

Statistiques et modélisation de maladies infectieuses

Clara Champagne

Avril 2022

clara.champagne@swisstph.ch

Avec le covid-19, la modélisation des maladies infectieuses est entrée dans notre quotidien ...

Orange F 4G 3 % 11:32

Chiffres clés

Mis à jour : Avant-hier à 14:00

45422

Nouveaux cas

Ce nombre correspond au nombre de nouveaux cas confirmés par un test positif, sur les dernières 24 heures.

Mis à jour : Il y a 6 jours à 14:00

1.4

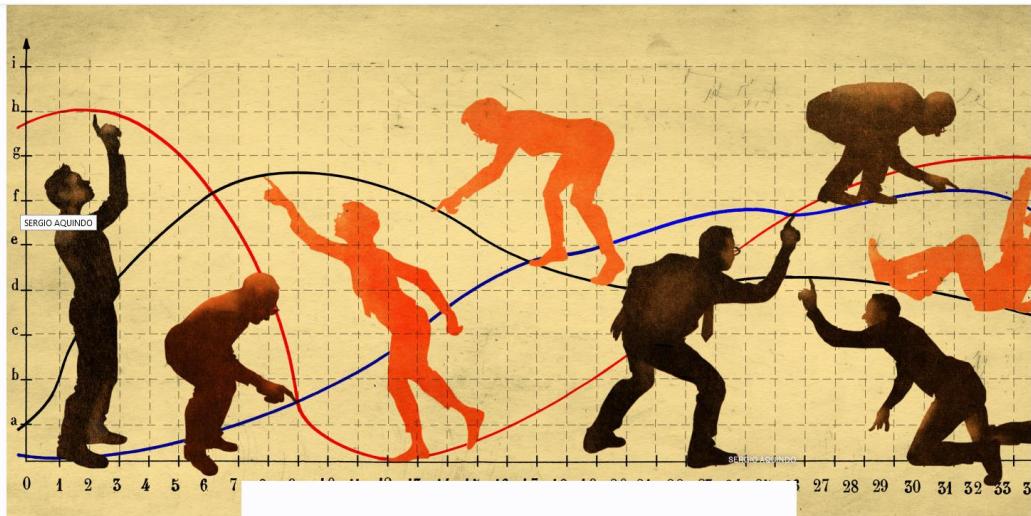
R effectif

Le nombre de reproduction du virus : c'est le nombre moyen de personnes qu'un malade peut contaminer. Actualisé régulièrement, cet indicateur mesure l'intensité de la circulation du virus. Si le R effectif est supérieur à 1, l'épidémie se développe ; s'il est inférieur à 1, l'épidémie régresse.

Mis à jour : Il y a 4 jours à 14:00

340.3

Taux d'incidence



Covid-19 : ces modélisateurs qui anticipent la pandémie

Le Monde, 5 janvier 2021

Support the Guardian
Available for everyone, funded by readers
[Contribute](#) [Subscribe](#)

The Guardian view Columnists Cartoons Opinion videos Letters

News **Opinion** **Sport** **Culture** **Lifestyle** More ▾

Paradigm shift: how the pandemic changed science Coronavirus

This article is more than 1 year old

How modelling Covid has changed the way we think about epidemics Adam Kucharski

Most viewed

Live Russia-Ukraine war latest: Moscow says it will drastically reduce military activity around Kyiv and Chernihiv - live

Russia compares sanctions

The Guardian, 4 janvier 2021

Outline

- Présentation du domaine
- Un exemple d'application: prioritisation des distributions de moustiquaires en Haïti
- Example d'application des statistiques: estimation par MCMC

Outline

- **Présentation du domaine**
- Un exemple d'application: prioritisation des distributions de moustiquaires en Haïti
- Example d'application des statistiques: estimation par MCMC

Un peu d'histoire

Daniel Bernoulli et la variole, 1760



Ronald Ross et le paludisme, 1900's

PLOS PATHOGENS

BROWSE PUBLISH ABOUT

OPEN ACCESS

REVIEW

Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens

David L. Smith, Katherine E. Battle, Simon I. Hay, Christopher M. Barker, Thomas W. Scott, F. Ellis McKenzie

Published: April 5, 2012 • <https://doi.org/10.1371/journal.ppat.1002588>

Article	Authors	Metrics	Comments	Media Coverage
---------	---------	---------	----------	----------------

| Abstract Abstract

MÉMOIRES
DE
MATHÉMATIQUE
ET
DE PHYSIQUE,
TIRÉS DES REGISTRES
de l'Académie Royale des Sciences;
De l'Année M. DCCLX.

ESSAI D'UNE NOUVELLE ANALYSE
De la mortalité causée par la petite Vérole, & des
avantages de l'Inoculation pour la prévenir.

Par M. DANIEL BERNOULLI

INTRODUCTION APOLOGÉTIQUE*.

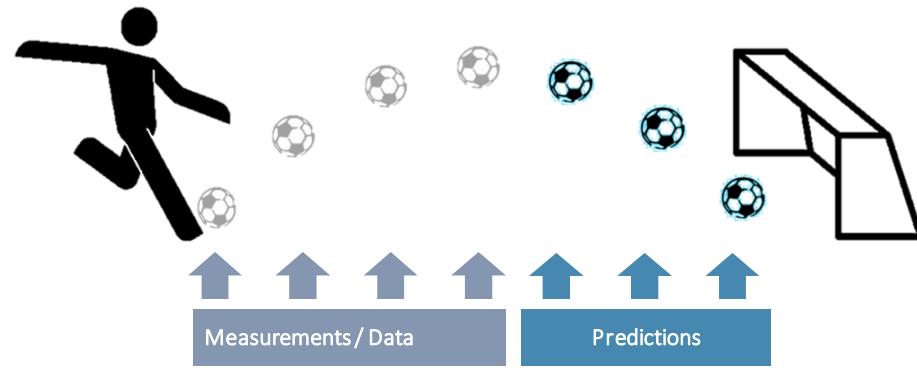
Ceux qui ont tenté tout l'avantage de l'Inoculation, ont
imaginé différentes façons de représenter cet avantage,
qui, quoique relevant au même, ne suffisent pas de faire une
* Cet Appendice n'a été fait que long-temps après le Mémoire,
jusqu'au 16 Avril 1765.

