

Projet de Modélisation en Épidémiologie

Pierre-Alexandre Quittet
Bastien Clémot

MODE 2023-2024

Lien GitHub du projet : <https://github.com/bClemot-Sc/MEPI-Projet>

Ce projet a été réalisé dans le cadre de la validation de l'unité d'enseignement MEPI (modélisation en épidémiologie) du Master 2 de modélisation en écologie de l'Université de Rennes 1. Ce travail est une introduction à l'analyse de sensibilité d'un modèle d'épidémiologie en temps discret **ModAppli()** qui sera détaillé dans la suite du rapport. Trois méthodes à complexité croissante ont été présentées : la méthode OAT standard, la méthode de Morris ((Morris, 1991)) et la méthode FAST étendue ((Cukier et al., 1973)). Après une description générale du modèle, les trois méthodes auront une partie consacrée à chacune. Elles seront introduites par une brève description du principe général et de la méthode, suivie d'une mise en pratique sur le logiciel R (R Core Team et al., 2022) et d'une interprétation des résultats. Les analyses de sensibilité ont été conduites sur le modèle *original* ainsi que sur un deuxième modèle issu du premier pour lequel nous avons intégré une nouvelle fonction de transmission.

Afin d'illustrer cette analyse de modèle épidémiologique, nous utiliserons l'exemple concret d'un agent pathogène et de son hôte associé. Nous prendrons le cas des Ranavirus et de l'espèce chez laquelle ils ont été historiquement décrits pour la première fois, la grenouille léopard *Lithobates pipiens* (Granoff et al., 1966). Nous considérerons ainsi la dynamique démographique des trois stades de développement suivant : **Larvaire**, pour les têtards, **Juvénile**, pour les jeunes individus matures, et enfin **Adulte**.

Les Ranavirus sont des virus à ARN qui contaminent notamment les animaux aquatiques et provoquent l'apparition de nécroses. Dans la littérature, ces virus présentent une forte infectiosité et virulence. Cependant, nous considérerons ici un mutant hypothétique ayant une plus faible virulence. Au sein de la population, ces virus vont pouvoir se transmettre par contact direct avec des individus infectés (Gray et al., 2009). Nous définirons donc au sein de la population quatre états de santé :

- **Sensible** : Individus sans pathologie, pouvant être infectés.
- **Exposé** : Individus infectés par le virus, en période d'incubation.
- **Infecté** : Individus infectés par le virus et exprimant la pathologie.
- **Remis** : Individus ayant survécu à la pathologie et en ayant acquis temporairement l'immunité.

1. Description du modèle

Par souci de clarté, les paramètres du modèle ont été renommés. Les nouveaux noms, en comparaison avec ceux du script original sont disponible dans le tableau 2.

a. Définissez le type de ce modèle

Le modèle présenté ici est de type **déterministe** car il comprend des paramètres fixes et prédéterminés n'étant pas soumis à des variations aléatoires en fonction du temps. C'est aussi un modèle à **compartiments** puisque les effectifs de la population sont sub-divisés en classes d'âges, entrecroisées d'états de santé. Enfin, c'est un modèle à **temps discret**, car à chaque pas de temps les effectifs $U_n + 1$ des compartiments sont calculés à partir des effectifs au temps précédent U_n .

b. Quels sont les processus biologiques en jeu ?

Processus démographiques

Le modèle consiste en une population structurée en 3 classes d'âges : Larvaire (L), Jeune (J) et Adulte (A). La croissance de cette population est logistique ($1 - \frac{N}{K}$) et seule les classes d'âges J et A peuvent se reproduire (taux de fécondité f_2 et f_3). Le nombre de descendants par reproduction est caractérisé par le paramètre Θ et seules les femelles sont considérées (sex ratio : $\psi = 0.5$). Le passage d'une classe d'âge i à la suivante se fait selon un taux de maturation τ_i . Chaque classe d'âge est soumise à un processus de mort naturelle modélisé par le taux m_i .

Processus épidémiologiques

Ce modèle décrit également différents états : Sensible (S), Infecté (E), Infectieux (I) et Immunisé (R). Les individus sensibles ont une chance d'être infectés par la rencontre avec des individus infectieux (β) et passeront à l'état infecté. Une fois infectés, les individus disposent d'un temps d'incubation (σ) avant de passer à l'état infectieux. L'état infectieux implique une mortalité supplémentaire liée à la maladie (μ) et une probabilité de devenir immunisé (γ). Dans l'état immunisé, les individus peuvent perdre leur immunité (λ) et ainsi redevenir sensibles.

c. Ecrivez les équations associées

Soit la population structurée autour des 3 classes d'âges : Larvaire (L), Juvénile (J) et Adulte (A), pouvant exprimer les états Sensible (S), Exposé / Infecté (E), Infectieux (I) et Immunisé / Résistants (R).

$$\left\{ \begin{array}{l}
L(t+1) = \left\{ \begin{array}{l}
L_S(t+1) = L_S(t) + max \left[0; \psi \Theta \left(f_J J(t) + f_A A(t) \right) \cdot \left(1 - \frac{N}{K} \right) \right] + \lambda L_R(t) - L_S(t) \cdot \left(m_L + \tau_L + \beta \frac{I(t)}{N} \right) \\
L_E(t+1) = L_E(t) + \beta L_S(t) \frac{I(t)}{N} - L_E(t) \cdot (m_L + \tau_L + \sigma) \\
L_I(t+1) = L_I(t) + \sigma L_E(t) - L_I(t) \cdot (m_L + \tau_L + \mu + \gamma) \\
L_R(t+1) = L_R(t) + \gamma L_I(t) - L_R(t) \cdot (m_L + \tau_L + \lambda)
\end{array} \right. \\
\\
J(t+1) = \left\{ \begin{array}{l}
J_S(t+1) = J_S(t) + \tau_L L_S(t) + \lambda J_R(t) - J_S(t) \cdot \left(m_J + \tau_J + \beta \frac{I(t)}{N} \right) \\
J_E(t+1) = J_E(t) + \tau_L L_E(t) + \beta J_S(t) \frac{I(t)}{N} - J_E(t) \cdot (m_J + \tau_J + \sigma) \\
J_I(t+1) = J_I(t) + \tau_L L_I(t) + \sigma J_E(t) - J_I(t) \cdot (m_J + \tau_J + \mu + \gamma) \\
J_R(t+1) = J_R(t) + \tau_L L_R(t) + \gamma J_I(t) - J_R(t) \cdot (m_J + \tau_J + \lambda)
\end{array} \right. \\
\\
A(t+1) = \left\{ \begin{array}{l}
A_S(t+1) = A_S(t) + \tau_J J_S(t) + \lambda A_R(t) - A_S(t) \cdot \left(m_A + \beta \frac{I(t)}{N} \right) \\
A_E(t+1) = A_E(t) + \tau_J J_E(t) + \beta A_S(t) \frac{I(t)}{N} - A_E(t) \cdot (m_A + \sigma) \\
A_I(t+1) = A_I(t) + \tau_J J_I(t) + \sigma A_E(t) - A_I(t) \cdot (m_A + \mu + \gamma) \\
A_R(t+1) = A_R(t) + \tau_J J_R(t) + \gamma A_I(t) - A_R(t) \cdot (m_A + \lambda)
\end{array} \right.
\end{array} \right.$$

Avec :

$$L(t) = L_S(t) + L_E(t) + L_I(t) + L_R(t)$$

$$J(t) = J_S(t) + J_E(t) + J_I(t) + J_R(t)$$

$$A(t) = A_S(t) + A_E(t) + A_I(t) + A_R(t)$$

$$N(t) = L(t) + J(t) + A(t)$$

$$I(t) = L_I(t) + J_I(t) + A_I(t)$$

d. Fournissez le schéma des transitions entre états concordant avec ces équations

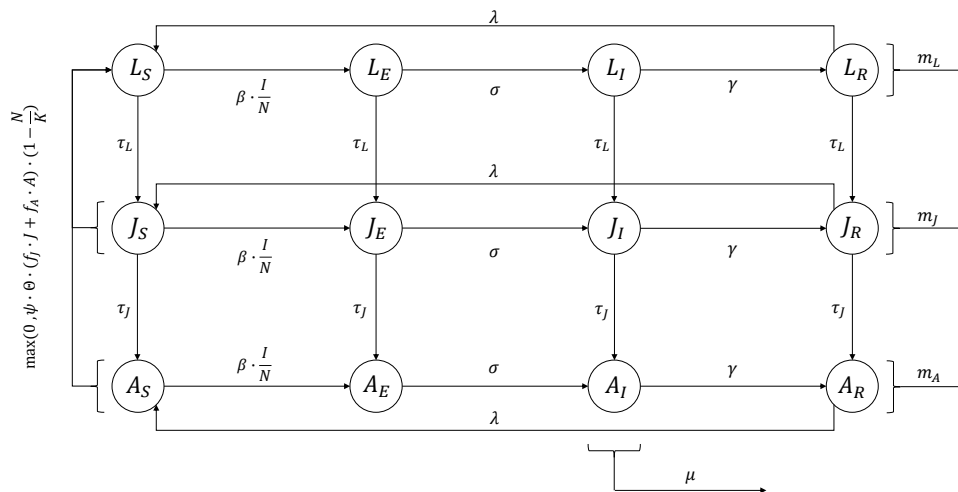


Figure 1 – Schéma de transition du modèle initial.

e. Listez les hypothèses principales de ce modèle (quel type de force d'infection, population variable ou constante, etc)

- Infection différée par état exposé non infectieux
- Infection fréquence dépendance
- Force d'infection forte et virulence faible
- L'immunité peut se perdre
- Population structurée en âge
- Spatial homogène (même environnement pour tout le monde)

La population, structurée en classe d'âge, est variable avec un flux entrant (naissance) et un flux sortant (mort naturelle et mort liée à la maladie). Pour éviter un taux de naissance négatif, le maximum entre 0 et le produit du sexe-ratio, du nombre de descendant par portée et du taux de fécondité avec les effectifs des classes reproductrices est sélectionné. Le modèle fait également l'hypothèse que les mortalités naturelles de la classe d'âge 1 ($m_1 = 0.14\%$) et 3 ($m_2 = 0.19\%$) sont plus élevées que celle de la classe 2 ($m_2 = 0.029\%$).

La force d'infection de la maladie est relativement élevée avec 30% ($\beta = 0.3$). La mortalité liée à la maladie est en revanche assez faible, avec un taux de mortalité de 0.1%. Le modèle décrit donc une maladie fortement contagieuse et peu létale. De plus, les individus passent d'un état intermédiaire exposé non-infectieux avant de devenir infectieux avec une période de latence et ainsi transmettre l'infection. Les individus immunisés peuvent perdre cette immunité et revenir à l'état sensible.

f. Quelles sont les conditions initiales ?

Les conditions initiales définissent la répartition de la population au sein des différents âges et états de santé, au temps 0 du modèle. On se base donc ici sur une population initiale de 87 individus comprenant 27 individus larvaires sensibles L_S , 23 individus juvéniles sensibles J_S , 36 individus adultes sensibles A_S et 1 individu adulte infectieux A_I , qui va servir de point de départ à l'infection de la population.

g. Listez et définissez les sorties possibles du modèle

En sachant que la simulation du modèle se déroule en jours sur deux ans, soit 730 pas de temps. La fonction du modèle propose alors 4 sorties :

- **Sortie 1** : $\frac{E(t=730)+I(t=730)}{N(t=730)}$ correspondant au taux de morbidité de la maladie au dernier jour
- **Sortie 2** : $\frac{\beta S(t=730) \cdot I(t=730)}{N}$ correspondant à l'incidence de la maladie au dernier jour
- **Sortie 3** : $\max_{i=1}^{730} I(t=i)$ correspondant au pic infectieux de la maladie
- **Sortie 4** : $\sum_{t=1}^{365} I(t=i)$ correspondant à la la prévalence de la maladie lors de la première année

h. Listez les paramètres du modèle, donnez leurs définitions et leurs valeurs

Paramètre	Code R	Valeur	Définition
K	K	100	Capacité d'accueil du milieu
ψ	sr	0.5	Sex-ratio
m_L	m_1	0.0014	Mortalité naturelle de la classe d'âge 1
m_J	m_2	0.00029	Mortalité naturelle de la classe d'âge 2
m_A	m_3	0.0019	Mortalité naturelle de la classe d'âge 3
f_J	f_2	0.0019	Taux de fécondité de la classe d'âge 2
f_A	f_3	0.0082	Taux de fécondité de la classe d'âge 2
Θ	portee	5	Nombre de descendants par reproduction
τ_L	t_1	1/365 jours ⁻¹	Taux de maturation de la classe d'âge 1
τ_J	t_2	1/365 jours ⁻¹	Taux de maturation de la classe d'âge 2
β	trans	0.3	Taux de transmission de la maladie
σ	lat	1/5 jours ⁻¹	Taux d'incubation
γ	rec	1/20 jours ⁻¹	Taux de récupération
λ	loss	1/100 jours ⁻¹	Taux de perte d'immunité
μ	madd	0.001	Mortalité liée à la maladie

Table 1 – Paramètres de la modélisation

i. Réalisez quelques figures (avec titres explicites et légendes) pour illustrer au mieux le comportement du modèle pour les valeurs nominales des paramètres (valeurs fournies)

Voir figure 2.

j. Interprétez les figures (de manière concise)

Le modèle atteint un état de stabilité vers les 200 premiers jours (**Sortie 1** : Taux de morbidité final = 0.183). À cet état stable, la majorité des individus se trouvent dans l'état résistant (figure 2.a). Cette prédominance s'explique par le fait que le temps de perte d'immunité est plus long que les temps d'incubation et de récupération. Par conséquent, les individus passent nettement plus de temps dans l'état résistant que dans les autres états.

