que Q_k soit entre ces deux valeurs est de 90%.

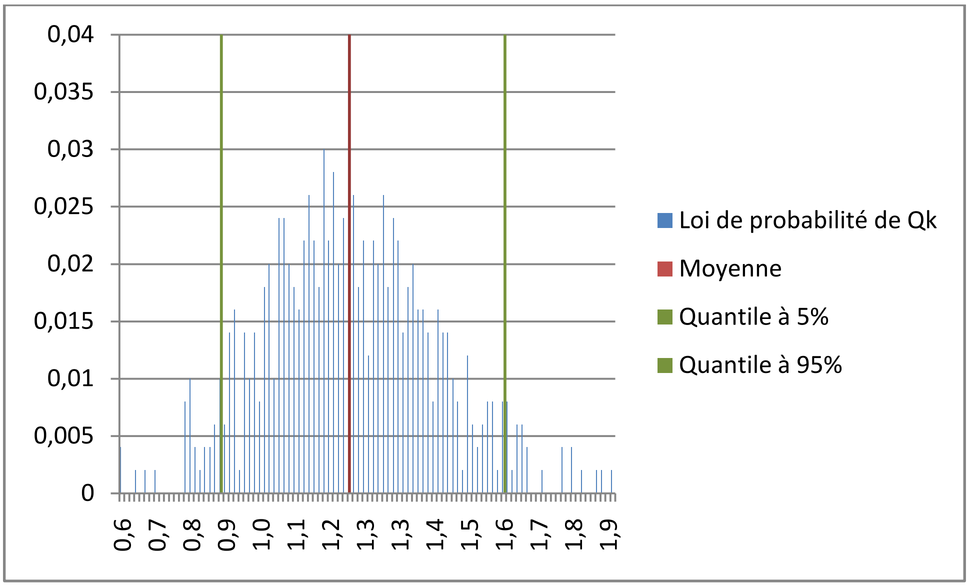


Figure 3 : exemple de loi de probabilité de Qk (désagrégation spatiale, commune 01065)

L'incertitude est ensuite caractérisée par le maximum de l'écart relatif à l'espérance de ces deux valeurs. L'outil indique les valeurs des deux quantiles et de l'incertitude ainsi calculée.

Dans l'exemple ci-dessus, la moyenne de la quantité de solvant utilisée dans la commune 01065 en 2004 est de 1.23 tonnes de solvant. L'incertitude associée à cette valeur est de 33%.

2. Désagrégations doubles

Il s'agit de désagréger la quantité de solvant par commune et par jour de l'année, ou bien par jour et heure de l'année. Plaçons-nous dans le cas d'une désagrégation spatio-temporelle.

On commence par réaliser une désagrégation spatiale, en suivant la méthode décrite dans le paragraphe précédent. A l'issue de cette étape, nous disposons des lois de probabilité de Q_k, $k = 1 \dots K$. Il s'agit de lois discrètes.

On réalise ensuite une seconde désagrégation, à l'aide des indicateurs journaliers. Notons :

- L le nombre de jours ($L = 366$ dans notre exemple) ;
- T_l, $l = 1, \dots, L$ l'indicateur local correspondant au $l$ème jour ;

- $T = \sum_{l=1}^{L} T_l$ la somme des indicateurs journaliers locaux.

De la même façon que pour la désagrégation simple, on détermine les lois de probabilité de T_l' et T' pour la prise en compte de l'erreur attendue.

On réalise la désagrégation de la quantité de solvant par commune Q_k à l'aide des indicateurs journaliers T_l' :

$$Q_{k,l} = Q_k \frac{T'_l}{T'} \quad (5)$$

Cette désagrégation se fait par une simulation de Monte-Carlo.

On obtient ainsi la loi de probabilité de la quantité de solvant $Q_{k,l}$ par commune k et par jour l. On en déduit la moyenne et l'incertitude à l'aide des quantiles à 5% et 95%.

3. Désagrégations triples

Il s'agit de désagréger la quantité globale de solvant par commune et par heure de l'année.

On commence par réaliser une désagrégation spatio-journalière. On dispose alors des lois de probabilité de la quantité de solvant $Q_{k,l}$ pour chaque commune k et chaque jour l.

Notons :

- M le nombre d'heures par jour ($H = 24$) ;

- H_m, $m = 1, \dots, M$ l'indicateur local correspondant à la $m^{ème}$ heure ;

- $H = \sum_{m=1}^{M} H_m$ la somme des indicateurs horaires locaux.

De la même façon que pour la désagrégation simple, on détermine les lois de probabilité de H_m' et H' pour la prise en compte de l'erreur attendue.

On réalise la désagrégation de la quantité de solvant par commune et par jour $Q_{k,l}$ à l'aide des indicateurs journaliers H_m' :

$$Q_{k,l,m} = Q_{k,l} \frac{H'_m}{H'} \quad (5)$$

Cette désagrégation se fait par une simulation de Monte-Carlo.

On obtient ainsi la loi de probabilité de la quantité de solvant $Q_{k,l,m}$ par commune k, par jour l et par heure m. On en déduit la moyenne et l'incertitude à l'aide des quantiles à 5% et 95%.