Mais, 1766.

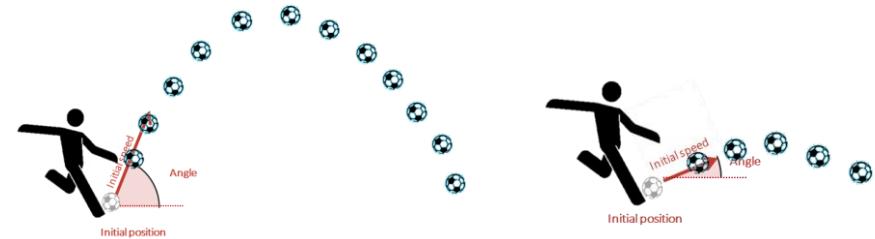
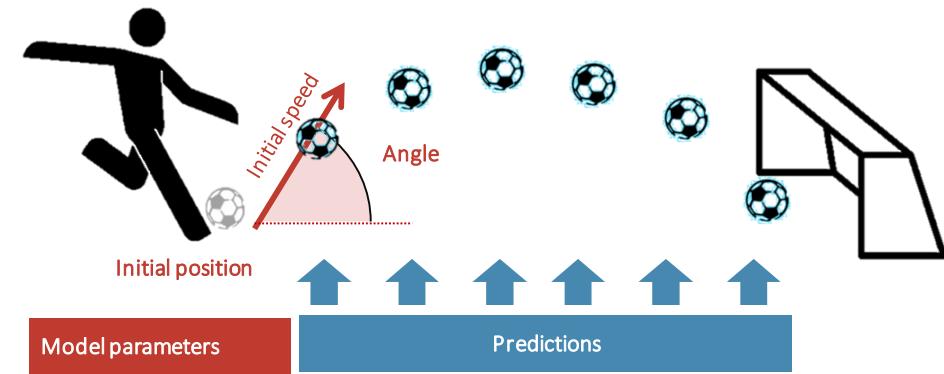
A

Différents types de modèles

Modèles statistiques



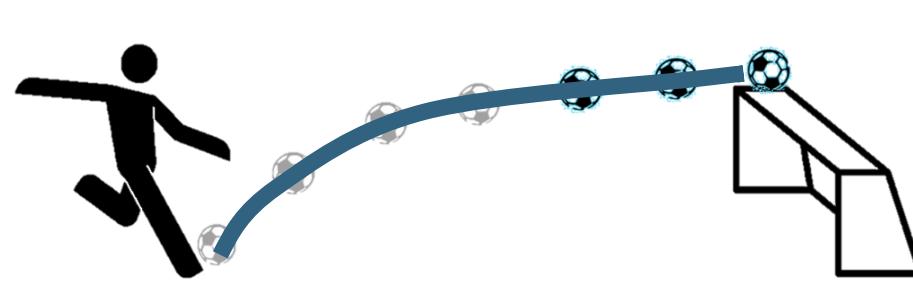
Modèles mathématiques/dynamiques



12

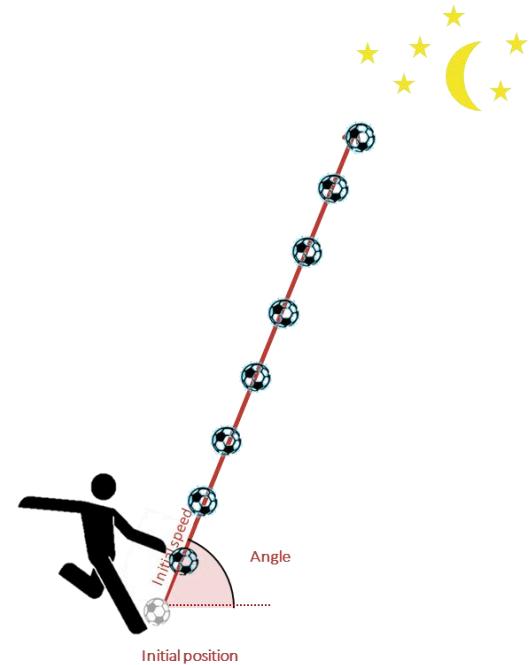
Différents types de modèles ... mais qui sont liés !

Modèles statistiques



Les modèles statistiques ont parfois besoin de structure, en plus des données elles-mêmes

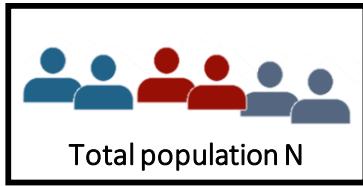
Modèles mathématiques/dynamiques



Les modèles mathématiques ont besoin d'être confrontés aux données pour être réalistes

→ besoin de statistiques !

Construire un modèle simple: le modèle SIR



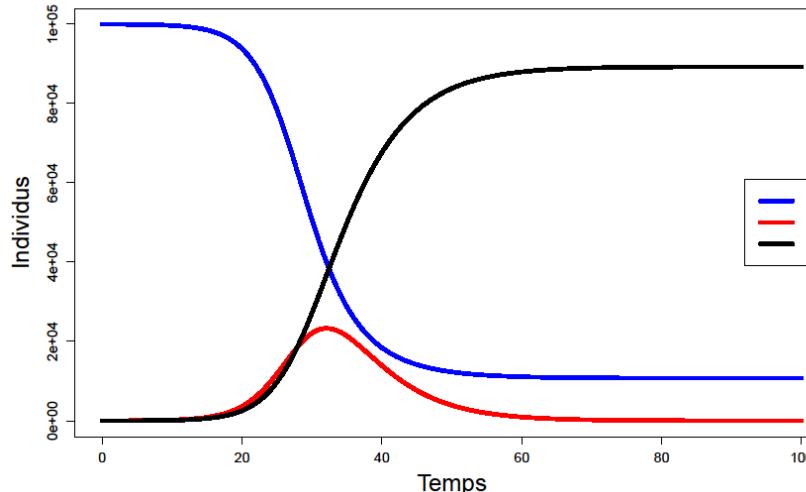
$$\frac{dS}{dt} = -b * \frac{I * S}{N}$$