Ce raisonnement s'applique également aux compartiments des exposés et des infectieux, où le temps d'incubation est bien plus court que le temps de récupération, entraînant ainsi une majorité d'individus dans le compartiment infectieux. Le nombre d'individus sensibles est quant à lui compensé par la natalité.

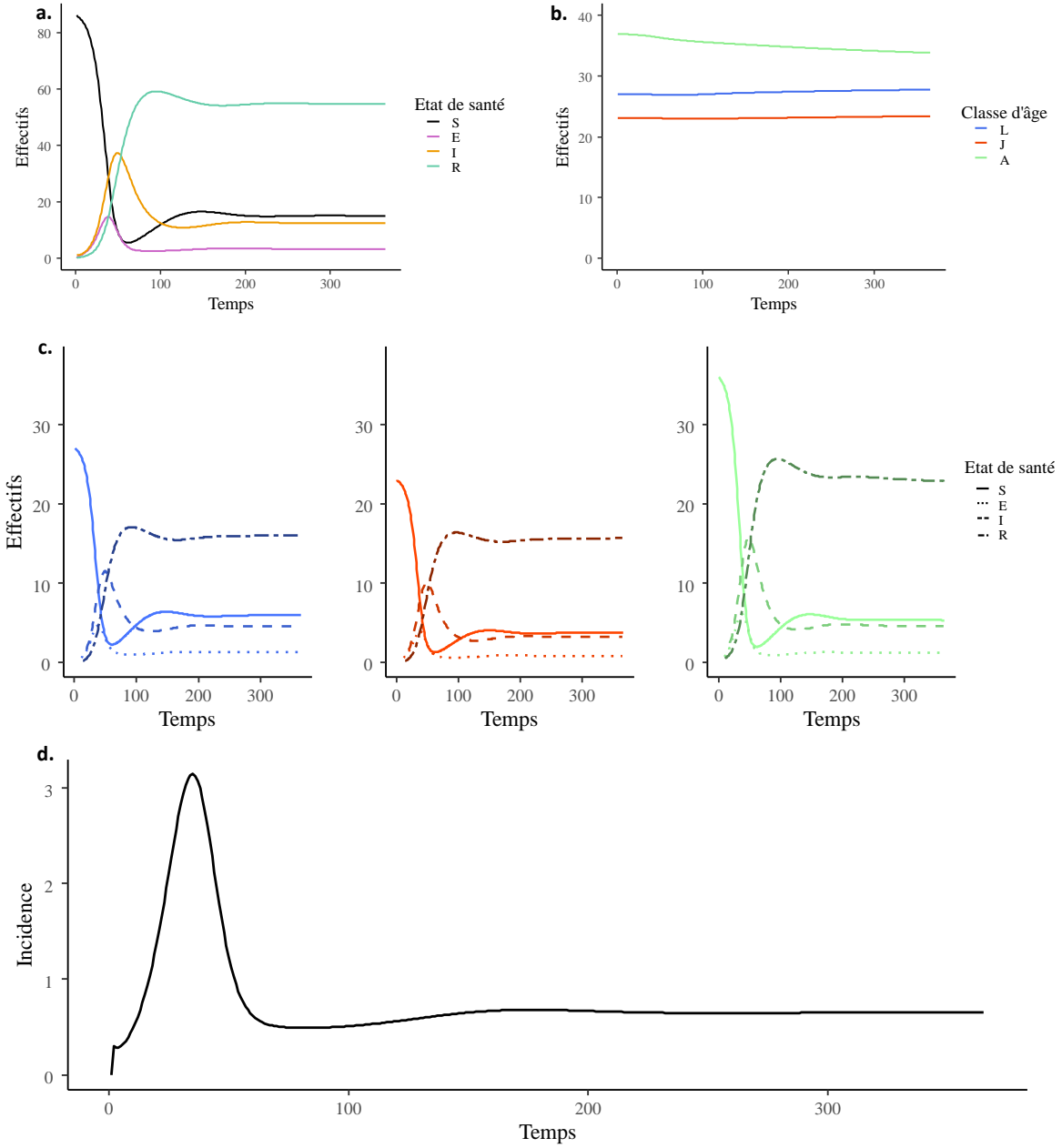


Figure 2 – Comportement des effectifs du modèle. Du fait de la stabilisation rapide du modèle, seule la première année de simulation a été représentée. **a.** Évolution des effectifs totaux pour chaque état de santé. **b.** Évolution des effectifs totaux pour chaque classe d'âge. **c.** Évolution des effectifs totaux pour chaque état de santé, en fonction des classes d'âge (Larvaire, Juvenile et Adulte). **d.** Évolution de l'incidence.

Pour les trois classes d'âge, un comportement similaire est observé (figure 2.c). Il est également notable que la proportion d'individus adultes est plus élevée que celle des larvaires et des juvéniles (figure 2.b et .c). Cette observation peut être expliquée par le fait que les individus quittent rapidement les stades larvaires et juvéniles en raison des taux de maturation, restant ensuite dans l'état adulte jusqu'à leur mort.

Finalement, le modèle révèle un pic d'incidence maximal aux alentours de 35 jours (figure 2.d), accompagné de manière cohérente par un pic d'infectieux (**Sortie 3** : Pic infectieux = 37.205 individus) et une diminution du nombre d'individus sensibles (figure 2.a). Ce pic d'incidence diminue rapidement, accompagné d'une augmentation du nombre d'individus immunisés. La population devient donc majoritairement infectée en l'espace d'une trentaine de jours. Ensuite, en raison de la courte période de récupération et de la faible mortalité, les individus acquièrent rapidement une immunité, limitant ainsi et stabilisant l'incidence (**Sortie 2** : Incidence finale = 0.645 ; **Sortie 4** : Prévalence de la première année = 280.944).

2. Analyse de sensibilité : OAT

Une analyse de sensibilité d'un modèle revient à mettre en relation les informations entrantes et sortantes du modèle. Elle est définie comme "l'étude de la manière dont l'incertitude de la sortie d'un modèle (numérique ou autre) peut être attribuée aux différentes sources d'incertitude dans les données d'entrée du modèle" (Saltelli et al., 2019). Elle est différente de l'analyse d'incertitude qui quantifie la variabilité des sorties mais ne les relie pas à la variabilités des paramètres d'entrées. En d'autres termes, l'analyse d'incertitude répond à la question *Quel est le niveau d'incertitude de $\mathcal{F}(x)$ induite par l'incertitude de x ?* alors que l'analyse de sensibilité répond à la question *Quelles sont les principales sources d'incertitude parmi les différents paramètres x_1, \dots, x_i ?*. Bien que ces deux questions soient complémentaires, la distinction conceptuelle est importante.

Une analyse de sensibilité est généralement organisé autour de 4 étapes majeures (Faivre et al., 2013) :

- Définition des gammes de variations des paramètres incertains
- Génération du plan d'échantillonnage des valeurs de paramètres (e.g. tirage aléatoire, simulation numérique...)
- Calcul des sorties du modèle avec le plan d'échantillonnage
- Calcul d'indice de sensibilité des paramètres (e.g. indices de Sobol, indice de Baueur et Hamby...)

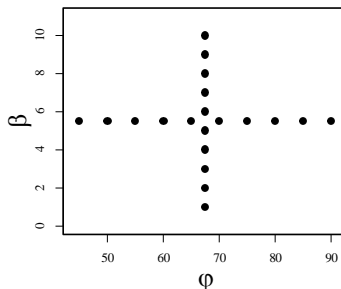


Figure 3 – Exploration de l'espace des paramètres avec la méthode OAT standard

La méthode OAT (one-at-a-time) fait partie des analyse de sensibilité locale. L'objectif d'une analyse de sensibilité locale est d'étudier la variabilité des sorties du modèle induite par une petite variation d'un unique paramètre x_i autour d'une valeur de référence x_0 . On va faire varier les paramètres un par un. On étudie ensuite comment ces variations se répercutent-elles sur la sortie du modèle.

Cette méthode est simple et peu coûteuse en simulations. Elle est adaptée à la détection de relation linéaire entre les paramètres et les sorties du modèle. Elle nécessite de définir des gammes de valeurs plausibles pour chaque paramètres et ne permet pas la détection d'interaction entre les paramètres.

L'analyse OAT standard ici présentée mesure la sensibilité des variations de la sortie du modèle en ne faisant varier qu'une seule variable d'entrée avec les autres valeurs étant fixées à leurs valeurs initiales (figure 3) permettant ainsi d'étudier comment de petites perturbations autour d'une valeur de référence des entrées influent la sortie du modèle.

Dans cette analyse, nous avons testé 10 valeurs par paramètres. Afin d'évaluer la contribution de chaque paramètre à la variation des 4 sorties du modèle, nous avons utilisé l'indice de sensibilité I_{x_i} de Bauer & Hamby (Bauer and Hamby, 1991) :

$$I_{x_i} = \frac{\max_{x_i} [\mathcal{F}(x_1, \dots, x_i, \dots, x_K)] - \min_{x_i} [\mathcal{F}(x_1, \dots, x_i, \dots, x_K)]}{\max_{x_i} [\mathcal{F}(x_1, \dots, x_i, \dots, x_K)]}$$

Cette indice était calculé pour chaque paramètre pour chaque sortie du modèle, avec x_i la variante i du paramètre x (i.e. une des 10 valeurs prises par x sur l'ensemble de sa gamme de variation $[\min; \max]$), K le nombre de paramètre total du modèle et $\mathcal{F}(x_1, \dots, x_K)$ la valeur de sortie de la fonction. L'indice de Bauer & Hamby met en rapport le minimum et maximum de variation obtenu par les variants i de chaque paramètre x_k , tous les autres paramètres étant fixé à leur valeur nominale. C'est un indice simple borné entre 0 et 1 qui permet de quantifier la contribution indépendante de chaque paramètre à la variation de sortie du modèle.

a. Justifiez les valeurs testées

Pendant la génération des gammes de variations pour chaque paramètres; nous avons fait en sorte que les valeurs soient biologiquement valides. Des considérations supplémentaires ont été appliquées pour certains paramètres : le sex-ratio (ψ ou sr) ainsi que les différents taux (transmission, immunité, fécondité...) ne pouvaient pas être supérieurs à 1. Pour tous les paramètres, nous avons fait en sorte que les variations soient symétriques (i.e. aucun a priori sur la direction de l'incertitude). C'est pourquoi nous avons fait varier l'ensemble chacun des paramètres initiaux p_0 de façon symétrique à plus ou moins 25% autour de chaque valeur initiale. Ainsi nous avons considéré la gamme de variation suivante : $\text{gamme}_{\hat{p}} = p_0 \pm 0.25 \cdot p_0$. Cette ordre de grandeur est classique dans les analyses de sensibilités (e.g. Charron et al., 2011 : $\pm 10\%$; Beaunée et al., 2015 : $\pm 25\%$ et Lupo et al., 2019 : $\pm 25\%$).