III. Utilisation de l'outil

Nous avons mis en place un outil sous VBA qui permet le calcul des différentes désagrégations. Les données nécessaires sont la donnée à désagréger, les indicateurs par commune, par jour et par heure ; elles sont stockées dans des feuilles Excel.

L'outil utilise les données contenues dans les onglets « Quantite_a_desagreger », « indicateurs_spatiaux », « profil_journalier » et « profil_horaire ». Ces feuilles contiennent les données relatives aux indicateurs utilisés pour la désagrégation : nom de la zone (ou du jour), valeur moyenne (effectifs ou part journalière), incertitude et erreur attendue. Le nombre de zones est modifiable. Le nom des onglets peut être modifié : il faut toutefois veiller à ne pas déplacer ces onglets.

L'outil ne peut être lancé si des données manquent. Dans ce cas, un message d'erreur s'affiche, invitant l'utilisateur à remplir correctement les données.

Le lancement de l'outil est décrit ci-dessous.

Cliquer sur le bouton « Lancer les calculs » de l'onglet « Quantite_a_desagreger » :

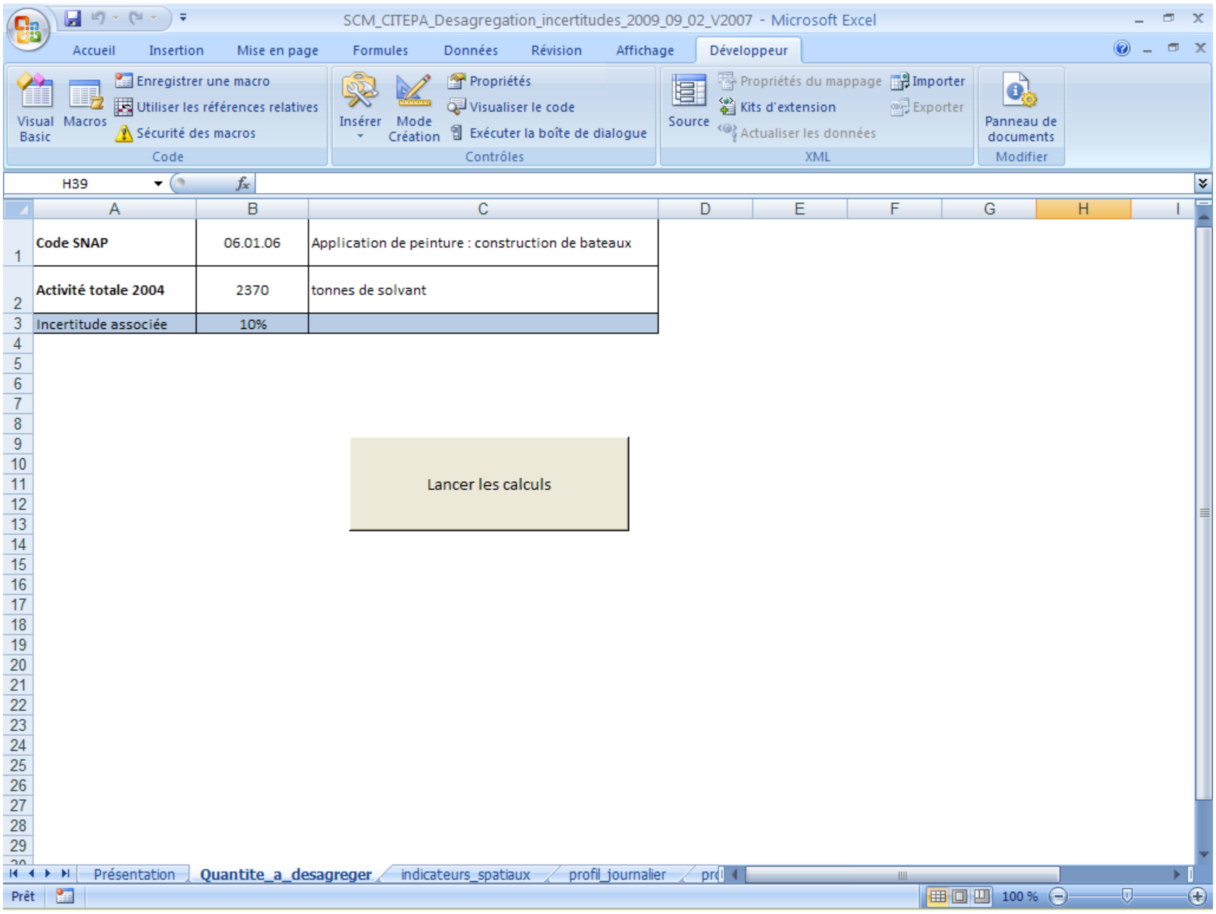


Figure 4 : lancement des calculs

Une fenêtre s'affiche. Elle permet de sélectionner le type de désagrégation (spatiale, temporelle ou spatio-temporelle) et le type de lois (lois normales ou lois uniformes).