$$\frac{dI}{dt} = b * \frac{I * S}{N} - r * I$$

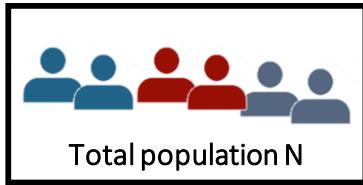
$$\frac{dR}{dt} = r * I$$

$$S + I + R = N$$

$\beta = 0.5/\text{jour}$, $\gamma = 0.2/\text{jour}$



Construire un modèle simple: le modèle SIR



$$\frac{dS}{dt} = -b * \frac{I * S}{N}$$

$$\frac{dI}{dt} = b * \frac{I * S}{N} - r * I$$

$$\frac{dR}{dt} = r * I$$

$$S + I + R = N$$



> 0 : l'épidémie augmente



< 0 : l'épidémie recule

$$bI \frac{S}{N} > rI$$

$$\left(\frac{S}{N} \right) \left(\frac{r}{b} \right) = \frac{1}{R_0}$$

au début de l'épidémie $\cong 1$



$$\begin{aligned} &\rightarrow 1 > 1/R_0 \\ &\rightarrow R_0 > 1 \end{aligned}$$

$$\rightarrow R_0 < 1$$

On peut vouloir vacciner une proportion $x=1-S/N$ de la population pour empêcher la maladie de se transmettre:

$$\frac{S}{N} < \frac{r}{b}$$

$$1 - x < \frac{r}{b}$$

$$x > 1 - 1/R_0$$

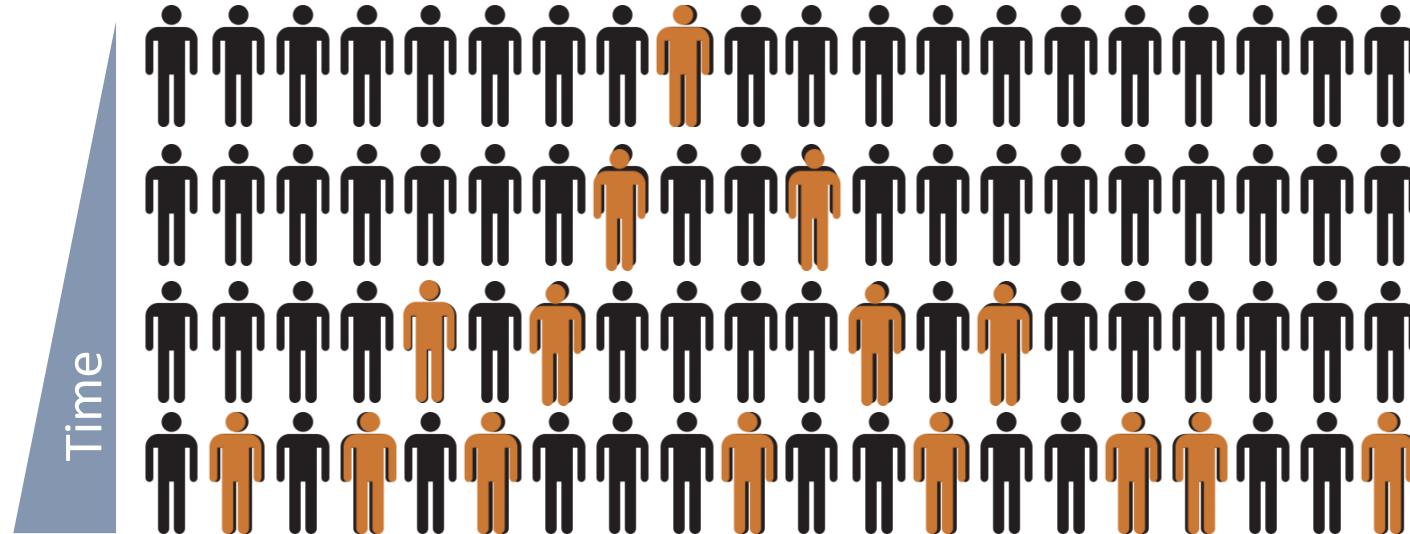
Le nombre de reproduction de base (R_0) représente le potentiel intrinsèque de transmission

R_0 : en l'absence d'interventions

Nombre d'individus infectés lorsqu'un individu infecté est introduit dans une population entièrement susceptible

$R_0 > 1$: l'épidémie se propage

Exemple: $R_0=2$



Courtesy of Emilie Pothin

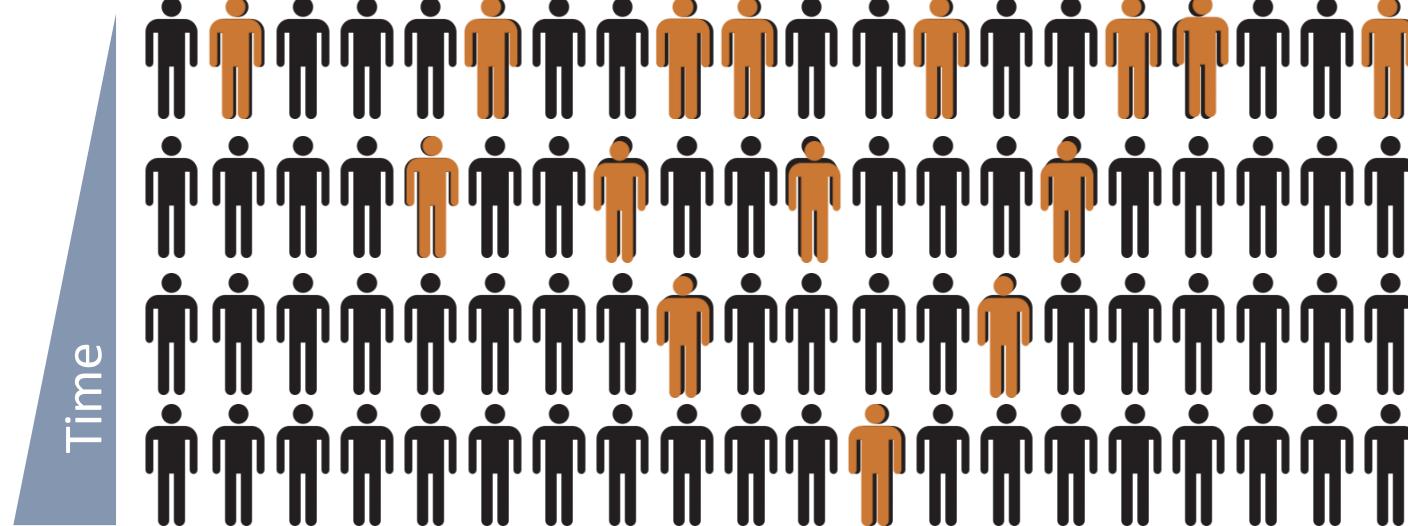
Le nombre de reproduction avec contrôle(R_c) représente le potentiel de transmission intrinsèque en présence d'interventions de contrôle

R_c : en présence d'interventions

Nombre d'individus infectés lorsqu'un individu infecté est introduit dans une population entièrement susceptible mais protégée par certaines interventions.

