

## Analyse de sensibilité de modèles à sorties stochastiques et fonctionnelles

H. M. KOUYE  
*Université Paris-Saclay*

**Supervisor(s):** Dr. Elisabeta Vergu (INRAE), Dr. Gildas Mazo (INRAE), Prof. Clémentine Prieur (Université Grenoble-Alpes, Inria)

**Ph.D. expected duration:** Oct. 2019 - Sep. 2022

**Address:** UR1404 INRAE MaIAGE (Mathématiques et Informatique Appliquées du Génome à l'Environnement), 78352 Jouy-en-Josas.

**Email:** henri-mermoz.kouye@inrae.fr

**Abstract:** Dans le contexte des modèles de complexité croissante, on note l'émergence des modèles stochastiques, permettant de décrire des phénomènes aussi bien à petite échelle qu'à grande échelle grâce à des outils issus des probabilités. Pour cette classe de modèles, deux appels du modèle évalué à une même entrée peuvent donner lieu à deux sorties différentes (voir exemple de modèle de dynamique épidémique figure 1). Lorsque l'on tire une entrée au hasard et que l'on évalue la sortie du modèle, il y a donc deux sources d'aléa : l'une due au tirage des entrées et l'autre due au modèle lui-même. Il est commode, alors, de supposer l'existence d'une fonction  $f$  et d'un élément aléatoire  $Z$  tels que

$$Y = f(X_1, \dots, X_p, Z),$$

où  $X_1, \dots, X_p$  sont les entrées et l'élément aléatoire  $Z$ , qui n'est pas observable, représente le bruit du modèle.

Les modèles épidémiologiques de type SIR (partitionnant la population en Sains, Infectés et Retirés) sont à la base de nombreuses modélisations de dynamiques épidémiques. L'analyse de sensibilité de type Sobol' d'un modèle SIR stochastique doit être capable de séparer la variabilité due à l'aléa intrinsèque au modèle de la variabilité due à la méconnaissance des paramètres incertains  $X_1, \dots, X_p$ . Par ailleurs, elle doit prendre en compte le caractère fonctionnel des sorties d'intérêt. Jusqu'à récemment, en épidémiologie, les méthodes d'analyse de sensibilité utilisées étaient relativement empiriques [2], sans garanties théoriques. Quelques articles récents ont proposé un cadre formel de décomposition de la variance dans le cadre de modèles stochastiques (voir par exemple l'introduction de Etoré *et al* [3] pour une revue bibliographique récente). Récemment, Mazo [6] a proposé une définition formelle des indices de Sobol', des estimateurs associés dans ce cadre et un plan d'échantillonnage optimisé avec des garanties statistiques.

Dans ce travail, nous proposons dans un premier temps de séparer les sources d'aléa intrinsèque au modèle SIR des sources d'aléa dues à une méconnaissance des paramètres de taux d'infection et de taux de guérison. La construction usuelle du SIR-stochastique repose sur une vision markovienne. La dynamique de l'épidémie est modélisée par une chaîne de Markov à temps continu caractérisée par deux composantes : la chaîne de Markov incluse (discrète) qui génère les transitions d'états et un processus ponctuel qui définit les instants où surviennent ces transitions. Cependant, l'interaction entre ces deux composantes rend difficile la caractérisation de l'aléa intrinsèque  $Z$ . Pour cette raison, nous avons préféré adopter la construction de Sellke [1],[7]. La construction de Sellke décrit la dynamique de l'épidémie en faisant intervenir deux variables aléatoires : la résistance de chaque individu de la population étudiée face à la pression de l'infection et la durée que chacun mettra à se rétablir indépendamment les uns des autres. Un avantage de cette construction est qu'elle repose sur les individus et sur l'indépendance de leurs caractéristiques face

à l'épidémie. La construction de Sellke est équivalente à la construction markovienne modulo un choix de lois exponentielles décrivant le temps passé dans chaque état. De plus elle permet de caractériser l'aléa intrinsèque  $Z$ . Dans un second temps, nous pouvons appliquer le formalisme proposé par Mazo [6] pour définir les indices de Sobol' temps par temps des sorties fonctionnelles que sont les nombres d'individus infectés, retirés et sains en fonction du temps. Nous proposons également un calcul des indices agrégés introduits par Lamboni *et al* [5] (voir aussi [4]).