b. Fournissez votre script

Script 1 : Script de l'analyse OAT standard

```
# Packages
library(sensitivity)
library(ggplot2)
library(tidyverse)
library(latex2exp) # Texte latex

# IMPORTATION FONCTION DE BASE ET DES VALEURS INITIALES
source("FONCTION_BASE.R")

# INITIALISATION -----
# Gamme de variation de chaque parametre
gamme_params <- vector(mode = "list", length = 15)
for (i in 1:length(ValNominale)) {
  gamme_params[[i]] <- seq(from = 0.75 * ValNominale[i],
                           to = 1.25 * ValNominale[i],
                           length.out = 10)
}

# Variable contenant les noms des parametres
names_para <- names(ValNominale)

# Matrice avec en ligne les scenarii non-modifis et en colonne les parametres
mat_scenarios <-
```



```

matrix(
  data = ValNominale,
  nrow = 15 * 10,
  ncol = 15,
  dimnames = list(paste0("scenario", 1:(15 * 10)), names_para),
  byrow = T
)

# Implementation des scenarii modifis
ligne <- 1

for (i in 1:length(gamme_params)){
  mat_scenarios[ligne:(ligne+10-1), i] <- gamme_params[[i]][1:10]
  ligne <- ligne + 10
}

# MODELISATION -----
# On modlise chaque scenario avec la fonction modAppli
oat <- modAppli(mat_scenarios)
dimnames(oat) <- list(c(1:150), c("tx_morbidite", "incidence_t730", "pic_infectieux", "prevalence_annee_1"))

# On modlise avec les paramtres de rfrence le modle m0
m0 <- modAppli(matrix(ValNominale, nrow = 1))

# OAT STANDARD : CALCUL SENSIBILITE -----
# --- Fonction apply pour calculer l'indice de sensibilit par sortie du modle par paramtre par scenario
list_as <- apply(
  X = oat,
  MARGIN = 1,

  # Indice de sensibilite = (output_initial - output_modifie_i) / (output_initial + output_modifie_i)
  FUN = function(x){abs(m0 - x)/(x + m0)}
)

list_as <- t(list_as) # on transpose le dataframe

# --- Inititalisation des matrices
as_mean <-
  as.data.frame(matrix(
    data = 0,
    nrow = 15,
    ncol = 4,
    dimnames = list(names_para, c("tx_morbidite", "incidence_t730", "pic_infectieux", "prevalence_annee_1"))
  ))

as_se <-
  as.data.frame(matrix(
    data = 0,
    nrow = 15,
    ncol = 4,
    dimnames = list(names_para, c("tx_morbidite", "incidence_t730", "pic_infectieux", "prevalence_annee_1"))
  ));

ligne <- 1 # incrmentation ligne pour prendre les paramtres des 10 valeurs de sd pour chaque paramtre

# --- Remplissage du df
for (i in 1:15){
  as_mean[i, 1:4] <- apply(list_as[ligne:(i*10),1:4], MARGIN = 2, FUN = function(x) mean(x, na.rm = T))
  as_se[i, 1:4] <- apply(list_as[ligne:(i*10),1:4], MARGIN = 2, FUN = function(x) sd(x, na.rm = T))

  ligne <- i*10 + 1
} # fin boucle 'parametre'

```

c. Réalisez des figures illustrant le comportement du modèle

La distribution des valeurs des 4 sorties du modèle avec les 150 scénarios des 15 paramètres variant un à un sur une gamme de 10 valeur montre une distribution très réduite quelque soit la sortie considérée (figure 4). En effet, on observe une grande fréquence de valeurs très proches des valeurs initiales des différents indicateurs. Compte tenu du fait que la méthode d'échantillonnage de la méthode OAT standard ne permet de tester que

10 valeurs par paramètres en maintenant les autres fixés à leur valeur nominale, la variation potentielle en terme de sortie du modèle est donc assez minime, d'où la distribution presque unimodale.

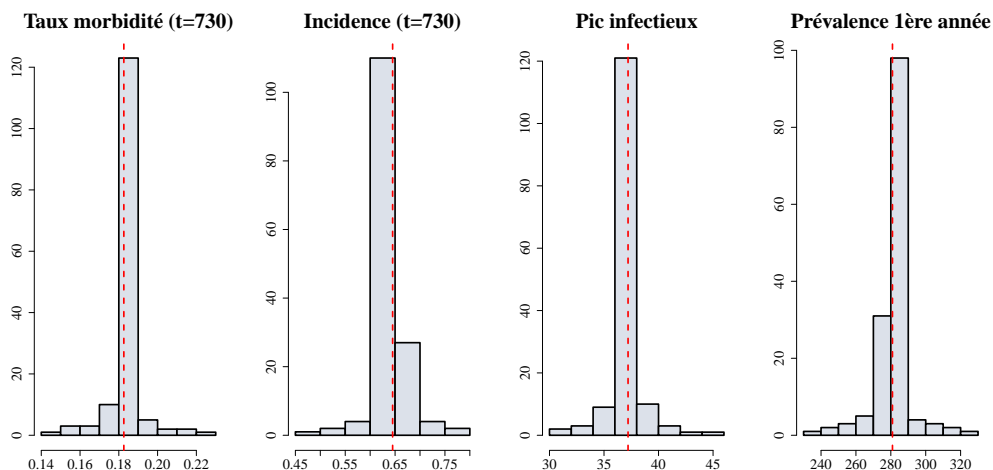


Figure 4 – Distribution des variations de sortie des 4 indicateurs épidémiologiques de l'échantillonnage OAT standard. La ligne rouge représente la valeur initiale de la sortie.

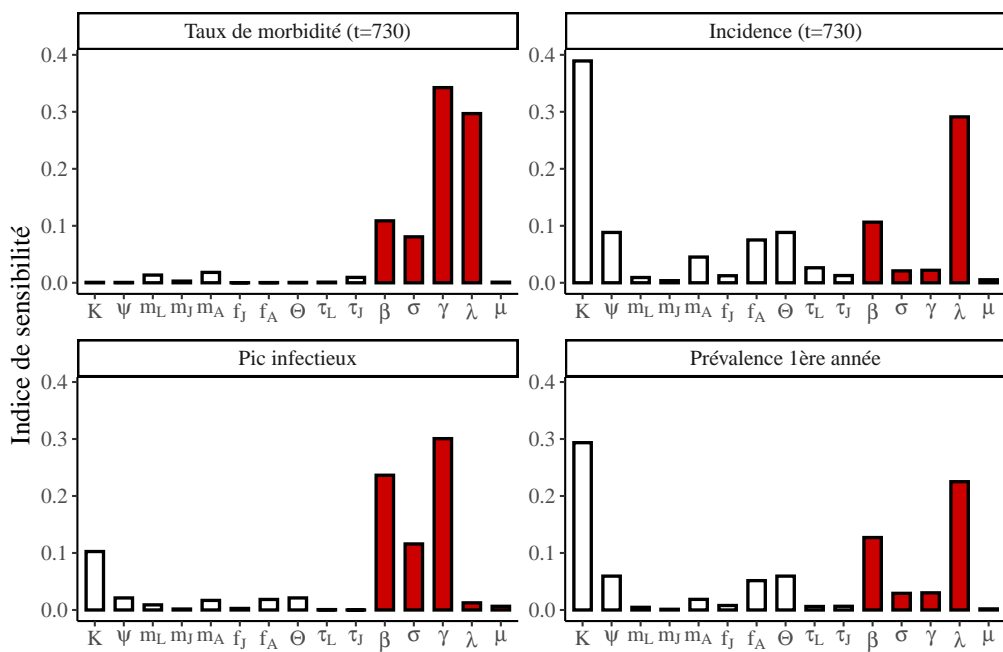


Figure 5 – Indice de sensibilité de Bauer & Hamby (Bauer and Hamby, 1991) de chacun des paramètres du modèle pour les 4 sorties du modèles.

Pour l'analyse des indices de sensibilité de Bauer & Hamby, nous avons considéré qu'un paramètre contribuait significativement lorsque son indice était supérieur ou égal à 0.2. De manière générale, les paramètres épidémiologiques ont montré un impact plus prononcé que les paramètres démographiques, avec trois paramètres en particulier, à savoir β , γ et λ , qui ont démontré une contribution significative à la variation de l'une ou plusieurs sorties du modèle (voir la Figure 5). Plus précisément, la variation du temps de récupération de la maladie, représentée par γ , a eu une influence majeure sur le taux de morbidité et sur le pic infectieux.

Le taux de transmission, noté β , a également exercé une forte influence sur le pic infectieux. L'incidence du dernier jour et la prévalence de la première année ont montré leur sensibilité au taux de perte d'immunité λ et à la capacité d'accueil de la population K , étant le seul paramètre démographique d'influence.

d. Interprétez vos résultats biologiquement et explicitez les limites possibles

Comme vu dans l'introduction du modèle, celui-ci va rapidement atteindre une stabilité du fait de la transition longue des individus dans l'état résistant, après une vague de contamination de la population. Le taux de morbidité à l'équilibre va ainsi être sensible au temps de récupération γ puisqu'il va dicter à quelle vitesse les individus vont quitter le compartiment infecté pour transiter dans celui des résistants. Le pic infectieux, correspondant quant à lui à la vague d'infection au début de la simulation, va principalement dépendre du rapport entre le taux de transmission β et le temps de récupération γ . Un taux de transmission β très élevé induira une très forte vague de contamination des sensibles, et à l'inverse un temps de récupération γ très court va limiter le pic infectieux car individus vont très rapidement sortir des compartiments infectés. Enfin, l'incidence finale et la prévalence de la première année vont beaucoup être influencé par la perte d'immunité λ et la capacité d'accueil K car ces deux paramètres vont beaucoup jouer sur le renouvellement des individus sensibles. Plus le renouvellement de ces derniers est fort, plus le nombre d'individus à infecter sera grand à l'équilibre du modèle.

Néanmoins, ces résultats sont à interpréter avec précaution. La limite principale de cette analyse est liée à (i) la méthode d'échantillonnage de l'espace des paramètres et (ii) l'indice de sensibilité, qui ne permettent d'analyser la variation que d'un unique paramètre séparément de la valeur des autres paramètres. Cela peut avoir une influence importante sur la validité des résultats puisque le plan d'échantillonnage ne permet ni de tester les interactions entre paramètres ni de tester des relations non-linéaires entre les paramètres et les valeurs de sortie du modèle. De plus, l'indice de Bauer & Hamby ne permettrait pas non plus de rendre compte d'effets d'interactions puisque l'effet de chaque paramètre est analysé avec les autres maintenus constants. Par exemple, si on considère la fonction $\mathcal{G}(x, y) = x e^y$, si x est fixé à 0, alors l'indice de sensibilité de Bauer & Hamby de y sera de 0 pour toute variation de y , alors que si $x \neq 0$, les variations de y auront un impact considérable sur la sortie de \mathcal{G} . La méthode de Morris présentée dans la partie suivante, permet de régler le premier problème et de considérer un espace de paramètre plus important.

2. Analyse de sensibilité : Morris

Présentation de la méthode

La méthode de criblage de Morris (Morris, 1991) est une analyse de sensibilité faisant partie des méthodes d'analyse OAT vues précédemment. Elle fait partie des premières approches de sélection des facteurs entrants en fonction de leur impact sur les facteurs sortants et nécessite un nombre de simulation relativement important.

Cette méthode s'applique sur un modèle au sein duquel sont choisis K facteurs à analyser. Les paramètres K du modèle seront uniformément discrétisés en un nombre choisi de *levels*, représentant leur nombre de pas de variation δ (identiquement proportionnelle pour tous les paramètres) de la valeur initiale du paramètre. L'échantillonnage est ainsi constitué d'un ensemble d'une suite de trajectoires dont le point de départ (noeud) est aléatoire et le pas est δ . Les paramètres du modèles et leurs *levels* seront ainsi conceptuellement représentés sur une grille à K dimensions ou espace des paramètres (e.g. figure 6a pour deux paramètres). Chaque noeud de cette grille, correspond ainsi à une valeur de sortie de la fonction \mathcal{F} correspond pour la valeur échantillonnée dans l'espace des paramètres pour le facteur X_k .

La méthode de Morris est basée sur l'étude des variations de sortie de la fonction \mathcal{F} entre deux points issus d'une même trajectoire (voir figure 6a). Soit deux points successifs dans l'espace d'échantillonnage de coordonnées $A^{x,y}$ et $A^{x,y+1}$ associés à une variation δ du facteur X_k . Cette variation δ induit $\Delta_k^{(i)} \mathcal{F}$ dans la sortie du modèle. Cette variation dans la sortie est appelée effet élémentaire du facteur X_k de la fonction \mathcal{F} .