$R_c < 1$: l'épidémie recule

Exemple: $R_c=0.5$

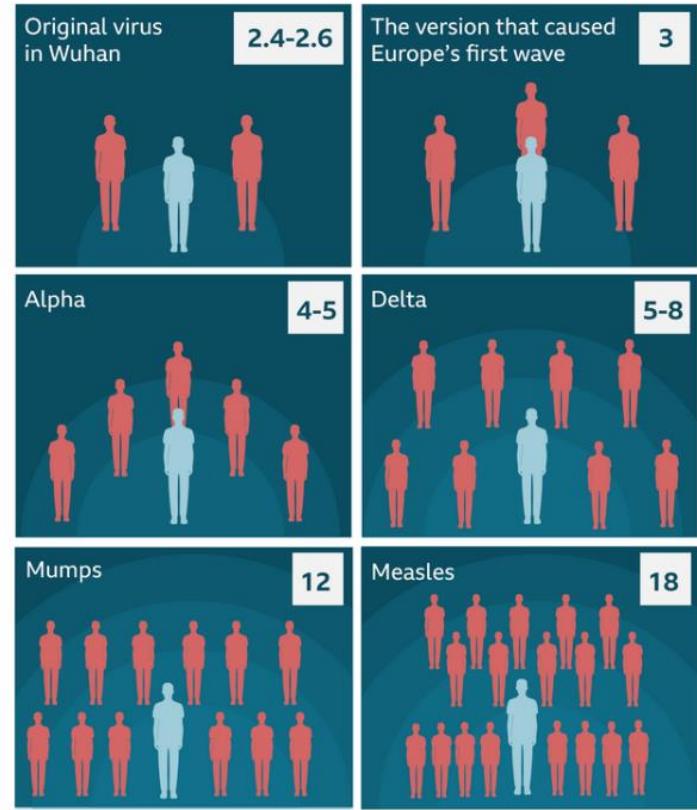


R_0 in the news



<https://twitter.com/BBCNews/status/1255912782610006016/photo/1>

How the R_0 numbers of Covid-19 variants and other diseases compare The more contagious, the higher the R_0 number



Source: Imperial College, Lancet, Australian government

BBC

<https://www.bbc.com/news/health-57431420>

12

Le nombre de reproduction peut aussi être suivi au cours du temps

R_0



Correspondent au moment (fictif) du tout
début d'une épidémie

Au cours de l'épidémie, le nombre de reproduction évolue, pour plusieurs raisons:

- Les interventions ou les paramètres peuvent changer au cours du temps
- Une partie de la population n'est plus susceptible

Le nombre de reproduction peut aussi être suivi au cours du temps

R_0



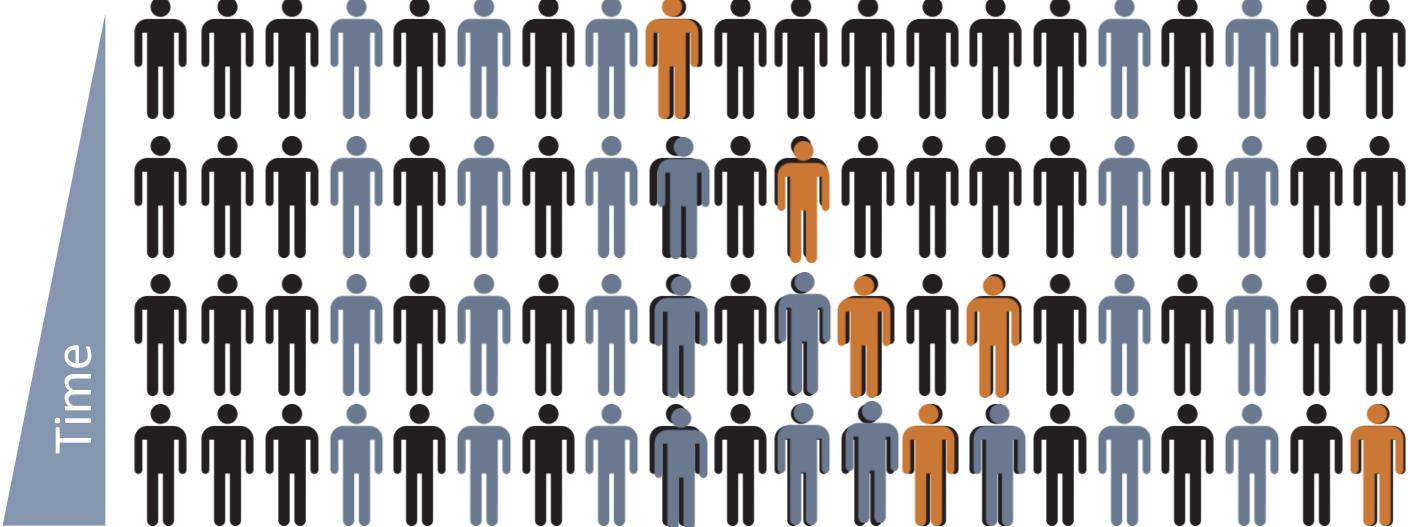
Correspondent au moment (fictif) du tout début d'une épidémie

Au cours de l'épidémie, le nombre de reproduction évolue, pour plusieurs raisons:

