

Incertitude et variabilité : la nécessité de les intégrer dans les modèles

M. L. Delignette-Muller

Laboratoire de Biométrie et Biologie Evolutive

VetAgro Sup - Université de Lyon - CNRS UMR 5558

24 novembre 2011



Recommandations internationales

La prise en compte de la variabilité et de l'incertitude dans les modèles d'appréciation quantitative des risques alimentaires est recommandée par les instances internationales depuis déjà quelques années.

- 1995 : Application of Risk Analysis to Food Standards Issues, a Joint FAO/WHO Expert Consultation
- 1999 : Principles and Guidelines for the Conduct of Microbial Risk Assessment, Codex alimentarius
- 2003 : Risk Assessment of Food Borne Bacterial Pathogens : Quantitative Methodology Relevant for Human Exposure assessment. European Commission
- ...

La pratique

En pratique subsiste une grande disparité quant aux méthodes utilisées en appréciation quantitative des risques

- Simulations de Monte Carlo à une dimension ne prenant en compte que la variabilité
- Simulations de Monte Carlo à une dimension ne différenciant pas incertitude et variabilité
- Simulation de Monte Carlo à deux dimensions différenciant incertitude et variabilité
- ...

Quel impact sur le résultat ?

Plan de l'exposé

- 1 Variabilité et incertitude
 - Définitions
 - Exemple
 - Besoins du gestionnaire
- 2 Modélisation et simulation
 - Modélisation et inférence
 - Simulation de Monte Carlo à deux dimensions
 - Méthodes alternatives de simulation
- 3 Discussion
 - Distinction variabilité incertitude parfois non triviale
 - Critique courantes des simulations MC2D
 - Conclusion

La variabilité

- résulte de l'**hétérogénéité** des composantes du système à modéliser.
- **ne peut être réduite** sans modifier ce système.
- est liée au caractère aléatoire (stochasticité) des phénomènes étudiés.
- est parfois appelée incertitude de type A (“**aleatory uncertainty**”) ou incertitude objective.
- correspond en langage courant à
“ Je sais que cela ne se passe pas toujours de la même façon ”.

L'incertitude

- résulte du **manque de connaissance**
- **peut être réduite** par l'acquisition de nouvelles connaissances.
- est liée au niveau d'ignorance du modélisateur.
- est parfois appelée incertitude de type B (“**epistemic uncertainty**”) ou incertitude subjective.
- correspond en langage courant à
“ Je ne sais pas exactement comment cela se passe ”.

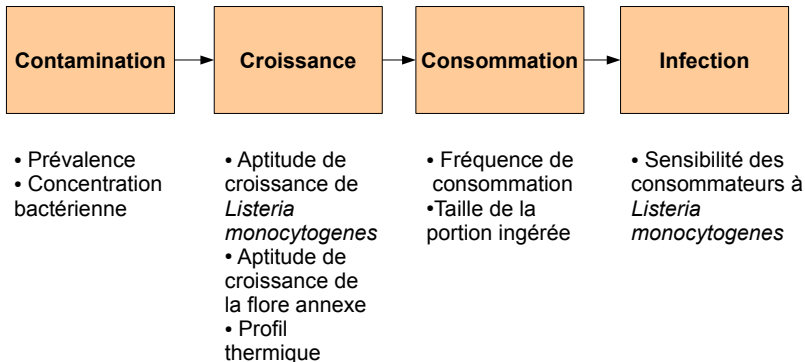
Prenons un exemple

Exemple issu d'un projet précédent (projet LMPR) :

Appréciation quantitative des risques liés à *Listeria monocytogenes* dans le saumon fumé en France.

- Delignette-Muller et al., 2006. Use of bayesian modelling in risk assessment : application to growth of *Listeria monocytogenes* and food flora in cold-smoked salmon.
- Pouillot et al., 2007. Quantitative risk assessment of *Listeria monocytogenes* in french cold-smoked salmon : I. Quantitative exposure assessment.
- Pouillot et al., 2009. Quantitative risk assessment of *Listeria monocytogenes* in french cold-smoked salmon : II. Risk characterization.

Schéma de l'appréciation quantitative des risques



Variabilité prise en compte dans l'exemple

- **concentration** en flore pathogène et en flore annexe dans les produits contaminés en sortie usine
(variabilité inter-produits)
- aptitudes de **croissance** des diverses souches pathogènes et de la flore annexe
(variabilité inter-souches et inter-produits)
- **profil thermique** de la distribution à la consommation
(variabilité inter-produits et inter-consommateurs)
- **taille de la portion** de saumon fumé ingérée
(variabilité inter-consommateurs)
- **sensibilité des consommateurs** au pathogène prise en compte en définissant 4 sous-populations de sensibilité homogène
(variabilité inter-consommateurs)

Incertitude prise en compte dans l'exemple

- **prévalence** : proportion de produits contaminés en sortie usine.
- **concentration** en flore pathogène et en flore annexe dans les produits contaminés en sortie usine
- aptitudes de **croissance** des diverses souches pathogènes et de la flore annexe
- **taille de la portion** de saumon fumé ingérée
- **sensibilité des consommateurs** au pathogène

Il arrive fréquemment qu'on associe à un même paramètre à la fois de la variabilité et de l'incertitude.