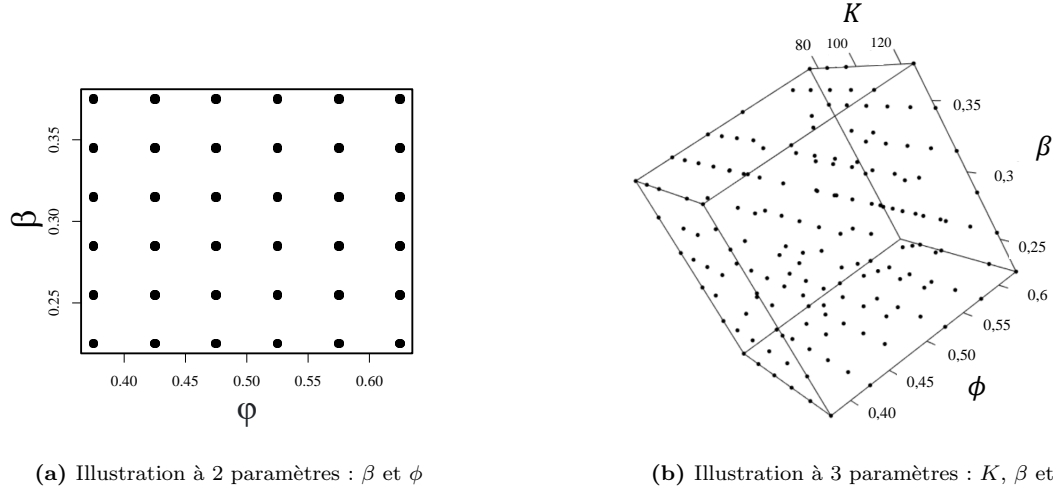


Figure 6 – Exploration de l’espace des paramètres avec la méthode Morris pour (a) 2 dimensions et (b) 3 dimensions

et est définie par :

$$\Delta_k^{(i)} \mathcal{F} = \frac{\mathcal{F}(x_1^{(i)}, \dots, x_k^{(i)} + \delta, \dots, x_K^{(i)}) - \mathcal{F}(x_1^{(i)}, \dots, x_k^{(i)}, \dots, x_K^{(i)})}{\delta}$$

Ici $x_k^{(i)}$ représente la valeur sur l’axe associée au paramètre X_k . Ainsi, la méthode de Morris r nombre de trajectoire en $N = r(K+1)$ simulations permettant le calcul de r effets élémentaires par K paramètres. L’indice de sensibilité μ_k^* relatif au paramètre X_k représente la moyenne des effets absolus des effets élémentaires :

$$\mu_k^* = \frac{1}{r} \sum_{i=1}^r \left| \Delta_k^{(i)} \mathcal{F} \right|$$

Morris propose un autre indice σ basé sur la variance des effets sur le nombre total de trajectoire r . Pour le facteur X_k :

$$\sigma_k = \sqrt{\frac{1}{r-1} \sum_{i=1}^r \left(\Delta_k^{(i)} \mathcal{F} - \mu_k^* \right)^2}$$

Ainsi, μ^* évalue l’impact globale (effet principal) d’un paramètre sur la réponse (i.e. un changement de valeur élevé induit un changement de valeur élevé dans la sortie du modèle) alors qu’un σ élevé représente une mesure de non-linéarité et/ou d’interaction avec d’autres paramètres. Une petite valeur de σ indique que l’effet du paramètre sur la sortie du modèle dépend fortement de la valeur relative prise par ce paramètre tandis que qu’un σ faible indique un effet similaire quelque soit la valeur relative du paramètre impliquant que ce paramètre est indépendant des valeurs des autres paramètres.

Les paramètres peuvent être classés en 3 catégories d’effet : (i) négligeable, (ii) additif et linéaire et (iii) non-linéaire et/ou interactifs. A fin de trier et catégoriser les facteurs d’entrée, un diagramme de Morris peut être obtenu en traçant σ_k en fonction de μ_k^* avec un ratio $\sigma_k/\mu_k^* \approx 1$ indiquant un effet non-linéaire et/ou d’interaction et $\sigma_k/\mu_k^* \approx 0.2$ indiquant un effet additif et linéaire et les paramètres proches de $\mu_k^* = 0$ sont considérés comme négligeables (Corrado and Mechri, 2009; Menberg et al., 2016).

Utilisation sur R

Afin de faciliter l'utilisation de l'analyse de sensibilité Morris, il est possible d'utiliser le package R 'sensitivity'. Ce dernier comprend une fonction *morris()*, avec dans notre cas les paramètres suivants :

- **model** = : la fonction du modèle
- **factors** = : le nombre de paramètre du modèle
- **r** = : le nombre de trajectoires
- **design** = : une liste d'argument comprenant :
 - *type* = : le type d'analyse, ici 'oat'
 - *levels* = : le nombre de valeurs pour chaque facteur testé
 - *grid.jump* = : le nombre de sauts de valeurs réalisés à chaque trajectoire (en général de valeur $\text{levels}/2$)
- **binf** = : un vecteur des valeurs bornes inférieures des paramètres
- **bsup** = : un vecteur des bornes supérieures des paramètres
- **scale** = : mise à l'échelle du sein d'exploration en fonction des bornes inférieurs et supérieures de scale

Script 2 : Script de l'analyse de MORRIS (Morris, 1991) avec le package sensitivity (Pujol et al., 2015)

```
# PACKAGES
library(sensitivity)
library(ggplot2)
library(latex2exp)
library(gridExtra)
library(ggrepel)

# IMPORTATION FONCTION DE BASE ET DES VALEURS INITIALES
source("FONCTION_BASE.R")

# ANALYSE DE SENSIBILITE MORRIS -----
# Cration des bornes minnum et maximum
bornes <- vector(mode = "list", length = 2)
bornes[[1]] <- .75 * ValNominale # borne infrieure
bornes[[2]] <- 1.25 * ValNominale # borne suprieure

# Fonction morris
AS.morris <- morris(
  model = modAppli,
  factors = c("K", "sr", "m1", "m2", "m3", "f2", "f3", "portee", "t1", "t2", "trans", "lat", "rec", "loss", "madd"),
  r = 100, # Nombre d'itration
  design = list(
    type = 'oat',
    levels = 6,
    grid.jump = 3
  ),
  # Gamme de variation [min; max]
  binf <- bornes[[1]],
  bsup <- bornes[[2]]
)
```

a. Fournissez une figure équivalente à celle donnée dans le cours et interprétez-la.

Le code de l'analyse de Morris est visible ci-dessus (Script 2). La figure figure 8 représente les valeurs de σ (la variance des effets élémentaires de chaque paramètre sur leurs gammes de valeurs) en fonction des valeurs de μ^* (la moyenne des valeurs absolues des effets élémentaires. D'après les droites du ratio σ_k/μ_k^* , la totalité des paramètres varient de façon linéaire avec les sorties du modèle ($\sigma_k/\mu_k^* \approx 0.2$ (Corrado and Mechri, 2009; Menberg et al., 2016)). Nous avons considéré pour que les paramètres ayant un μ^* ayant un effet global supérieur à 20% de la sortie initiale ($\mu^* > 0.2 \cdot \mathcal{F}(x_1, \dots, x_K)$) contribuait significativement (soit 0.04, 0.13, 7.4 et 56.2 respectivement pour le taux de morbidité, l'incidence, le pic infectieux et la prévalence). L'analyse de sensibilité de Morris sur notre modèle montre que le taux de récupération de la maladie γ et λ le taux de perte d'immunité participent le plus à la variation du taux de morbidité. L'incidence est de la maladie au dernier jour est sensible aux variations de la capacité d'accueil K et λ tandis que le pic infectieux

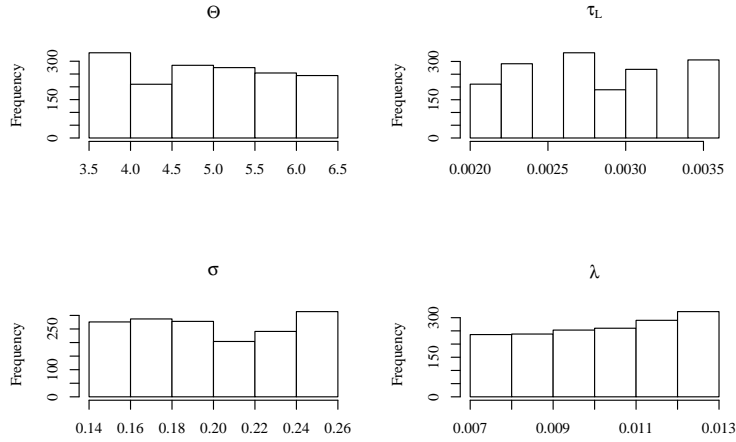


Figure 7 – Répartition des valeurs des paramètres Θ , τ_L , σ et σ lors du plan d'échantillonnage de Morris

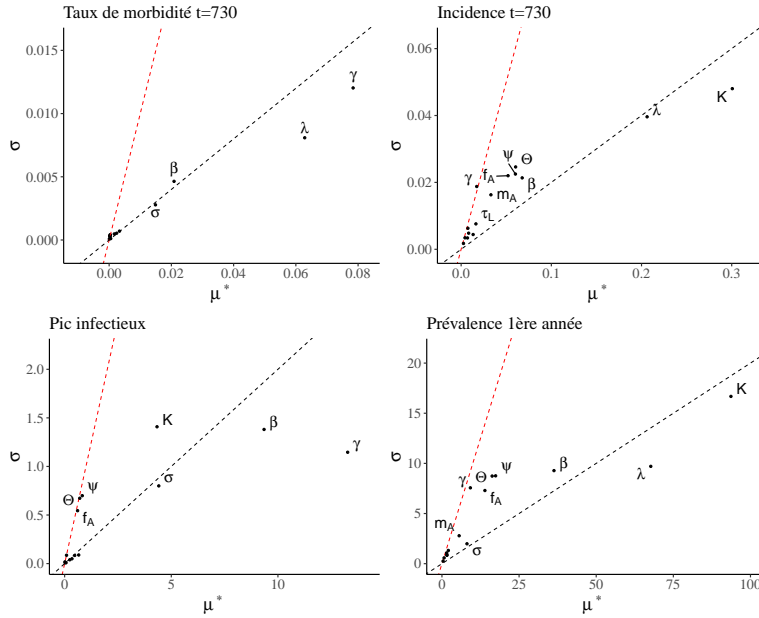


Figure 8 – Moyennes et variances des contributions des paramètres dans la méthode Morris. La ligne rouge correspond au ratio $\sigma/\mu^* = 1$ et la ligne noire au ratio $\sigma/\mu^* = 0.2$

est sensible aux variations des paramètres β et γ . Enfin, la prévalence de la première année est sensible aux variations de K et λ .

b. Qu'en concluez-vous si vous comparez les deux approches et les résultats obtenus ?

Les conclusions des deux analyses OAT standard et Morris sont semblables. Ce sont les mêmes facteurs épidémiques et démographiques (tout particulièrement ceux liés à l'immunité) qui influencent les mêmes indicateurs. La méthode OAT standard et Morris ne diffèrent en soi que par leur exploration de l'espace des paramètres, ce qui résulte en une plus grande variance dans les sorties du modèle malgré des bornes de va-

riations similaires (figure 9 comparé à la figure 4). L'analyse de Morris permet donc d'être plus précise dans l'estimation de la sensibilité car elle peut inclure une partie des effets d'interactions d'un paramètre dans son indice et surtout d'estimer quels sont les paramètres sujet à des effets d'interactions et/ou non-linéaires. En revanche, la méthode de Morris ne permet pas de décomposer la part d'effet imputable à l'effet principal d'une variable des effets d'interactions. Elle peut néanmoins identifier les variables potentiellement sujettes à ces effets d'interactions et/ou non-linéaires à l'aide du graphique de type figure 8 selon la proximité des paramètres avec le ratio σ_k/μ_k^* dans le plan (μ_k^*, σ_k) . En se basant sur le graphique de Morris (figure 8), la totalité des paramètres d'influence ont une relation linéaire avec les sorties du modèle. Il n'est donc pas surprenant que les deux résultats soient identiques puisque seul un effet principal était à estimer pour les paramètres en l'absence d'interaction et de non-linéarité. Il est donc probable que les résultats de ces deux analyses soient fiables puisqu'elles parviennent à quantifier des indices de sensibilités valides lorsque seuls des effets principaux / globaux de paramètres sont à estimer.