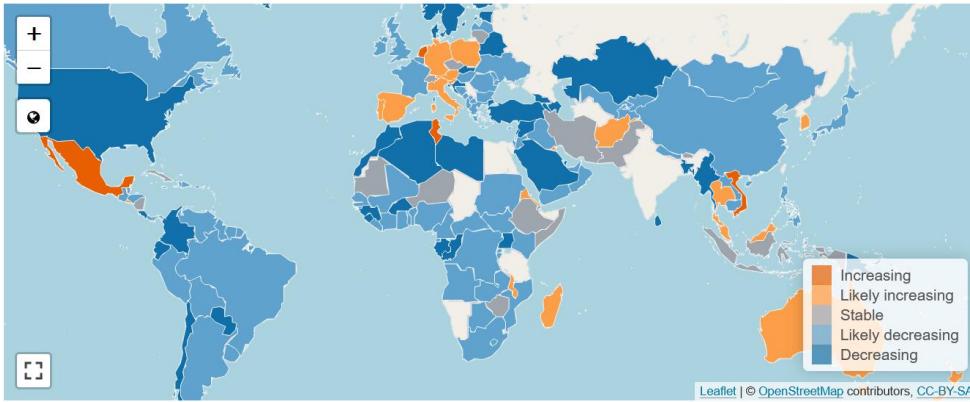
- Les interventions ou les paramètres peuvent changer au cours du temps
- Une partie de la population n'est plus susceptible

R_t correspond au nombre de reproduction de la maladie à un instant donné.

Il y a de nombreuses méthodes pour le calculer à partir de données de cas, mais nous n'allons pas en discuter aujourd'hui.

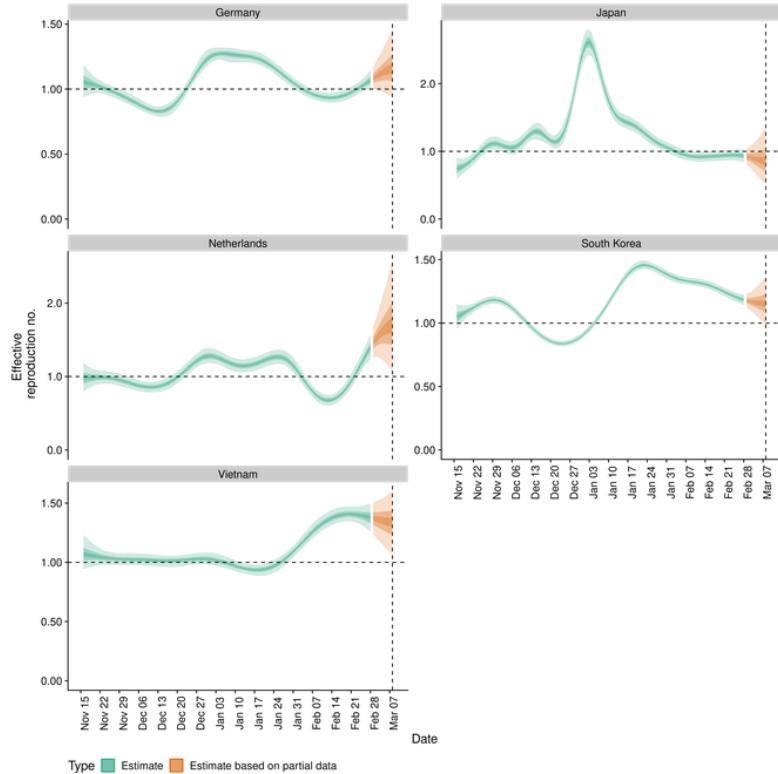


R_t et la Covid-19

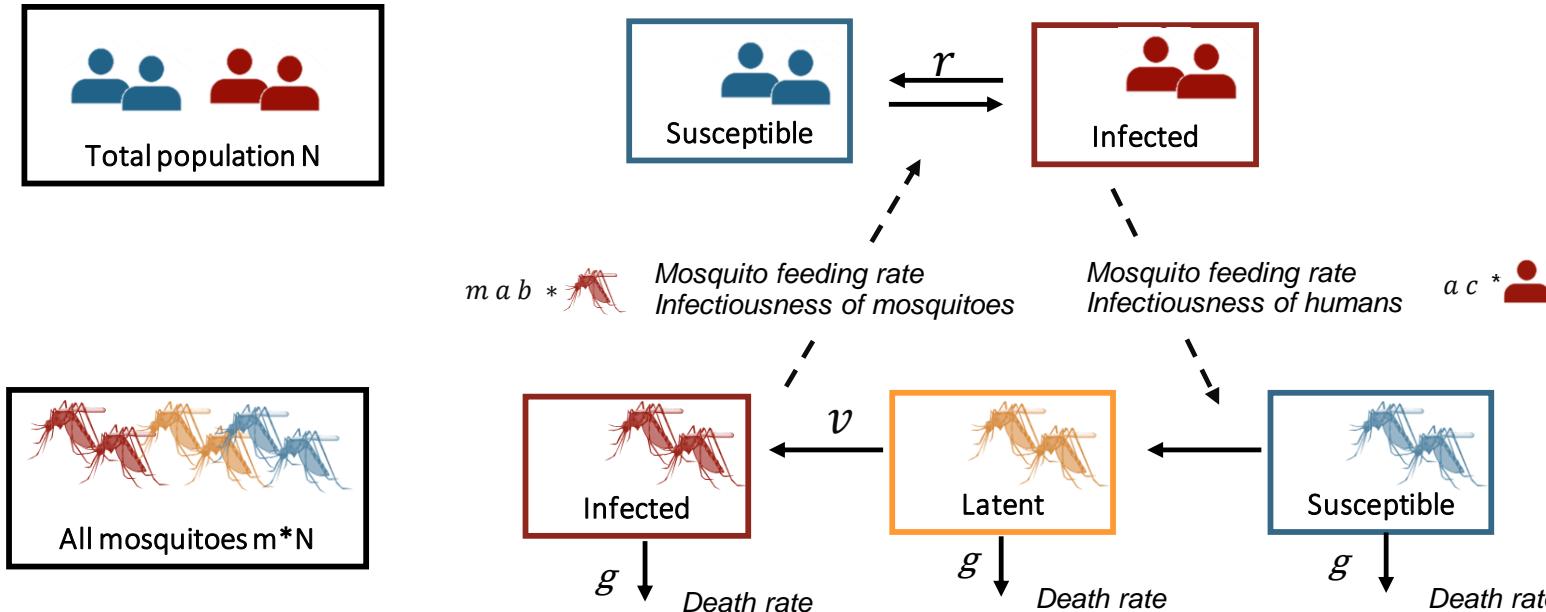


<https://epiforecasts.io/covid/posts/global/>

Reproduction numbers over time in the six regions expected to have the most new confirmed cases