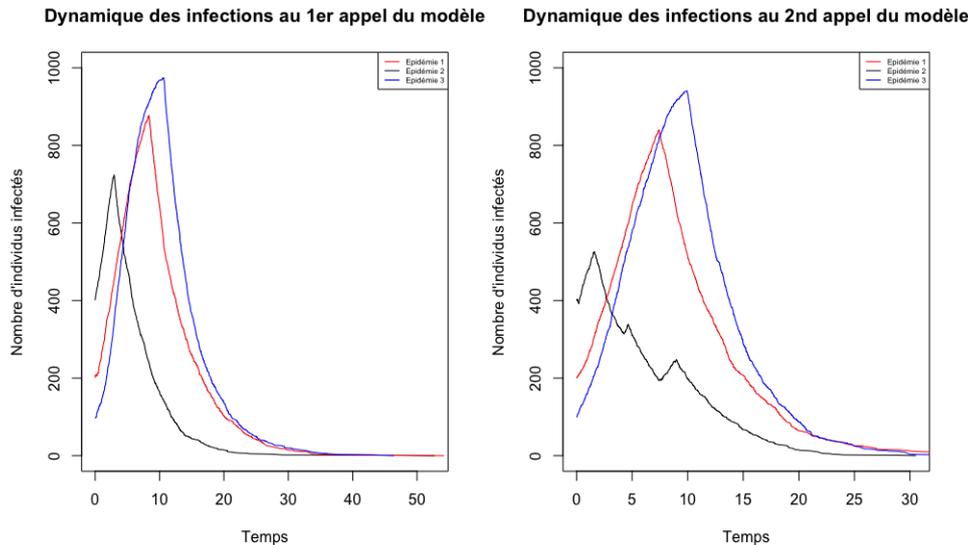


Figure 1: Courbes du nombre d'infectés du modèle SIR-stochastique pour trois scénarios d'épidémies, c'est-à-dire pour trois jeux de paramètres d'entrées différents (taux d'infection et taux de guérison). Les différences entre les courbes des deux appels (à comparer les mêmes couleurs) illustrent l'effet de l'aléa intrinsèque.

## References

- [1] Hakan Andersson and Tom Britton. *Stochastic epidemic models and their statistical analysis*. Springer lectures notes in statistics, 2000.
- [2] Aurélie Courcoul, Hervé Monod, Mirjam Nielen, Don Klinkenberg, Lenny Hogerwerf, François Beaudeau, and Elisabeta Vergu. Modelling the effect of heterogeneity of shedding on the within herd *coxiella burnetii* spread and identification of key parameters by sensitivity analysis. *Journal of Theoretical Biology*, 284(1):130 – 141, 2011.
- [3] Pierre Etoré, Clémentine Prieur, Dang Khoi Pham, and Long Li. Global sensitivity analysis for models described by stochastic differential equations. *Methodology and Computing in Applied Probability*, pages 1–29, 2019.
- [4] Fabrice Gamboa, Alexandre Janon, Thierry Klein, and Agnès Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *Electron. J. Statist.*, 8(1):575–603, 2014.
- [5] Matieyendou M. Lamboni, Herve H. Monod, and David D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering and System Safety*, 96(4):450–459, 2011.
- [6] Gildas Mazo. An optimal tradeoff between explorations and repetitions in global sensitivity analysis for stochastic computer models. July 2019.
- [7] Thomas Sellke. On the asymptotic distribution of the size of a stochastic epidemic. *Journal of Applied Probability*, 20(2):390–394, 1983.

**Short biography** – Après avoir suivi un cursus de mathématiques fondamentales et appliquées, j'ai obtenu un master de mathématiques appliquées aux sciences du vivant à l'université Paris-Saclay grâce à la bourse d'excellence de la fondation mathématique Jacques Hadamard. Ma thèse, financée par l'INRAE et l'Inria, est destinée à répondre aux besoins d'outils d'analyse de sensibilité pour des modèles stochastiques complexes.