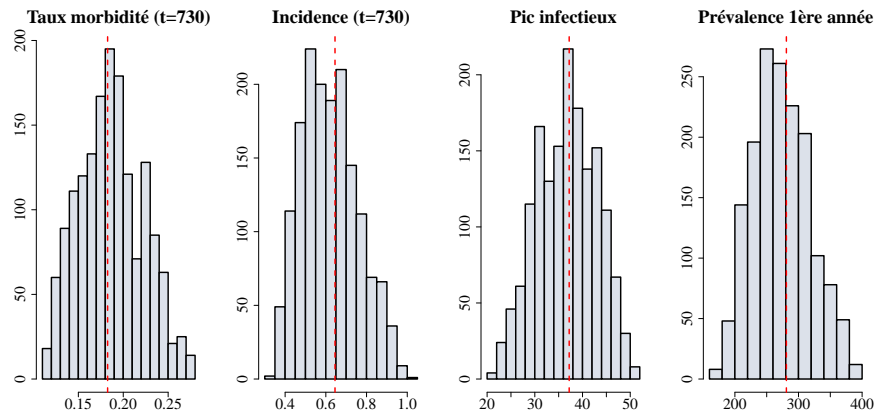


Figure 9 – Distribution des variations de sortie des 4 indicateurs épistémologiques de l'échantillonnage de Morris. La ligne rouge représente la valeur initiale de la sortie. (Morris, 1991)

3. Analyse de sensibilité : FAST

La méthode FAST (*Fourrier Amplitude Sensitivity Technique*) fait partie des méthodes d'analyse globale (Cukier et al., 1973). Elle est similaire à une analyse de variance basé sur la décomposition des sorties du modèle et permet de discriminer et quantifier la contribution individuelle de chaque paramètre et ses interactions dans la variance totale du résultat (Faivre et al., 2013). La méthode FAST permet également de réaliser une analyse de sensibilité même quand la relation entre les variations de paramètres et les sorties du modèles sont non-linéaires (Faivre et al., 2013; Saltelli et al., 1999). Elle est basée sur une décomposition de Fourier pour la création de l'espace des paramètres.

Dans la méthode FAST, l'échantillonnage est constitué de trajectoires déterministes définie par un jeu de fréquence $\omega = \omega_1, \dots, \omega_K$ associé aux K paramètres généré en utilisant la transformation de Fourier pour générer des fréquences spécifiques pour simuler les valeurs des paramètres à tester. En d'autres termes, cela génère des oscillations (variations périodiques) dans les valeurs prises par les paramètres de façon à investir le plus grand espace possible dans l'espace des paramètres (figure 10). La répartition des valeurs d'échantillonnage dans l'espace des paramètres dépendra de leur loi de distribution (e.g. gaussienne, uniforme...). Dans cette méthode, tous les paramètres varient simultanément à chaque trajectoire. On a donc KN valeurs prises par les paramètres, avec N le nombre d'itération. Ici nous utiliserons une variante récente de FAST appelée 'méthode FAST étendue' ou '*extended FAST analysis*' dans laquelle on aura une trajectoire par paramètre X_i associé à une fréquence ω_i élevée, et une faible valeur pour les autres fréquences ω_i avec $1 \leq j \leq 15$ et $j \neq i$.

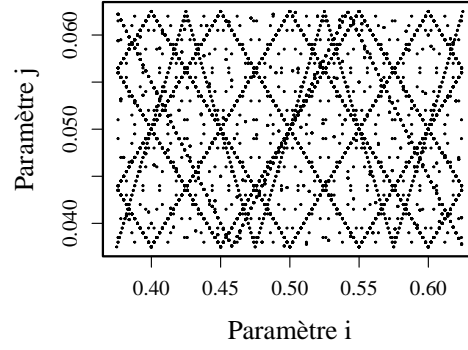


Figure 10 – Exploration de l’espace des paramètres avec la méthode FAST

Deux indices de sensibilité sont définis pour cette méthode : L’indice de sensibilité principal (S) :

$$S(x_i) = \frac{Var[E(y)|x_i]}{Var(y)}$$

et l’indice de sensibilité total (ST) :

$$ST(x_i) = \frac{Var[E(y)|x_j, j \neq i]}{Var(y)}$$

Avec $Var(y)$ la variance totale de sortie du modèle lorsque tous les facteurs varient. SI mesure l’importance de la variance de l’espérance de y conditionnellement au facteur x_i . Autrement dit, cet indice mesure la variabilité des simulations du modèle lorsque ces simulations sont moyennées pour les autres paramètres incertains que x_i . C’est une mesure de l’effet principale. L’indice totale TSI mesure l’espérance de la variance de y sur les paramètres incertains autres que x_i . Il correspond à une mesure de l’effet principal de x_i plus l’effet des interactions de ce facteurs avec les autres facteurs incertains. Cet indice peut-être interprété comme la fraction moyenne d’incertitude qui resterait si tous les facteurs incertains sauf x_i était parfaitement connus (i.e. seul x_i serait incertain). Plus précisément, l’indice total de sensibilité de Sobol (Sobol, 1993 ; Saltelli et al., 1999) est calculé comme la somme de tous les indices de sensibilité incluant ce paramètre. En guise d’illustration (exemple repris de Saltelli et al., 1999), on considère un modèle à 3 paramètres. On cherche à calculer l’indice de sensibilité total de Sobol S_T du paramètre 1. On a :

$$ST_1 = S_1 + S_{12} + S_{13} + S_{123}$$

avec S_1 l’index de sensibilité de première ordre du facteur 1, S_{1j} l’indice de sensibilité de second ordre issu de l’interaction entre le paramètre 1 et le paramètre j (avec $j \neq 1$). Cette indice permet donc un calcul de l’effet total de sensibilité pour chaque paramètre. Les deux indices sont compris entre 0 et 1. En pratique, une valeur de $ST(x_i)$ proche de zéro indique qu’il n’est pas nécessaire de connaître x_i avec une grande précision (inversement pour 1).

a. Identifiez et explicitez le package que vous utilisez pour cette analyse ; donnez les éléments les plus importants qui peuvent être modulés lors de cette analyse. Notamment, comment choisir le nombre de scénarios à réaliser ?

Pour réaliser la méthode FAST, nous avons utilisé le package ‘sensitivity’ (Pujol et al., 2015). Plus précisément, nous avons utilisé la fonction ‘fast99()’ que nous avons fait tourner sans modèle pour générer la matrice des scénarios à tester, puis nous avons lancé notre modèle ModAppli() sur l’ensemble des scénarios et récupérer les résultats. Ces résultats ont ensuite été intégrés dans la fonction ‘tell()’ qui a permis de calculer les indices de sensibilité.

La fonction fast99() a plusieurs paramètres important :

- '**model** =' correspondant au modèle dans lequel l'analyse doit être réalisé. Si on ne précise pas de modèle, alors la fonction générera simplement la matrice des scenarii pour chaque paramètre
- '**factors** =' correspondant au nombre de paramètres du modèle
- '**n** =' le nombre d'itération. En considérant que FAST est basée sur une analyse de variance, on considère que 5 valeurs minimum sont nécessaires pour évaluer un paramètre statistique. Ainsi, en comptant les effets principaux et d'interactions, le nombre d'itération nécessaire se situe dans l'ordre de $5K * (K - 1)$ itérations (avec K le nombre total de paramètre).
- '**q** =' la distribution de probabilité de l'incertitude des paramètres (ici, nous avons choisi une distribution uniforme puisque nous avons pas d'a priori sur la répartition de l'incertitude dans nos intervalles de valeurs définis)
- '**q.arg** =' qui correspond aux paramètres de la distribution précisé dans l'argument 'q' (e.g. la moyenne et la variance pour une loi gaussienne, le minimum et le maximum pour une loi uniforme)

Script 3 : Script de l'analyse FAST étendu (Cukier et al., 1973; Saltelli et al., 1999) avec le package sensitivity (Pujol et al., 2015)

```
# PACKAGES
library(sensitivity)

# IMPORTATION FONCTION DE BASE ET DES VALEURS INITIALES
source("FONCTION_BASE.R")

# INITIALISATION
# --- Noms des parametres
names_para <- c("K", "sr", "m1", "m2", "m3", "f2", "f3", "portee", "t1", "t2", "trans", "lat", "rec", "loss", "madd")

# --- Liste des parametres de distribution pour chaque parametre
q.arg <-
  apply(cbind(0.75 * unname(ValNominale),
             1.25 * unname(ValNominale)),
        MARGIN = 1,
        function(x) {
          list(min = x[1], max = x[2])
        })

# ANALYSE FAST
# FAST (100 iterations)
# --- Generation des valeurs de parametres pour les differents scenarii
as_fast_100 <- sensitivity::fast99(model = NULL,
                                  factors = names_para,
                                  n = 100,
                                  q = "qunif",
                                  q.arg = q.arg)

# --- On run le modele modAppli sur ces valeurs
sortie_100 <- modAppli(as_fast_100$X)
colnames(sortie_100) <- c("taux morbidit", "incidence t=730", "pic infectieux", "prvalene ire annee")

# --- Calcul des variances et indices de Sobol pour chaque sortie pour chaque sortie du modele
fast_100_sortie1 <- as_fast_100
fast_100_sortie2 <- as_fast_100
fast_100_sortie3 <- as_fast_100
fast_100_sortie4 <- as_fast_100

# (tell() calcul partir des scenarii les indices de Sobol et compte l'objet 'incomplet' x)
tell(x = fast_100_sortie1, y = sortie_100[,1]) # taux de morbidite
tell(x = fast_100_sortie2, y = sortie_100[,2]) # incidence t=730
tell(x = fast_100_sortie3, y = sortie_100[,3]) # pic infectieux
tell(x = fast_100_sortie4, y = sortie_100[,4]) # prevalence Aere annee

# -----

# FAST (1000 iterations)
# --- Generation des valeurs de parametres pour les differents scenarii
as_fast_1000 <- sensitivity::fast99(model = NULL,
                                   factors = names_para,
```

```

n = 1000,
q = "qunif",
q.arg = q.arg)

# --- On run le modele modAppli sur ces valeurs
sortie_1000 <- modAppli(as_fast_1000$X)
colnames(sortie_1000) <- c("taux morbidit", "incidence t=730", "pic infectieux", "prvalene 1re annee")

# --- Calcul des variances et indices de Sobol pour chaque sortie pour chaque sortie du modele
fast_1000_sortie1 <- as_fast_1000
fast_1000_sortie2 <- as_fast_1000
fast_1000_sortie3 <- as_fast_1000
fast_1000_sortie4 <- as_fast_1000

tell(x = fast_1000_sortie1, y = sortie_1000[,1]) # taux de morbidite
tell(x = fast_1000_sortie2, y = sortie_1000[,2]) # incidence t=730
tell(x = fast_1000_sortie3, y = sortie_1000[,3]) # pic infectieux
tell(x = fast_1000_sortie4, y = sortie_1000[,4]) # prevalence 1re annee

```

b. Réalisez un premier échantillonnage avec peu de scénarios par paramètre (100). Combien cela fait-il de simulations au total ?

Le nombre de simulation total est 1500 ($N * K = 100 * 15$ paramètres).

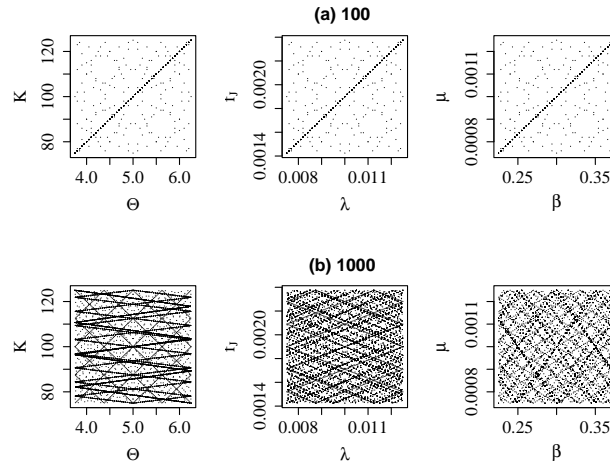


Figure 11 – Exploration de l'espace des paramètres avec la méthode FAST pour les couples de paramètres $[\Theta; K]$, $[\lambda; f_J]$ et $[\beta; \mu]$ pour (a) 100 et (b) 1000 itérations.

c. Sauvegardez le tableau obtenu et réalisez un graphique illustrant les variations des paramètres obtenues.