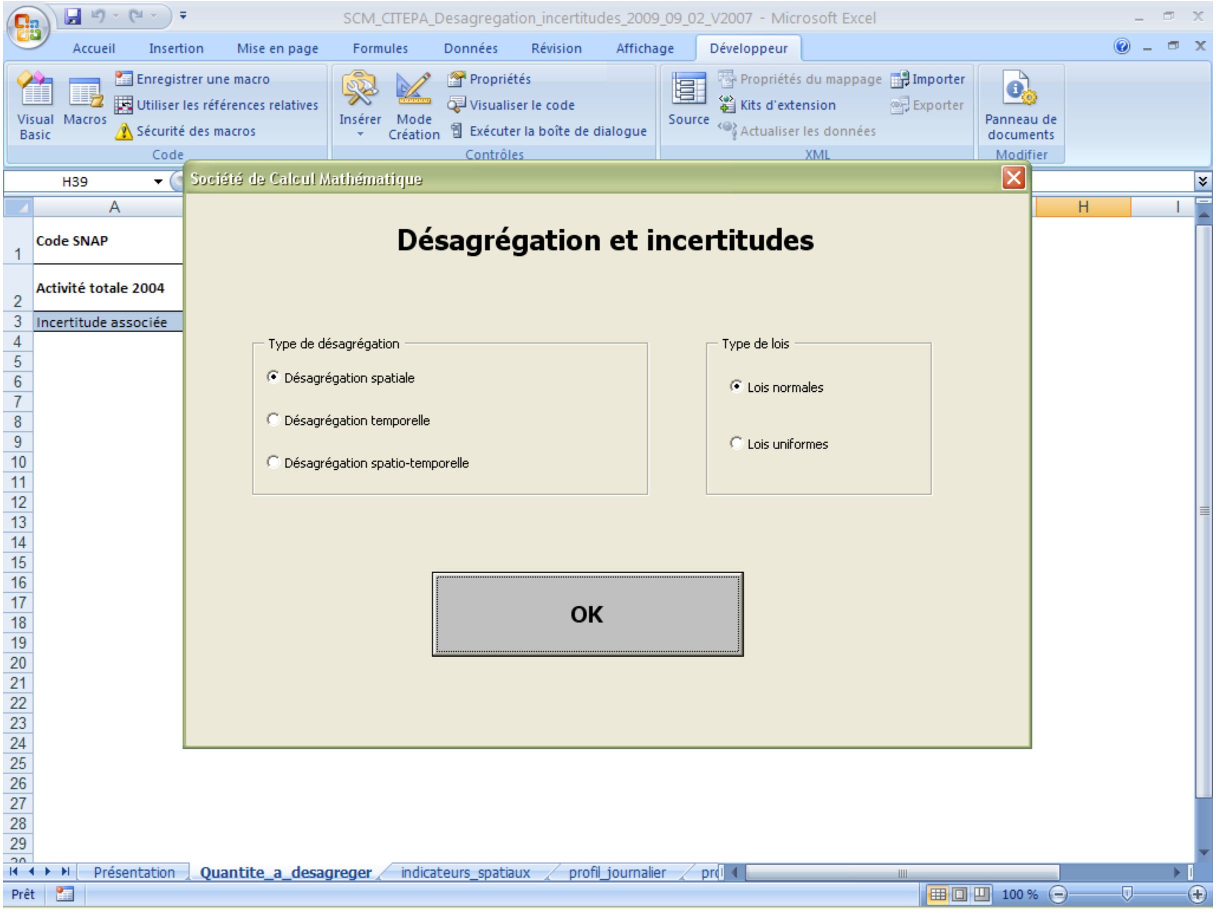


Figure 5 : choix du type de désagrégation et du type de lois

Une fenêtre de confirmation s'affiche.