Les modèles Ross - Macdonald pour le paludisme (1911, 1950's)



$$R_0 = \frac{m a^2 b c e^{-gv}}{rg}$$

Exemples d'utilisation de modèles de paludisme pour informer des décisions

1916



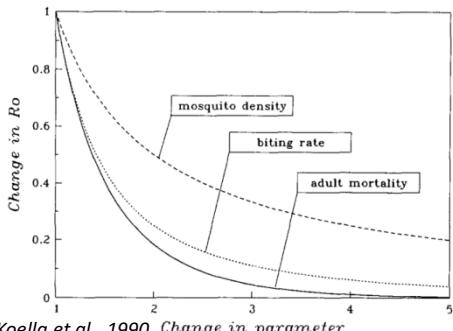
•“[...] from such reasoning alone we derive the very important practical conclusion that in order to counteract malaria anywhere we need not to bannish *Anopheles* there entirely – we need only to reduce their numbers below a certain figure”

•*Ronald Ross, 1916*



Source: National Library of Medicine

1956



•“This powerful influence of changes in the mortality rate of mosquitoes on the transmission of disease by them is the explanation of the brilliant success of residual insecticides.”

•*George McDonald, WHO bulletin, 1955*



Source: National Museum of Health and Medicine

Courtesy of Emilie Pothin

Limites des modèles présentés jusqu'à présent

- Représentation très simple de la maladie
→ *on peut ajouter d'autres compartiments (sévérité, incubation, immunité, etc.)*
- Pas d'interventions
→ *on peut ajouter des interventions comme le traitement, le contrôle vectoriel, la vaccination, etc.*
- Tous les individus sont identiques (**homogénéité**)
→ **compartimental models:** *tous les individus sont identiques dans un compartiment*



on peut ajouter d'autres compartiments pour des régions, des groupes d'âge, etc.
→ **individual-based models:** *chaque individu a ses propres caractérisques*



- Déterministe, pas d'aléa
→ *on peut ajouter de la stochasticité*

Inclure de la stochasticité

Ecriture du modèle sous forme de reactions:

Infection : $\{s,i\} \rightarrow \{s-1, i+1\}$

Guérison : $\{s,i\} \rightarrow \{s+1, i-1\}$

A chaque reaction, on associe un taux de transition:

Infection : $T(s-1, i+1 | s, i) = \frac{b i s}{N}$

Guérison : $T(s+1, i-1 | s, i) = r i$

Inclure de la stochasticité

Ecriture du modèle sous forme de reactions:

Infection : $\{s,i\} \rightarrow \{s-1, i+1\}$

Guérison : $\{s,i\} \rightarrow \{s+1, i-1\}$

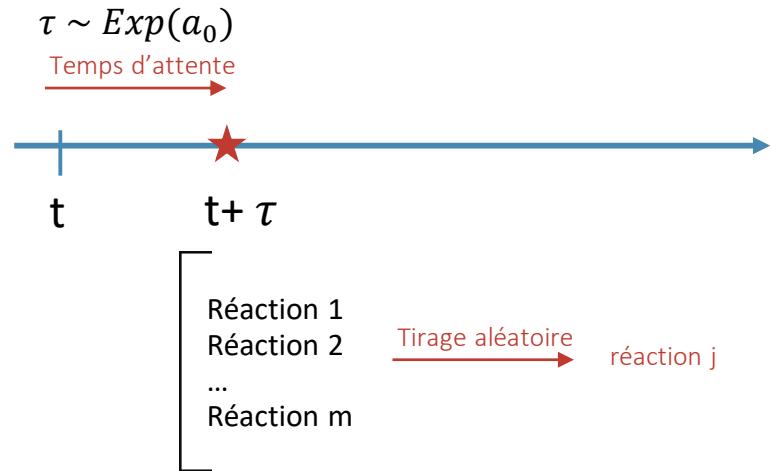
A chaque reaction, on associe un taux de transition:

$$\text{Infection : } T(s-1, i+1 | s, i) = \frac{b i s}{N}$$

$$\text{Guérison : } T(s+1, i-1 | s, i) = r i$$

Comment le simuler?

Algorithme de Gillespie



Des approximations moins coûteuses en temps de calcul existent, comme la méthode tau-leap

Inclure de la stochasticité

Ecriture du modèle sous forme de reactions:

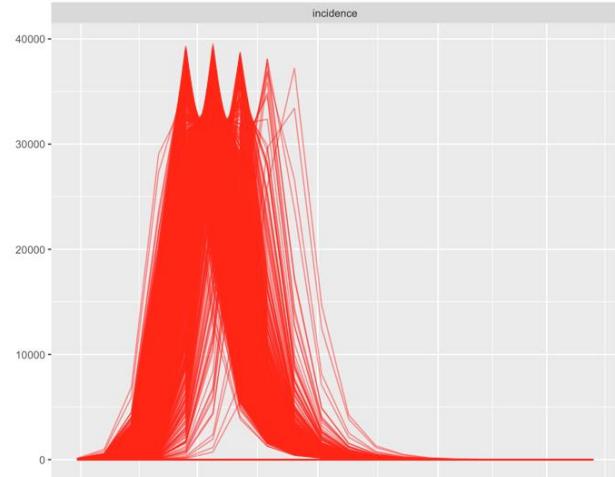
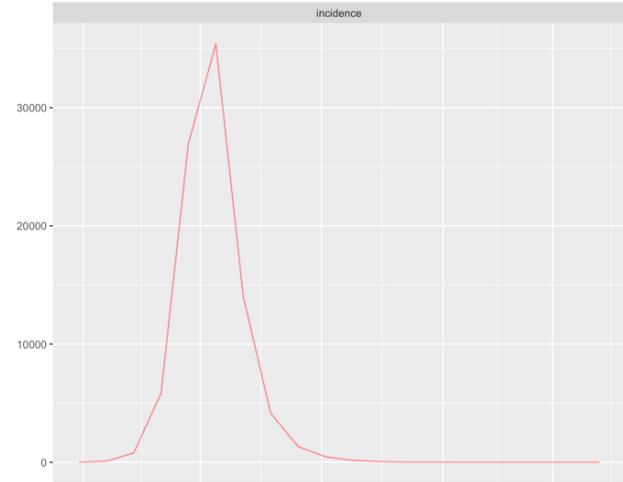
Infection : $\{s,i\} \rightarrow \{s-1, i+1\}$