Voir script 3 ci-dessus. Le graphique illustrant les variations des paramètres est visible sur la figure 11a avec 3 exemples de paramètres variant 2 à 2 en guise d'illustration.

d. Réalisez alors un second échantillonnage avec beaucoup de scénarios par paramètre (1000). De même, sauvegardez le tableau obtenu et illustrez graphiquement votre échantillonnage.

Voir script 3 ci-dessus. Le graphique illustrant les variations des paramètres est visible sur la figure 11b.

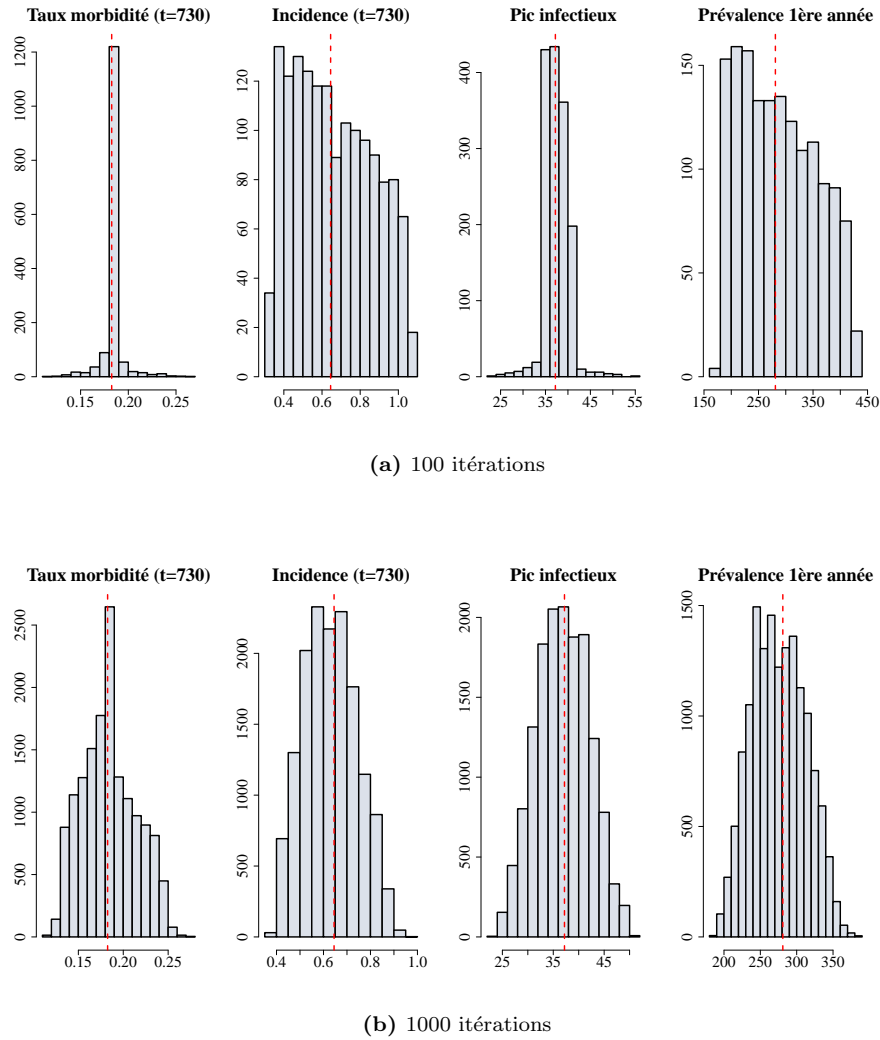


Figure 12 – Distribution des fréquences de variations de sortie des 4 indicateurs épistémologiques de l'échantillonnage FAST avec 100 et 1000 itérations. La ligne rouge représente la valeur initiale de la sortie.

e. Comparez les deux échantillonnages et discutez des implications possibles.

Les échantillonnages avec 100 répétitions (figure 12.a) illustrent un manque d'itération puisque l'espace des paramètres deux à deux est largement moins occupé qu'avec 1000 répétitions (figure 12.b). Il semble que les oscillations spécifiant la variation des paramètres ne peuvent pas avoir lieu contrairement au cas à 1000 répétitions. En terme d'implication, les calculs des effets de chaque paramètre sera bien moins puissant avec 100 itérations qu'avec 1000, surtout pour le calcul des effets interactions. En effet, la ligne droite de coefficient 0 pour les itérations à 100 correspond à des tests d'effets de 'relations d'interactions linéaire' (tous les paramètres varient de 1 pas à chaque fois), réduisant le champ des possible en terme d'interaction possible. Cela peut mener à des estimations biaisés dans l'estimation des effets des paramètres Avec aussi peu de simulations, il est possible que les limites similaires aux méthodes OAT standard et de Morris concernant le plan d'échantillonnage de variation s'appliquent à l'analyse de sensibilité et ainsi de rater de sous-estimer ou surestimer des effets d'interactions entre paramètre.

f. Lancez le modèle sur ces deux échantillonnages en sauvegardant bien à chaque fois les valeurs des sorties dans chaque cas.

Voir script 3 ci-dessus.

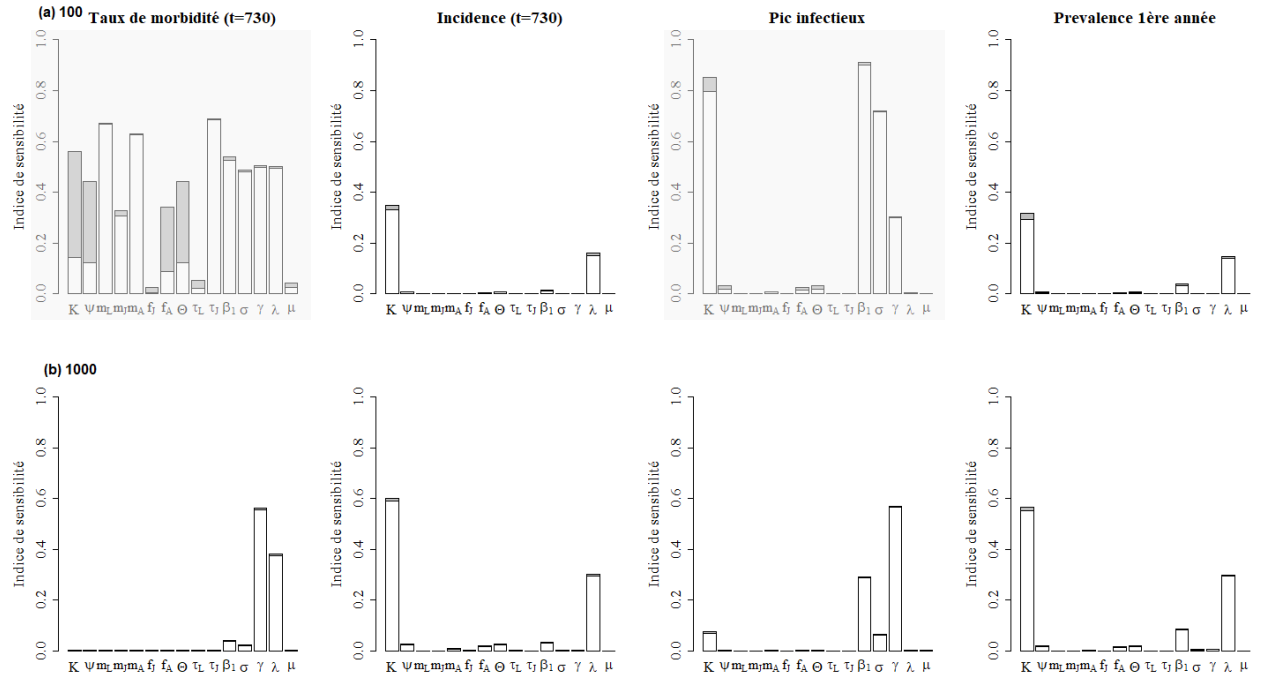


Figure 13 – Indice principal et indice total de chaque paramètre pour les indices interprétables pour (a) 100 itérations et (b) 1000 itérations FAST. Les plots grisés correspondent aux sorties non-interprétables. Blanc : effet principal, gris : effet d'interaction, la somme des deux représentant la contribution totale à la variance.

g. Visualisez la distribution des sorties obtenues : qu'en pensez-vous ? Quel impact dans le cadre d'une analyse de variance (telle que faite pour calculer les indices de sensibilité) ? Pouvez-vous donc interpréter toutes les sorties et pourquoi ?

Les distributions des valeurs de sorties permettent d'estimer la distribution de chaque indicateur épidémiologique. Les distributions de nos 4 sorties montrent un écart important entre l'échantillonnage à 100 itérations et 1000 itérations (figure 12). En effet, les distributions à 1000 itérations sont toutes quasi Gaussienne à l'exception peut-être du taux de morbidité, avec une sur représentation de fréquence de valeurs très proche de la moyenne. A échantillonnage à 100 itérations, les indicateurs de taux de morbidité et de pic infectieux dénotent particulièrement puisqu'il n'y a pratiquement pas de variance autour des valeurs de sorties avec une distribution quasi-uni modale pour le taux de morbidité. L'incidence et la prévalence quant à elle ont une distribution plus proche de la distribution uniforme. Cela peut avoir de fortes conséquences sur une analyse de variance puisqu'elle nécessite une certaine hétérogénéité des données pour pouvoir estimer correctement les paramètres. De plus, la méthode FAST repose sur l'hypothèse implicite que les réponses du modèle sont régulières en fonction des entrées et que les variations des paramètres couvrent l'ensemble du domaine d'incertitude possible pour une estimation fiable des paramètres ((Faivre et al., 2013)). Les écarts de distributions pour le taux de morbidité et pour le pic infectieux entre 100 et 1000 itérations est donc problématiques puisque cela signifie que 100 itérations n'étaient pas suffisante pour couvrir ces deux hypothèses sous-jacentes puisque manifestement la distribution des réponses du modèle n'étaient pas stabilisées à en juger par les changements de distributions entre les différents nombres d'itérations. L'estimation des paramètres auraient donc été faussés.

h. Pour les sorties interprétables, calculez les indices de sensibilité principaux (effet principal) et d'ordre 1 (interactions deux-à-deux) et visualisez graphiquement les résultats.

Les résultats sont disponibles dans la figure ?? . Les sorties non-interprétables ont été grisées mais conservées.

i. Comparez les résultats des deux échantillonnages : qu'en concluez-vous ?

Ici, nous avons considéré que toutes les variables contribuant à plus de 10% à la variance totale étaient majeures (Beaunée et al., 2015)). Les résultats des sorties interprétables montrent des résultats très similaires, avec les mêmes paramètres contribuant significativement à chaque sortie.

j. Interprétez biologiquement le cas à 1000 scénarios par paramètre.

Les mêmes paramètres influençant les mêmes sorties, l'interprétation biologique est identique à celle de la question d de la partie de l'analyse OAT standard.

6. Discussion des résultats des analyses de sensibilité.

les résultats sont-ils cohérents entre méthodes ? Quelles implications possibles ? Que feriez-vous ensuite en termes d'analyse de ce modèle ?

Trois méthodes d'analyse de sensibilité ont été appliquées à un modèle d'épidémiologie à temps discret type SEIR structuré en âge : OAT standard, la méthode de Morris et la méthode FAST étendue. Les résultats entre les 3 analyses étaient similaires et leurs conclusions semblables. Nous avons montré que l'ensemble des paramètres qui contribuaient significativement à la sensibilité du modèle variaient de façon linéaire avec ses réponses et aucune interaction n'a été détectée ni par le graphique de Morris (figure 8) ni par l'indice de sensibilité totale de la méthode FAST (pas de partie grisée dans la figure ??). Compte tenu du fait que la principale différence d'utilisation de ces techniques est la considération des effets d'interaction et/ou non-linéarité, il est alors possible que ce soit la raison pour laquelle ces 3 analyses aient généré des résultats similaires. En revanche, dans le cas d'un modèle où certaines réponses auraient co-variées avec plusieurs paramètres, on aurait observé une différence nette de résultats, notamment entre les méthodes locales OAT standard & Morris et la méthode FAST, cette dernière s'avérant plus pertinente dans ce cas de figure. Bien que la méthode FAST soit supérieure en tous points aux autres méthodes, elles nécessitent un nombre important de simulations et un effort de ressource de temps de calcul. Pour éviter cette sur-monopolisation de ressources, nous aurions pu avoir recours à une construction de métamodèles ((Faivre et al., 2013)), sorte d'émulateur du modèle initial, pour déterminer à moindre coût de calcul si des variables étaient en interactions et quelles étaient ces variables.