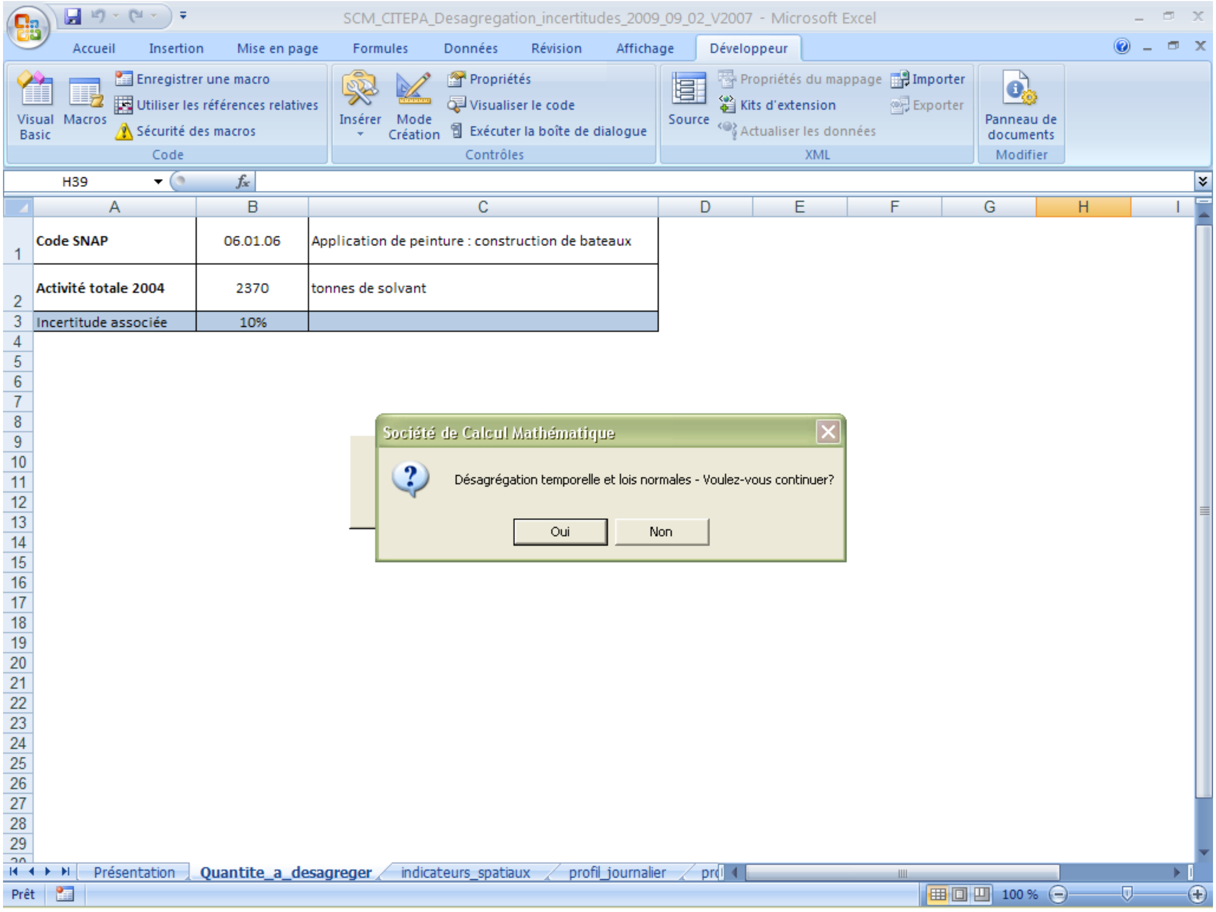


Figure 6 : fenêtre de confirmation

Dans le cas d'une désagrégation temporelle ou spatio-temporelle, une seconde fenêtre s'affiche : elle permet de choisir une désagrégation journalière ou horaire.