Estimation du risque utile au gestionnaire du risque

Une estimation du **risque pour une population donnée**
(par exemple un nombre de cas annuel)
intégrant toutes les sources de variabilité,
assortie de **l'incertitude sur cette estimation.**

Ex. : Estimation du nombre annuel de listérioses liées à la
consommation de saumon fumé,
attendu en France,
et précision associée à cette estimation ?

Quelles méthodes de modélisation et de simulation ?

Modélisation hiérarchique de l'incertitude et de la variabilité

- La **variabilité** est décrite par une **distribution de probabilité** caractérisée par des **paramètres**.
Ex. : température minimale de croissance de *Listeria monocytogenes*

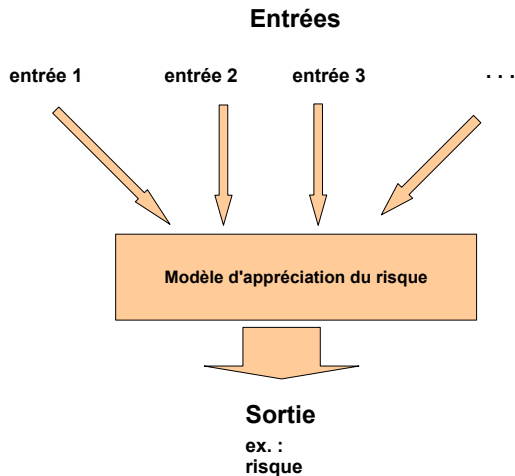
$$T_{min} | M_{Tmin}, S_{Tmin} \sim N(M_{Tmin}, S_{Tmin})$$

- Ces **paramètres** peuvent eux-mêmes être supposés **incertains** et caractérisés par des **distributions de probabilité** obtenues
 - par inférence bayésienne
 - ou à l'aide de techniques de rééchantillonnage (bootstrap)

Différents types d'incertitude

- L'association de la modélisation hiérarchique et de l'inférence bayésienne permet de modéliser facilement **l'incertitude sur les paramètres** liée
 - au manque de données,
 - aux données censurées,
 - aux données incertaines (incertitude de mesure)
- Il peut aussi être important, dans certains cas, de prendre en compte **l'incertitude sur le modèle lui-même**, en utilisant des arbres de décision.

Modèle d'appréciation quantitative des risques

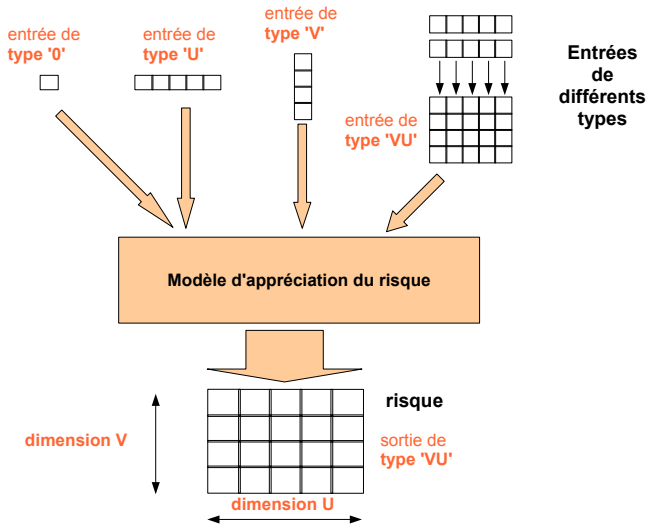


Différents types d'entrées d'un modèle

Formalisme adopté dans le package `Rmc2d` (Pouillot *et al.*)

- Entrée supposée constante et connue (`type="0"`)
- Entrée supposée uniquement incertaine (variabilité négligée) (`type="U"`)
- Entrée supposée uniquement variable (incertitude négligée) (`type="V"`)
- Entrée supposée variable et incertaine (avec modélisation hiérarchique de l'incertitude et de la variabilité telle que décrite précédemment) (`type="VU"`)

Transferts de l'incertitude et de la variabilité



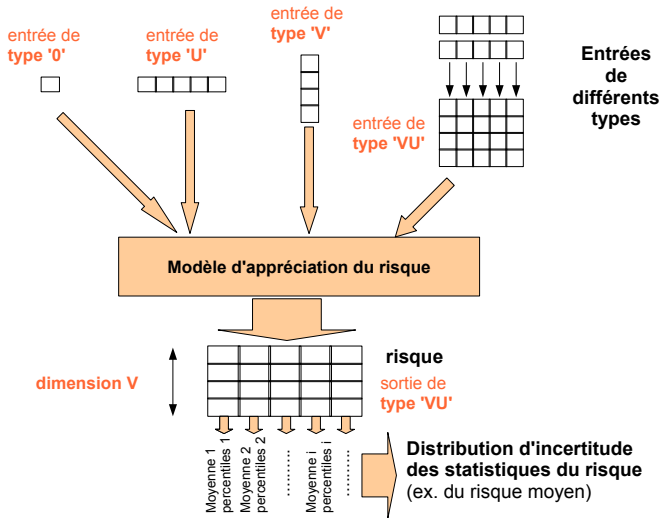
Monte Carlo à deux dimensions

pour i allant de 1 à N_U (incertitude)