À l'issue de ces analyses de sensibilité, nous aurions pu ensuite estimer la précision de l'estimation des indices de sensibilité obtenus. En effet, nous ne pouvons pas résoudre le modèle analytiquement, donc ces indices ont dû être estimés numériquement, ce qui par définition comporte une part d'erreur non-négligeable qui doit être explorée (Faivre et al., 2013; Saltelli et al., 2019). Toujours concernant la précision, il serait intéressant d'explorer dans quelle mesure il serait coûteux d'un point de vue méthodologique de préciser l'estimation des paramètres d'immunité (paramètres de contributions principales). Nous pourrions également aussi explorer comment les paramètres contributifs participent à la variation des réponses du modèle (sens de l'effet, taille de l'effet...).

6. Modification du modèle

a. Documenter cette modification

Jusqu'à présent, seule la transmission des Ranavirus par contact direct entre les individus a été considérée. Cependant, la littérature montre que ces virus se transmettent aussi de façon indirecte en étant libérés par les individus dans l'eau (Gray et al., 2009). Nous allons donc ajouter à notre système un compartiment pour les pathogènes sous forme libre, en se basant sur un modèle déjà existant et développé pour l'intégration des formes libres du Covid-19 (Mwalili et al., 2020). Nous considérerons aussi que les trois stades de développement de l'hôte n'ont pas le même degré d'interaction avec ce nouveau compartiment. En effet, le stade larvaire est beaucoup plus inféodé à l'eau et plus sensible au virus. Ces individus vont donc répandre des virus dans l'environnement de façon plus importante s'ils sont infectés. Il seront aussi en retour plus facilement infectés par ces derniers. Pour approfondir cette idée, des études ont montré que la sensibilité des têtards à la maladie était plus importante du fait de leur exposition à de nombreux stress, comme la compétition intraspécifique, la prédation ou encore la pollution (Gray et al., 2009). Les équations du modèle modifié sont ainsi les suivantes :

$$\left\{ \begin{array}{l} L(t+1) = \left\{ \begin{array}{l} L_S(t+1) = L_S(t) + \max \left[0; \psi \Theta \cdot \left(f_J J(t) + f_A A(t) \right) \cdot \left(1 - \frac{N}{K} \right) \right] + \lambda L_R(t) - L_S(t) \cdot \left(m_L + \tau_L + \beta_1 \frac{I(t)}{N} + \alpha_L \beta_2 P(t) \right) \\ L_E(t+1) = L_E(t) + \beta_1 L_S(t) \frac{I(t)}{N} + \alpha_L \beta_2 P(t) \cdot L_S(t) - L_E(t) \cdot (m_L + \tau_L + \sigma) \\ L_I(t+1) = L_I(t) + \sigma L_E(t) - L_I(t) \cdot (m_L + \tau_L + \mu + \gamma) \\ L_R(t+1) = L_R(t) + \gamma L_I(t) - L_R(t) \cdot (m_L + \tau_L + \lambda) \end{array} \right. \\ \\ J(t+1) = \left\{ \begin{array}{l} J_S(t+1) = J_S(t) + \tau_L L_S(t) + \lambda J_R(t) - J_S(t) \cdot \left(m_J + \tau_J + \beta_1 \frac{I(t)}{N} + \alpha_J \beta_2 P(t) \right) \\ J_E(t+1) = J_E(t) + \tau_L L_E(t) + \beta_1 J_S(t) \frac{I(t)}{N} + \alpha_J \beta_2 P(t) \cdot J_S(t) - J_E(t) \cdot (1 + m_J + \tau_J + \sigma) \\ J_I(t+1) = J_I(t) + \tau_L L_I(t) + \sigma J_E(t) - J_I(t) \cdot (m_J + \tau_J + \mu + \gamma) \\ J_R(t+1) = J_R(t) + \tau_L L_R(t) + \gamma J_I(t) - J_R(t) \cdot (m_J + \tau_J + \lambda) \end{array} \right. \\ \\ A(t+1) = \left\{ \begin{array}{l} A_S(t+1) = A_S(t) + \tau_J J_S(t) + \lambda A_R - A_S(t) \cdot \left(m_A + \beta_1 \frac{I(t)}{N} + \alpha_A \beta_2 P(t) \right) \\ A_E(t+1) = A_E(t) + \tau_J J_E(t) + \beta_1 A_S(t) \frac{I(t)}{N} + \alpha_A \beta_2 P(t) \cdot A_S(t) - A_E(t) \cdot (m_A + \sigma) \\ A_I(t+1) = A_I(t) + \tau_J J_I(t) + \sigma A_E(t) - A_I(t) \cdot (m_A + \mu + \gamma) \\ A_R(t+1) = A_R(t) + \tau_J J_R(t) + \gamma A_I(t) - A_R(t) \cdot (m_A + \lambda) \end{array} \right. \\ \\ P(t+1) = P(t) + \eta_L L_I(t) + \eta_J J_I(t) + \eta_A A_I(t) - P(t) \cdot \left(m_P + \alpha_L \beta_2 L_S(t) + \alpha_J \beta_2 J_S(t) + \alpha_A \beta_2 A_S(t) \right) \end{array} \right.$$

Avec :

$$L(t) = L_S(t) + L_E(t) + L_I(t) + L_R(t)$$

$$J(t) = J_S(t) + J_E(t) + J_I(t) + J_R(t)$$

$$A(t) = A_S(t) + A_E(t) + A_I(t) + A_R(t)$$

$$N(t) = L(t) + J(t) + A(t)$$

$$I(t) = L_I(t) + J_I(t) + A_I(t)$$

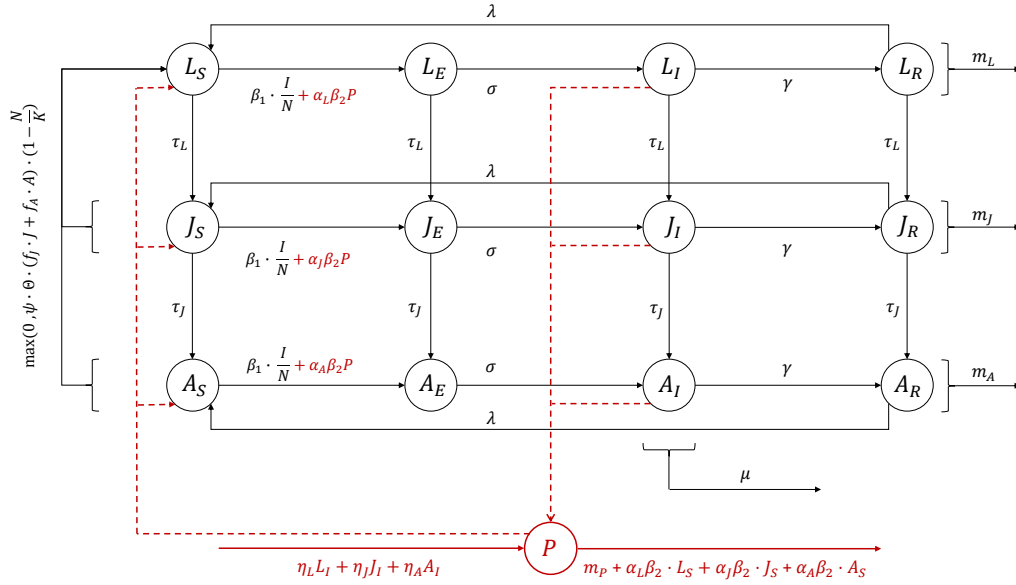


Figure 14 – Schéma de transition du modèle modifié.

Paramètre	Code R	Valeur	Définition
η_L	eta1	1	Taux de virus répandu dans l'environnement par les stades larvaires infectés
η_J	eta2	0.1	Taux de virus répandu dans l'environnement par les stades juvéniles infectés
η_A	eta3	0.2	Taux de virus répandu dans l'environnement par les stades adultes infectés
m_p	mpath	$1/58 \text{ jours}^{-1}$	Taux de mortalité du pathogène sous forme libre
β_2	trans2	0.003	Taux de transmission indirect de la maladie
α_L	a1	1	Proportion d'interaction des stades larvaires avec le milieu contaminé
α_J	a2	0.4	Proportion d'interaction des stades juvéniles avec le milieu contaminé
α_A	a3	5	Proportion d'interaction des stades adultes avec le milieu contaminé

Table 2 – Nouveaux paramètres ajoutés au modèle.

La modification du modèle inclue donc un nouveau compartiment P (figure 14 pour les formes libres du virus). Le flux entrant est composé du nombre d'unité virale déposés par chaque individu infectieux par pas de temps. Nous avons ainsi considéré que les stades larvaires, étant entièrement inféodés à l'eau, seraient ceux qui déposent le plus de virus. A l'opposé, nous avons considéré que les jeunes individus matures, plus actifs dans la recherche de partenaire en dehors des marres, déposeraient moins de virus (Gray et al., 2009). Ce postulat est ainsi représenté mathématiquement par les taux de virus rependus η_i (Mwalili et al., 2020). La même justification s'applique pour la proportion d'interaction des individus avec l'environnement α_i qui pondère leur contamination (Mwalili et al., 2020). Le flux sortant du compartiment des pathogènes libres prend en compte la durée de vie de ces derniers, qui a été prise à 58 jours en milieu humide (Nazir et al., 2012). Ce flux sortant part aussi de l'hypothèse que lorsqu'une unité virale infecte un individu sensible, elle est internalisée et sortie du compartiment libre. Enfin, la contamination des individus sensibles par les formes libres du virus se fait par une fonction de transmission densité dépendante.

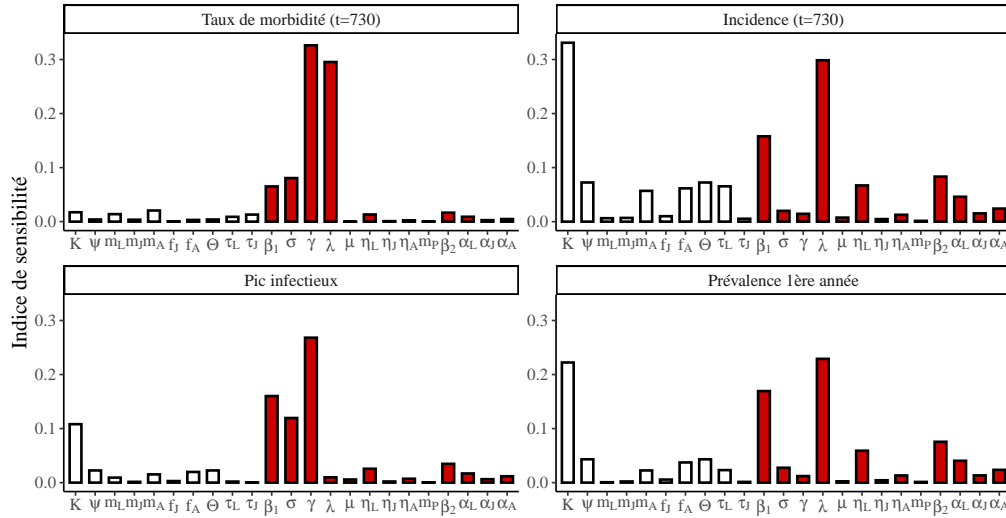


Figure 15 – Indice de sensibilité de Bauer & Hamby (Bauer and Hamby, 1991) de chacun des paramètres du modèle modifié pour les 4 sorties du modèles.

b. Réaliser les mêmes analyses (OAT, Morris, FAST)

On analyse donc avec un seuil de 0.2, comme précédemment, l'analyse de sensibilité OAT réalisée pour le modèle modifié (figure 15). Le taux de morbidité est ainsi sensible au taux de perte d'immunité λ et au taux de récupération γ . L'incidence, tout comme la prévalence de la première année, sont principalement sensibles à la capacité limite K et au taux de perte d'immunité λ . Rien ne change donc pour ces trois sorties comparé à l'analyse précédente. Cependant, le pic infectieux n'est plus sensible qu'au temps de récupération γ , et moins au taux de transmission par contact direct β_1 comme avant.