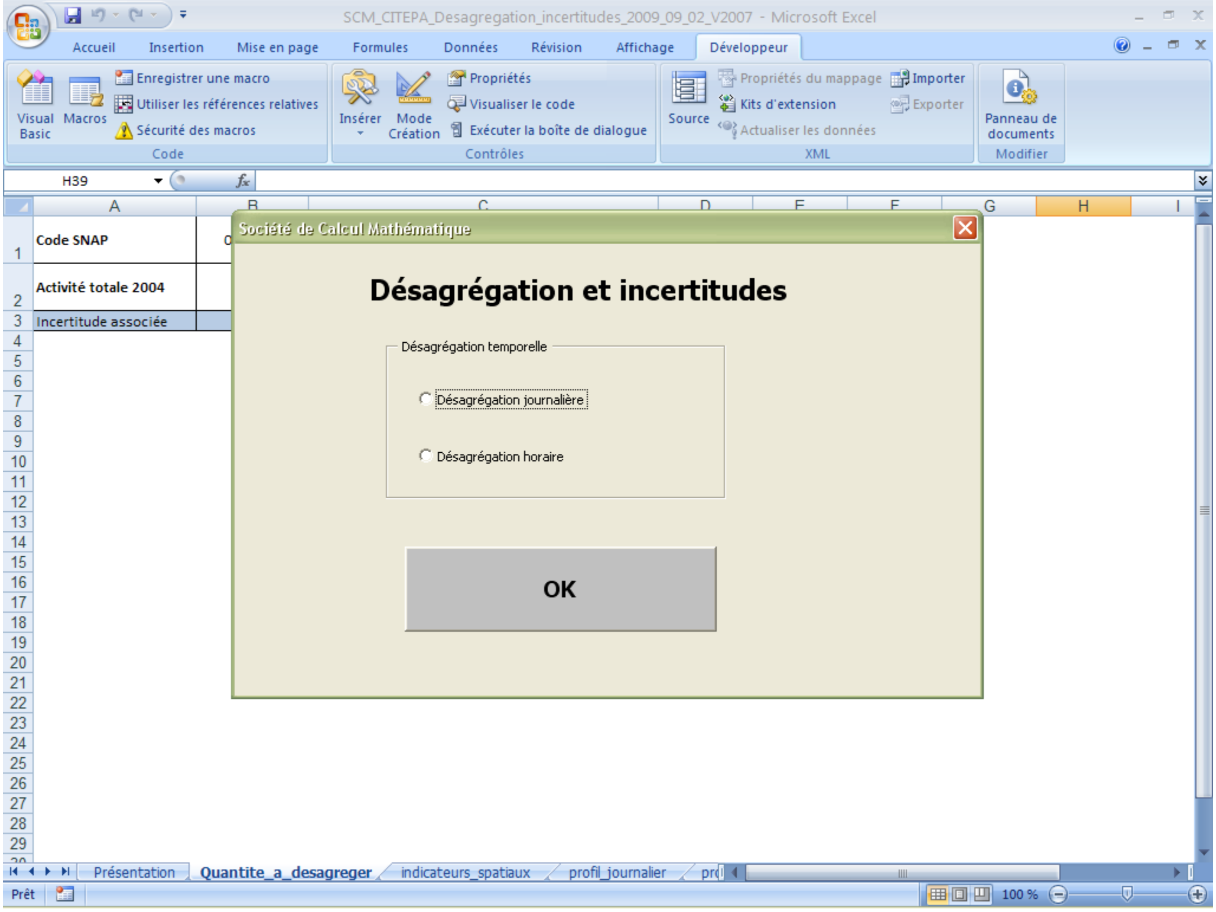


Figure 7 : choix du type de désagrégation dans le cas d'une désagrégation temporelle

Une fois le choix fait, une fenêtre de validation s'affiche.

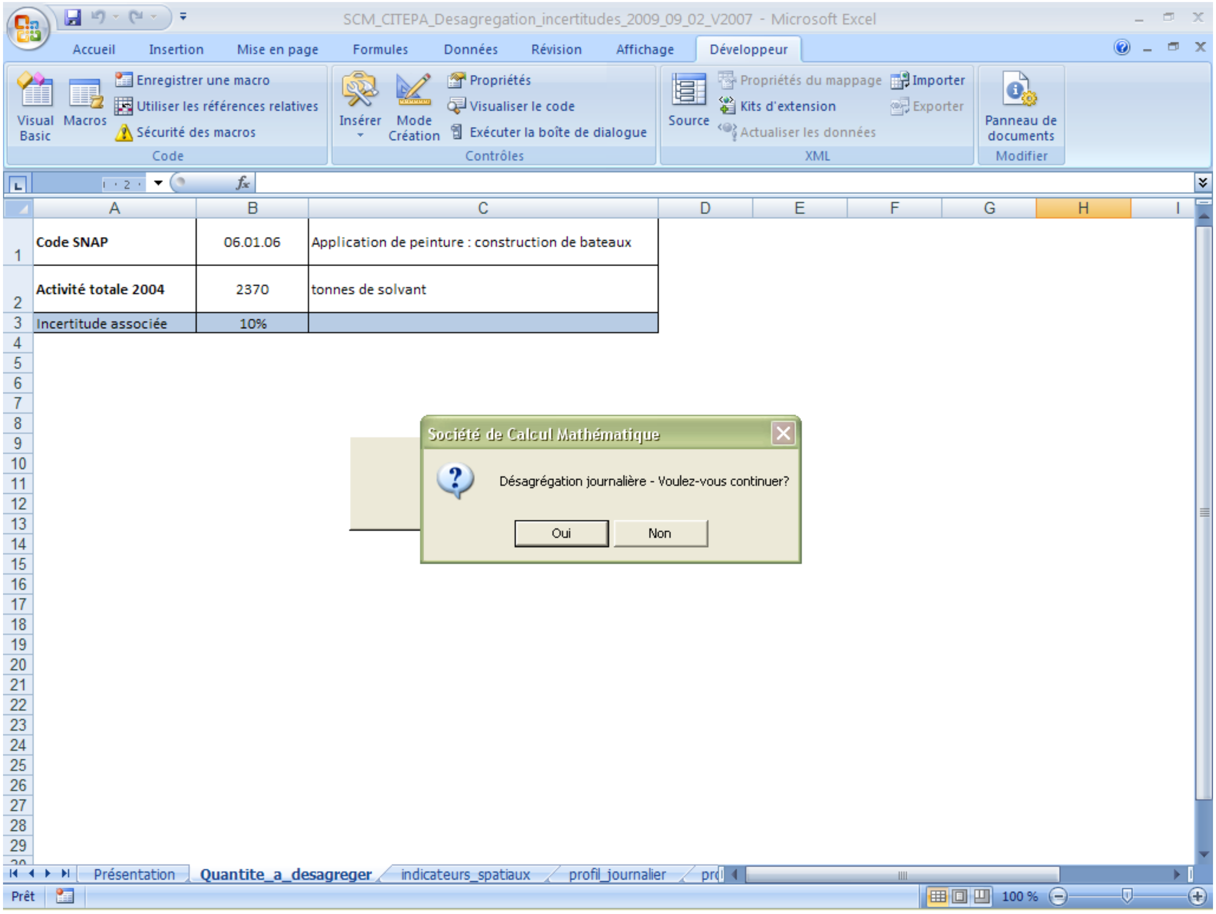


Figure 8 : fenêtre de confirmation

Une dernière fenêtre s'affiche : elle permet de sélectionner le nombre de tirages à réaliser pour les simulations de Monte-Carlo.