- tirage aléatoire des entrées incertaines dans leurs distributions respectives
- **Monte Carlo à une dimension pour j allant de 1 à N_V (variabilité)**
 - **tirages aléatoires des entrées variables dans leurs distributions respectives**
 - **calcul de la sortie du modèle**
- calcul et stockage d'un certain nombre de statistiques décrivant la distribution de variabilité de la sortie du modèle (ex. risque moyen)

Résultats : percentiles des statistiques décrivant la distribution de variabilité de la sortie.

Résultats des simulations



Résultats obtenus par Monte Carlo à deux dimensions sur l'exemple

Exemple

Résultat sur l'incidence annuelle chez les individus de référence (sous-population la moins sensible)
exprimée en nombre de cas prédits sur cette sous-population :
estimation ponctuelle (médiane) : environ 50 cas par an
intervalle de crédibilité à 95% : entre quelques cas et plus de 1000 cas

Calcul déterministe (cas le plus probable)

Que se passe-t-il si on fixe tous les paramètres aux valeurs les plus probables ?

Résultat sur le même exemple :

estimation ponctuelle à 0.01 cas par an.

Ce genre de calcul sous-estime très souvent le risque pour la population.

Prise en compte de la seule variabilité : MC1D

Que se passe-t-il si on néglige l'incertitude
(cas le plus probable uniquement sur la dimension
d'incertitude) ?

Résultat sur le même exemple :

estimation ponctuelle à environ 30 cas par an.

Le résultat peut fournir une bonne estimation du risque si
l'incertitude est peu importante par rapport à la variabilité
(pas facile à anticiper).

On ne peut pas associer d'incertitude à cette estimation.

Prise en compte de la variabilité et de l'incertitude dans une même dimension : MC1D

Que se passe-t-il si on ne sépare pas incertitude et variabilité dans les simulations ?

Résultat sur le même exemple :

médiane à 0 cas par an

intervalle de crédibilité à 95% : entre 0 et 200 cas

Que simule-t-on exactement ainsi ?

- un risque individuel : **intérêt pour le gestionnaire ?**
- un risque pour une population homogène sur tous les maillons de la chaîne depuis la contamination du produit jusqu'à l'infection : **réalisme ?**

Résultats difficiles à interpréter et souvent biaisés par excès de poids donné aux scénarios extrêmes (ici risque nul)

La méthode de simulation de Monte Carlo à deux dimensions semble donc la plus appropriée, mais sa mise en oeuvre révèle parfois quelques difficultés.

Quand la variabilité devient de l'incertitude

La distinction entre la variabilité et l'incertitude n'est pas toujours triviale et dépend de la question posée par le gestionnaire.

Parfois une part de la variabilité doit être considérée comme de l'incertitude.

Exemple de question :

Nombre de cas de listérioses prédit lors d'une épidémie liée à une seule production de saumon fumé, contaminée à un niveau donné, par une souche unique (inconnue) de *Listeria monocytogenes* ?

Pour répondre à une telle question, on devrait considérer la variabilité inter-souches et au moins une part de la variabilité inter-produits comme de l'incertitude.

Quand les données rendent difficiles les distinctions

Le nombre et la qualité des données rendent parfois difficiles les distinctions :

- entre incertitude et variabilité, notamment si l'incertitude est trop grande.
- entre les diverses sources de variabilité.
Il devient alors impossible de considérer une partie de la variabilité comme incertitude pour répondre à une question de type “nombre de cas lors d'une épidémie”.

Le Monte Carlo à deux dimensions critiqué

Les méthodes de Monte Carlo à deux dimensions sont parfois critiquées pour :

- la **lourdeur des calculs** : $10000 \times 10000 = 100000000$!
- des **sorties difficiles à interpréter**
- **incertitude décrite par une distribution de probabilité**
(probabilité au sens subjectif : degré de croyance, qui est néanmoins à la base de l'inférence bayésienne)

Certains ont proposé une méthode hiérarchique alternative, modélisant la variabilité sous forme de distributions de probabilité et l'incertitude sous forme d'intervalles ("P-Box approach", Karanki *et al.*, 2009) : méthode non encore explorée en risque alimentaire.

Conclusion

Même si la méthode utilisée le plus souvent dans ce but présente quelques inconvénients (modélisation hiérarchique suivie de simulations de Monte Carlo à deux dimensions), la prise en compte réfléchie et prudente de la variabilité et de l'incertitude de façon séparée semble nécessaire pour rendre un résultat non biaisé et pertinent au gestionnaire des risques.