Guérison : $\{s,i\} \rightarrow \{s+1, i-1\}$

A chaque reaction, on associe un taux de transition:

$$\text{Infection : } T(s-1, i+1 | s, i) = \frac{b i s}{N}$$

$$\text{Guérison : } T(s+1, i-1 | s, i) = r i$$



Limites des modèles présentés jusqu'à présent

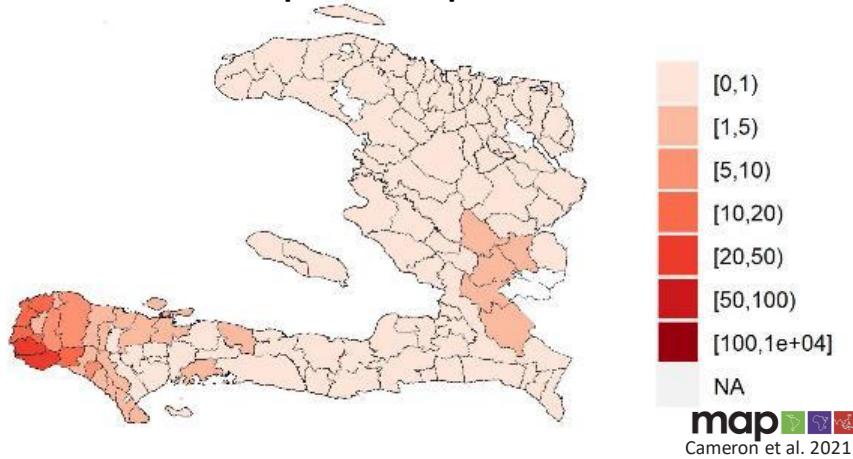
- Représentation très simple de la maladie
 - on peut ajouter d'autres compartiments (cavité, incubation, immunité, etc.)
-
- vac
- **PLUS COMPLEXE N'EST PAS TOUJOURS MIEUX!**
- - Si le modèle est trop complexe, il devient intractable et perd de son utilité en tant que modèle
 - Les modèles très complexes requièrent de très nombreux paramètres à calibrer
-

Outline

- Présentation du domaine
- Un exemple d'application: prioritisation des distributions de moustiquaires en Haïti
- Example d'application des statistiques: estimation par MCMC

Le paludisme en Haïti en 2019

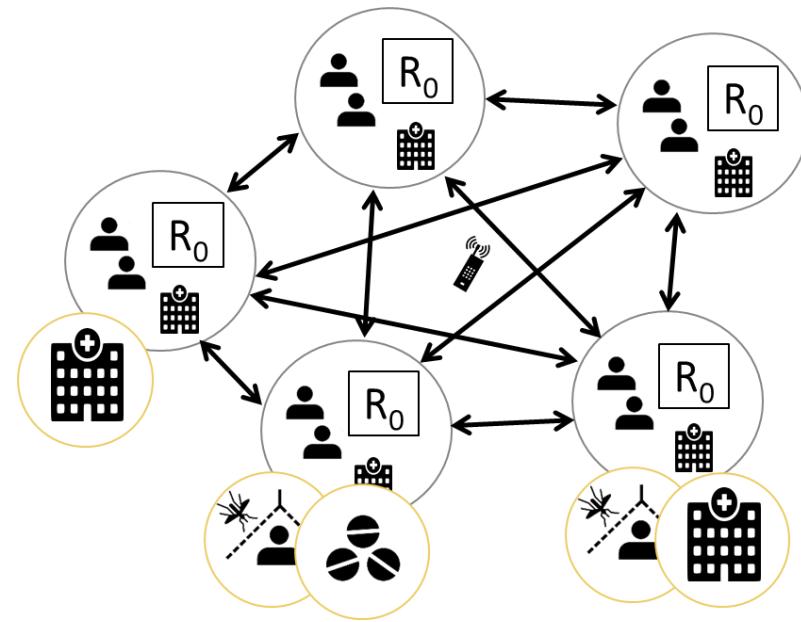
Nombre de cas de paludisme par commune en 2019