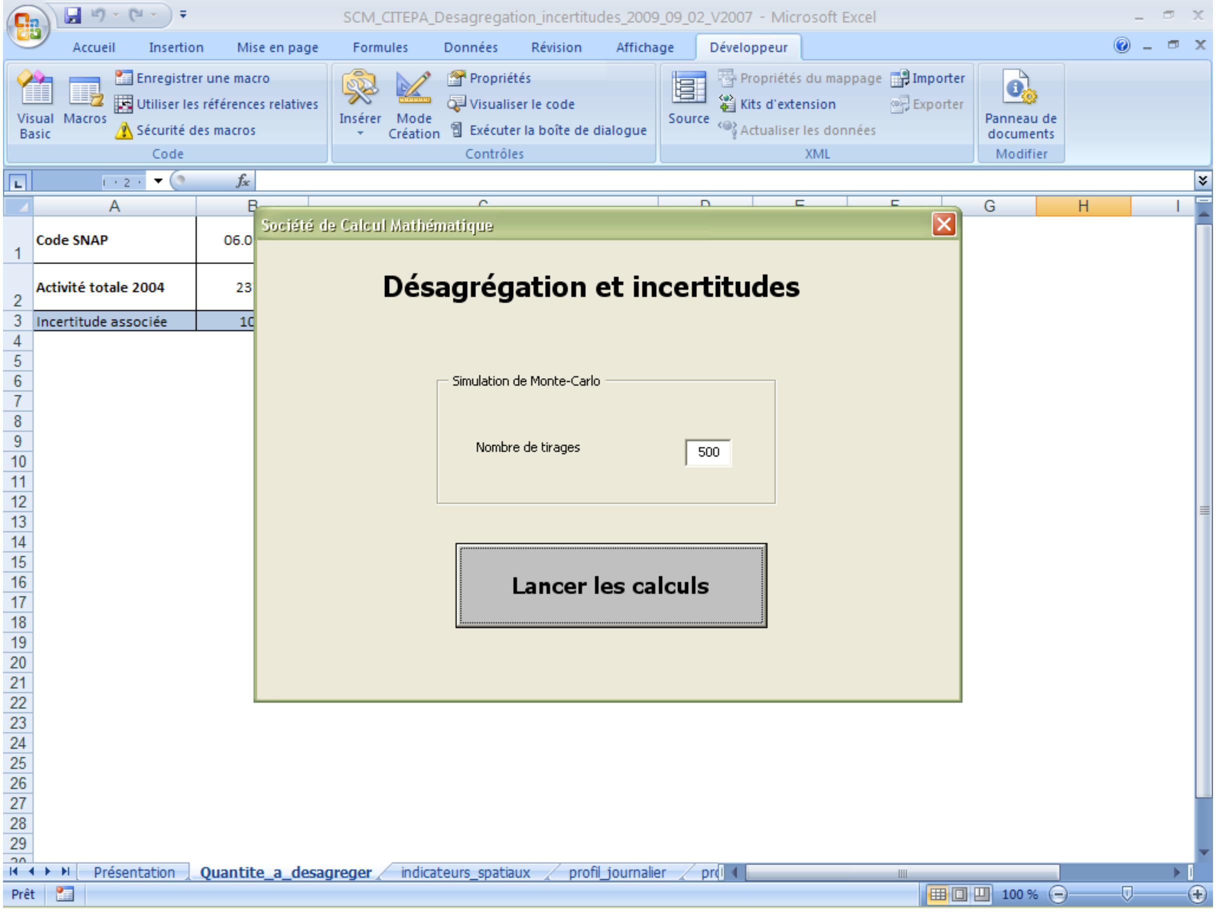


Figure 9 : choix du nombre de tirages pour les simulations de Monte-Carlo

Cliquer sur le bouton « Lancer les calculs ». Une fois les calculs effectués, une fenêtre indique le temps d'exécution du programme.