Nous pouvons donc comparer les trois analyses de sensibilité réalisée sur notre modèle modifié : OAT (figure 15, Morris figure 16 et FAST (figure 17 et figure ??), avec les analyses précédentes. On observe alors que l'incidence, la prévalence de la première année et le taux de morbidité, sont majoritairement sensibles aux mêmes paramètres qu'avant. Nous pouvons seulement constater que le pic infectieux, n'est plus autant sensible au taux de contamination par contact direct β_1 , comparé à avant.

c. Discutez les résultats obtenus, incluant une comparaison avec le modèle initial

La modification du modèle proposée dans cette étude, n'a au final apportée que très peu d'effets sur les analyses de sensibilité réalisée à l'issue. Parmi les 8 paramètres épidémiologiques ajoutés, aucun d'entre eux n'est ressorti particulièrement influant au sein des trois méthodes d'analyse.

Une principale raison pouvant expliquer cette observation, est que l'ajout complexe d'un compartiment de pathogènes libres, ne va faire que renforcer, à échelle plus ou moins légère, l'intensité du passage des individus sensibles à infectieux. Hors, comme nous avons pu l'évoquer ultérieurement, le comportement de notre modèle est principalement dépendant de la stagnation des individus dans le compartiment résistant, et de l'immunisation majoritaire de la population après la vague d'infection. L'ajout d'autres paramètres en lien avec les fonctions d'infection, n'ont donc rien apporté.

Nous avons cependant pu noter avant, que le pic infectieux était sensible au taux d'infection par contact direct β_1 , ce qui n'est plus autant le cas dans notre modèle modifié. Ceci est un constat intéressant, puisqu'en effet nous avons expliqué que le pic infectieux, était majoritairement influencé par les paramètres qui renforçaient la vague d'infection au départ du modèle. Nous pourrions donc nous attendre à ce que l'ajout d'un moyen de contamination supplémentaire ait une influence sur cette sortie. C'est ce qu'il se passe effectivement, puisque l'on peut voir que notre nouveau taux de contamination β_2 , est celui qui ressort souvent le plus parmi tous nos nouveaux paramètres. Il est donc possible que, la sensibilité liée au taux de contamination β_1 , soit "diluée" ou "répartie" avec celle du taux de contamination β_2 . Ceci est d'autant plus pertinent que bien que nous ayons ajouté une fonction de transmission de type densité-dépendance, alors que la première fonction de

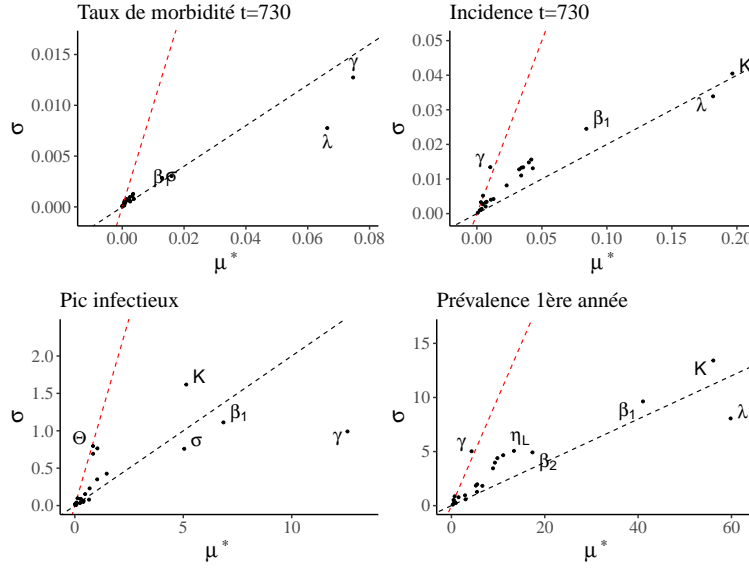


Figure 16 – Moyennes et variances des contributions des paramètres dans la méthode Morris pour le modèle modifié. La ligne rouge correspond au ratio $\sigma/\mu^* = 1$ et la ligne noire au ratio $\sigma/\mu^* = 0.2$

transmission était de type fréquence-dépendance, nous avons conservé une échelle de grandeur adéquate entre ces deux types par rapport à la taille totale générale de la population.

Pour finir, nous aurions pu envisager d'autres alternatives afin d'induire des modifications sur les analyses de sensibilité de notre modèle. Par exemple, nous aurions pu ajouter la dynamique démographique d'un second variant de Ranavirus, ou d'une autre espèce sensible à ce dernier et pouvant transmettre le virus à notre population initiale, comme une espèce de poisson ou d'autre amphibien. Enfin, il est important de noter que les analyses de sensibilité, quelque soit la méthode utilisée, sont entièrement dépendante de la sortie choisie. Il est donc assuré que les paramètres de la modification proposés sur le modèle, bien qu'ils n'aient pas eu d'effets sur nos quatre sorties actuelles, ait une influence importante sur d'autres types de sorties, notamment comme par exemple sur le nombre de reproduction de base, R_0 .

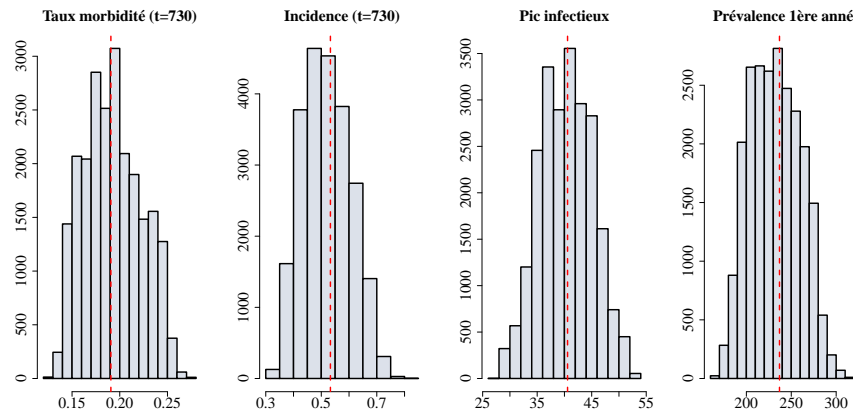


Figure 17 – Distribution des variations de sortie des 4 indicateurs épistémologiques de l'échantillonnage FAST. La ligne rouge représente la valeur initiale de la sortie.

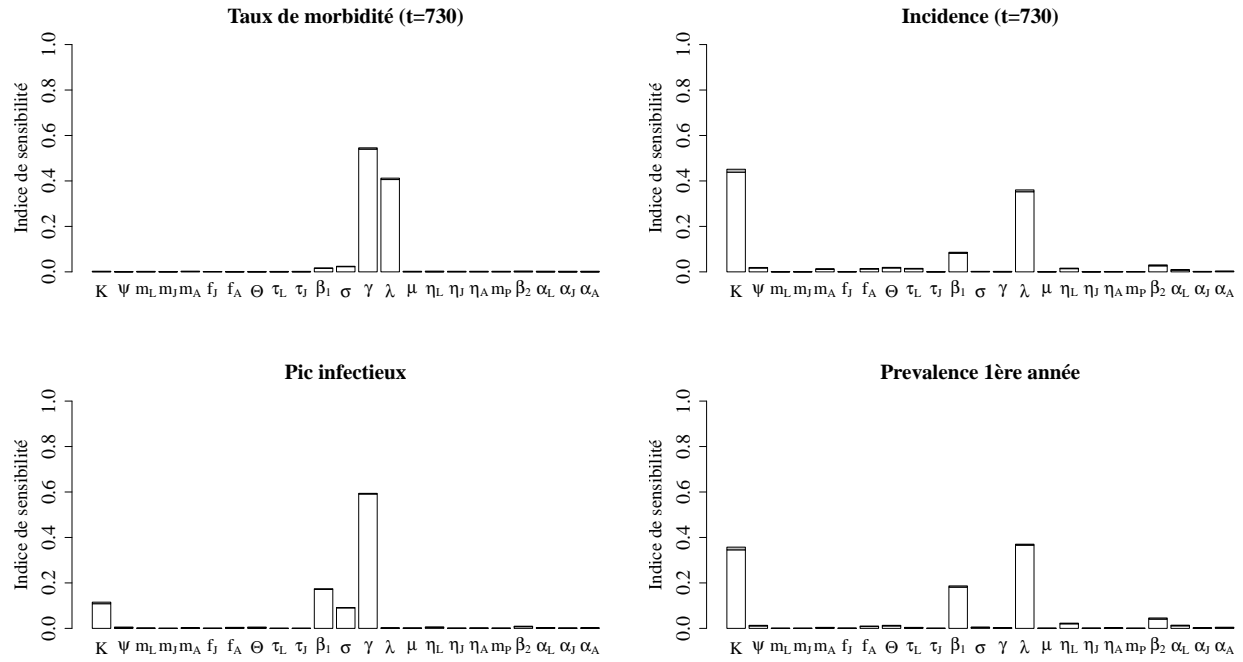


Figure 18 – Indice principal et indice total de chaque paramètre pour les indices interprétables pour 1000 itérations FAST. Les plots grisés correspondent aux sorties non-interprétables. Blanc : effet principal, gris : effet d'interaction, la somme des deux représentant la contribution totale à la variance.

Références

- Bauer, L. and Hamby, D. (1991). Relative sensitivities of existing and novel model parameters in atmospheric tritium dose estimates. *Radiation protection dosimetry*, 37(4) :253–260.
- Beaunée, G., Gilot-Fromont, E., Garel, M., and Ezanno, P. (2015). A novel epidemiological model to better understand and predict the observed seasonal spread of pestivirus in pyrenean chamois populations. *Veterinary Research*, 46(1) :1–11.
- Charron, M. V., Seegers, H., Langlais, M., and Ezanno, P. (2011). Seasonal spread and control of bluetongue in cattle. *Journal of theoretical biology*, 291 :1–9.
- Corrado, V. and Mechri, H. E. (2009). Uncertainty and sensitivity analysis for building energy rating. *Journal of building physics*, 33(2) :125–156.
- Cukier, R., Fortuin, C., Shuler, K. E., Petschek, A., and Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i theory. *The Journal of chemical physics*, 59(8) :3873–3878.
- Faivre, R., Makowski, D., Mahévas, S., and Iooss, B. (2013). Analyse de sensibilité et exploration de modèles : application aux sciences de la nature et de l’environnement. *Analyse de sensibilité et exploration de modèles*, pages 1–352.
- Granoff, A., Came, P. E., and Breeze, D. C. (1966). Viruses and renal carcinoma of rana pipiens : I. the isolation and properties of virus from normal and tumor tissue. *Virology*, 29(1) :133–148.
- Gray, M. J., Miller, D. L., and Hoverman, J. T. (2009). Ecology and pathology of amphibian ranaviruses. *Diseases of aquatic organisms*, 87(3) :243–266.
- Lupo, C., Travers, M.-A., Tourbiez, D., Barthélémy, C. F., Beaunée, G., and Ezanno, P. (2019). Modeling the transmission of vibrio aestuarianus in pacific oysters using experimental infection data. *Frontiers in Veterinary Science*, 6 :142.
- Menberg, K., Heo, Y., and Choudhary, R. (2016). Sensitivity analysis methods for building energy models : Comparing computational costs and extractable information. *Energy and Buildings*, 133 :433–445.
- Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2) :161–174.
- Mwalili, S., Kimathi, M., Ojiambo, V., Gathungu, D., and Mbogo, R. (2020). Seir model for covid-19 dynamics incorporating the environment and social distancing. *BMC Research Notes*, 13(1) :1–5.
- Nazir, J., Spengler, M., and Marschang, R. (2012). Environmental persistence of amphibian and reptilian ranaviruses. *Diseases of Aquatic Organisms*, 98(3) :177–184.
- Pujol, G., Iooss, B., Iooss, M. B., and DiceDesign, S. (2015). Package ‘sensitivity’.
- R Core Team, R. et al. (2022). R : A language and environment for statistical computing.
- Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., and Wu, Q. (2019). Why so many published sensitivity analyses are false : A systematic review of sensitivity analysis practices. *Environmental modelling & software*, 114 :29–39.
- Saltelli, A., Tarantola, S., and Chan, K.-S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1) :39–56.