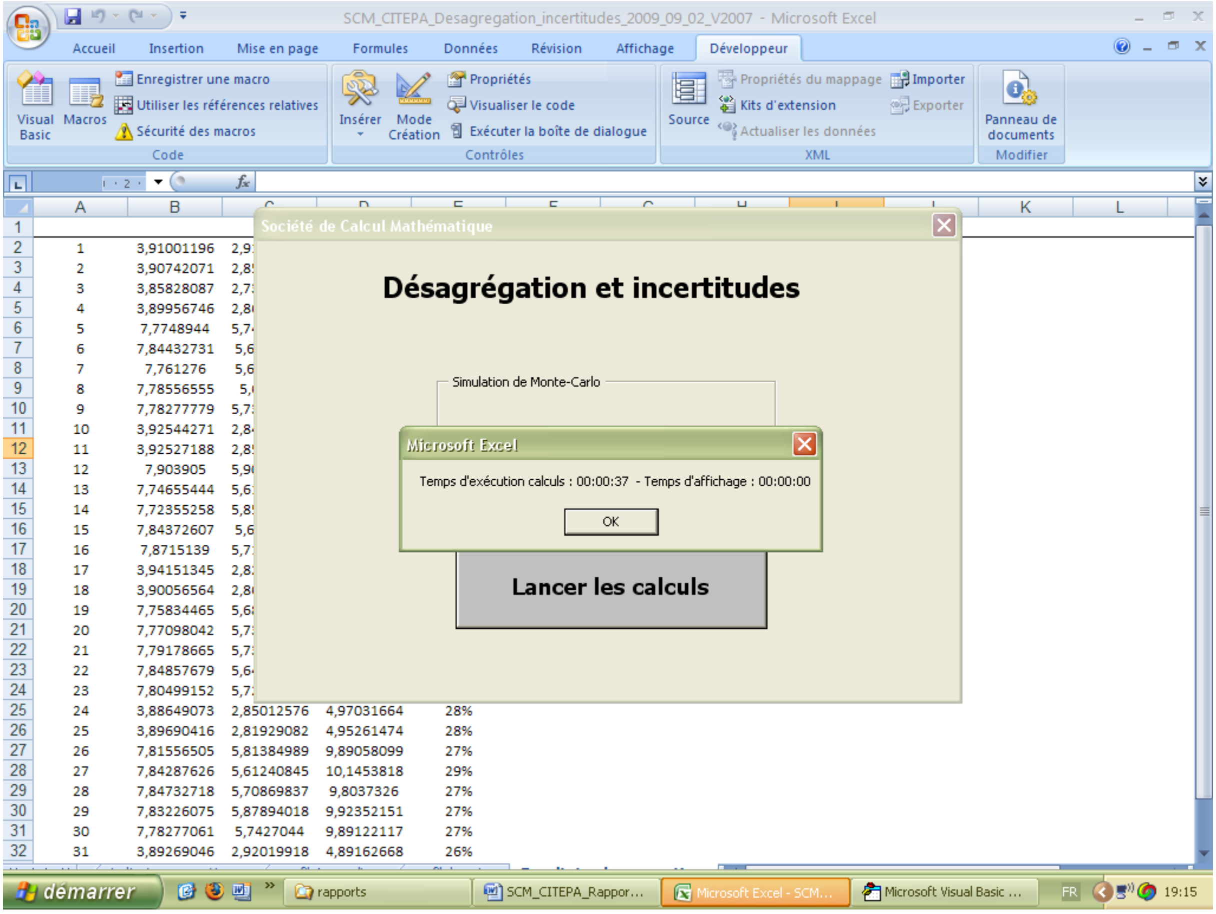
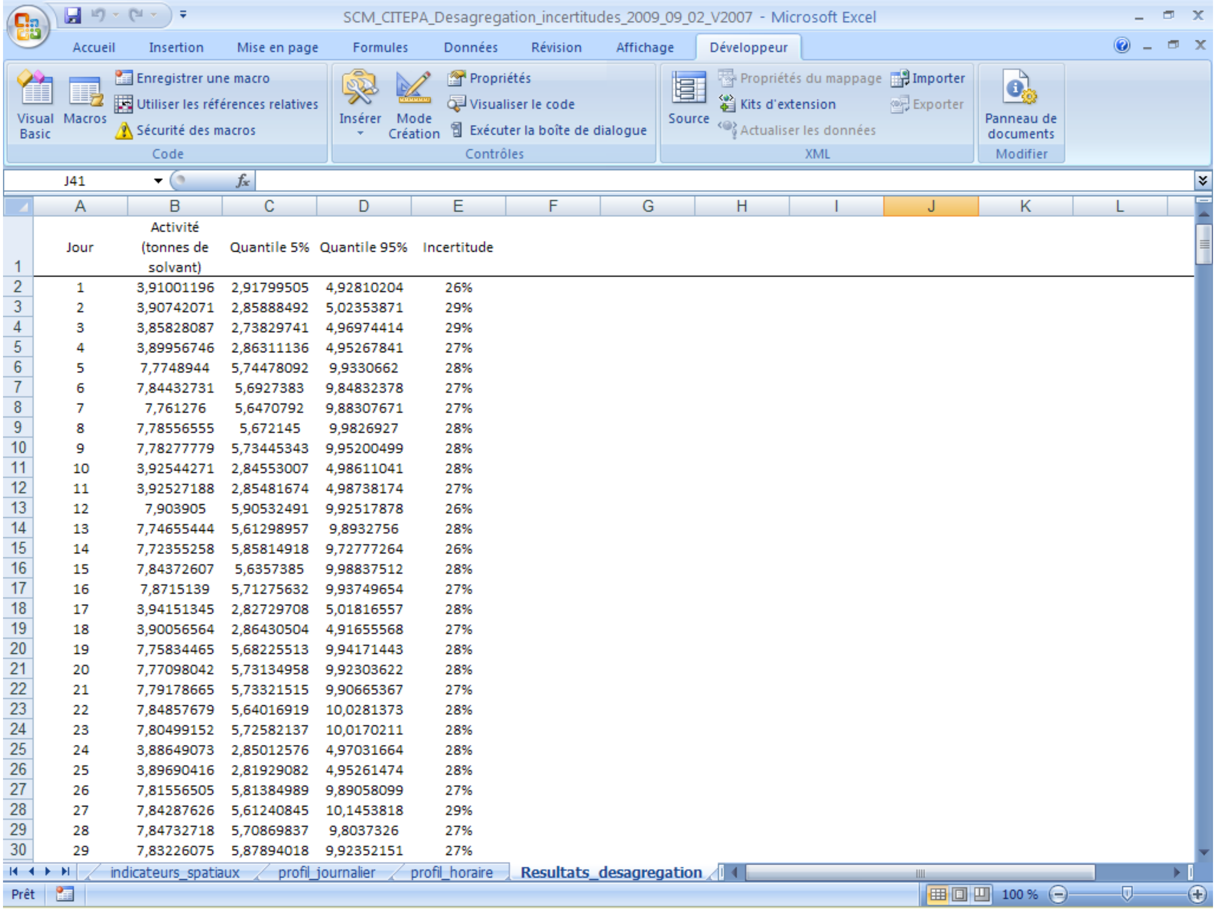


Figure 10 : fin des calculs

Les résultats sont affichés dans la feuille « Résultats_désagrégation ». Dans le cas d'une désagrégation spatio-horaire, les résultats sont contenus dans les feuilles « Résultats_désagrégation » à « Résultats_désagrégation4 ».



Jour	Activité (tonnes de solvant)	Quantile 5%	Quantile 95%	Incertitude
1	3,91001196	2,91799505	4,92810204	26%
2	3,90742071	2,85888492	5,02353871	29%
3	3,85828087	2,73829741	4,96974414	29%
4	3,89956746	2,86311136	4,95267841	27%
5	7,7748944	5,74478092	9,9330662	28%
6	7,84432731	5,6927383	9,84832378	27%
7	7,761276	5,6470792	9,88307671	27%
8	7,78556555	5,672145	9,9826927	28%
9	7,78277779	5,73445343	9,95200499	28%
10	3,92544271	2,84553007	4,98611041	28%
11	3,92527188	2,85481674	4,98738174	27%
12	7,903905	5,90532491	9,92517878	26%
13	7,74655444	5,61298957	9,8932756	28%
14	7,72355258	5,85814918	9,72777264	26%
15	7,84372607	5,6357385	9,98837512	28%
16	7,8715139	5,71275632	9,93749654	27%
17	3,94151345	2,82729708	5,01816557	28%
18	3,90056564	2,86430504	4,91655568	27%
19	7,75834465	5,68225513	9,94171443	28%
20	7,77098042	5,73134958	9,92303622	28%
21	7,79178665	5,73321515	9,90665367	27%
22	7,84857679	5,64016919	10,0281373	28%
23	7,80499152	5,72582137	10,0170211	28%
24	3,88649073	2,85012576	4,97031664	28%
25	3,89690416	2,81929082	4,95261474	28%
26	7,81556505	5,81384989	9,89058099	27%
27	7,84287626	5,61240845	10,1453818	29%
28	7,84732718	5,70869837	9,8037326	27%
29	7,83226075	5,87894018	9,92352151	27%

Figure 11 : Résultats dans un exemple de désagrégation journalière et lois normales

Affichage des résultats

Dans le tableau ci-dessous, nous indiquons le nombre de feuilles nécessaires à l'affichage des résultats.

Type de désagrégation	Nombre de zones de désagrégation	Nombre de feuilles en Excel 2003	Nombre de feuilles en Excel 2007
Spatiale	464	1 feuille	1 feuille
Journalière	366	1 feuille	1 feuille
Horaire	$366 * 24 = 8\,784$	1 feuille	1 feuille
Spatio-journalière	$464 * 366 = 169\,824$	3 feuilles	1 feuille
Spatio-horaire	$464 * 366 * 24 = 4\,075\,776$	63 feuilles	4 feuilles

Tableau 1 : Nombre de feuilles nécessaires à l'affichage des résultats

Compte-tenu du nombre élevé de données, le calcul d'une désagrégation spatio-horaire n'a été codé que pour Excel 2007.

Temps de calcul

Les différentes désagrégations ont été simulées sous Excel 2007, pour 500 tirages, afin d'estimer les temps de calcul. Les résultats obtenus sont donnés dans le tableau suivant.

Type de désagrégation	Type de lois	Temps de calcul
Spatiale	Lois normales	50s
	Lois uniformes	20s
Journalière	Lois normales	40s
	Lois uniformes	15s
Horaire	Lois normales	11min
	Lois uniformes	6min
Spatio-journalière	Lois normales	5h30
	Lois uniformes	2h
Spatio-horaire	Lois normales	>3 jours
	Lois uniformes	2 jours

Tableau 2 : Nombre de feuilles nécessaires à l'affichage des résultats

Ces temps de calcul ont été obtenus sur un PC présentant les caractéristiques suivantes :

- Processeur : Intel Core 2 Duo
- Fréquence : 2.66 GHz
- RAM : 2 Go