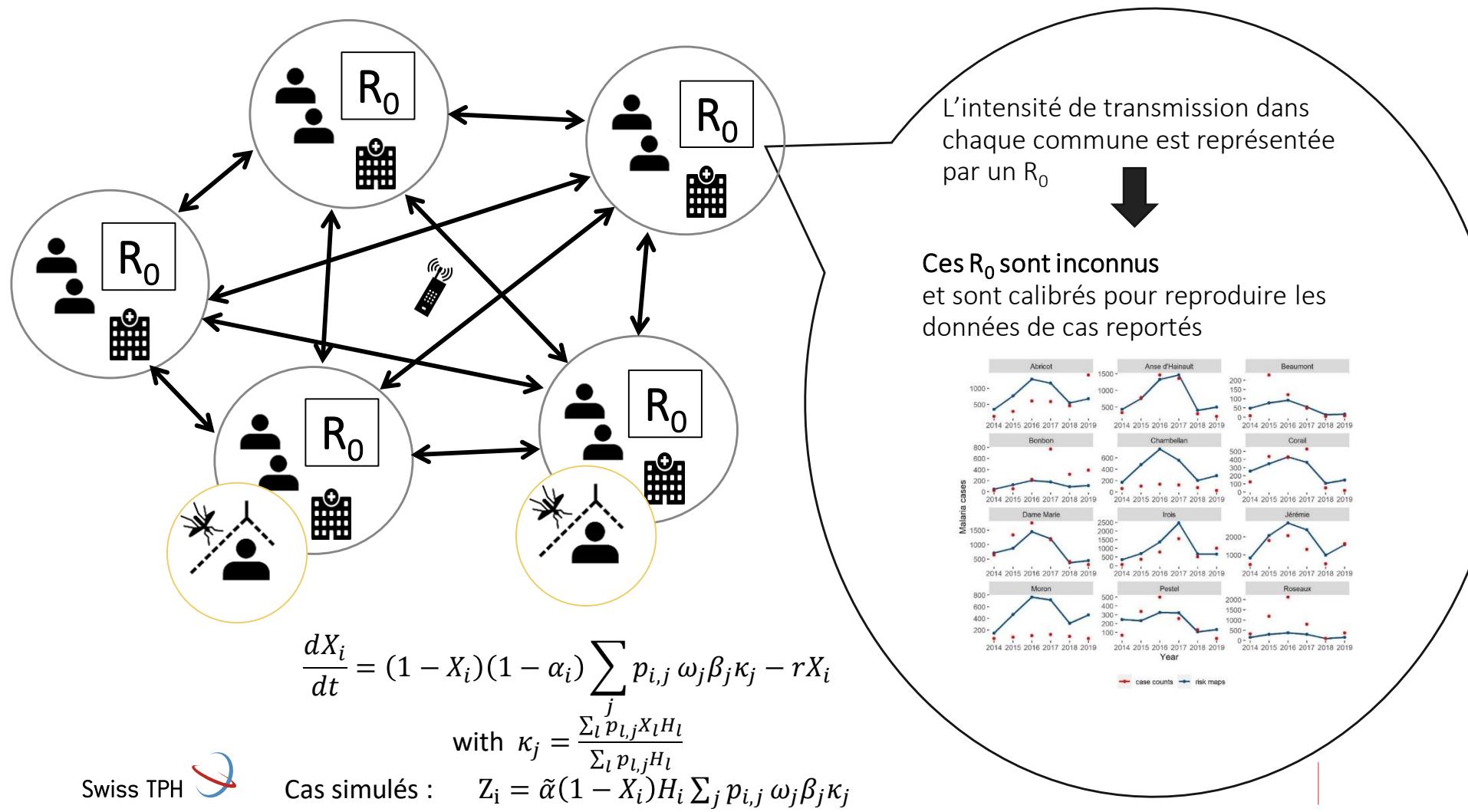


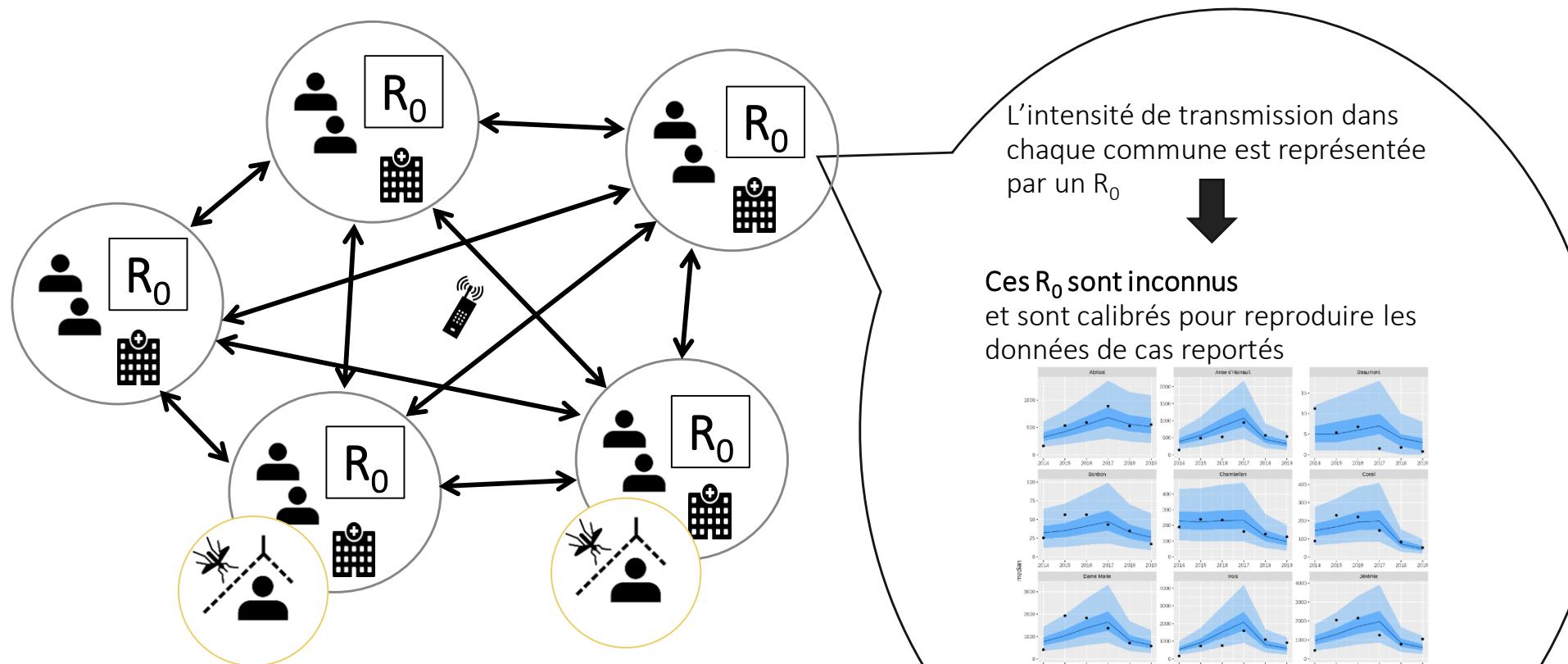
Situation hétérogène qui implique une prioritisation des interventions

→ Question: où prioriser les distributions de moustiquaires pour avoir le plus d'impact sur le paludisme?

Modèle: SIS connecté avec interventions



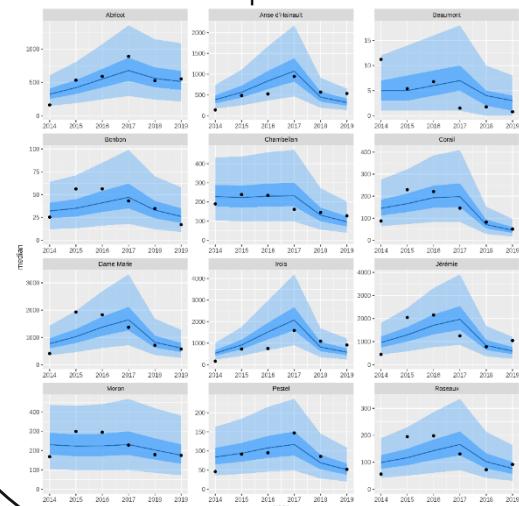




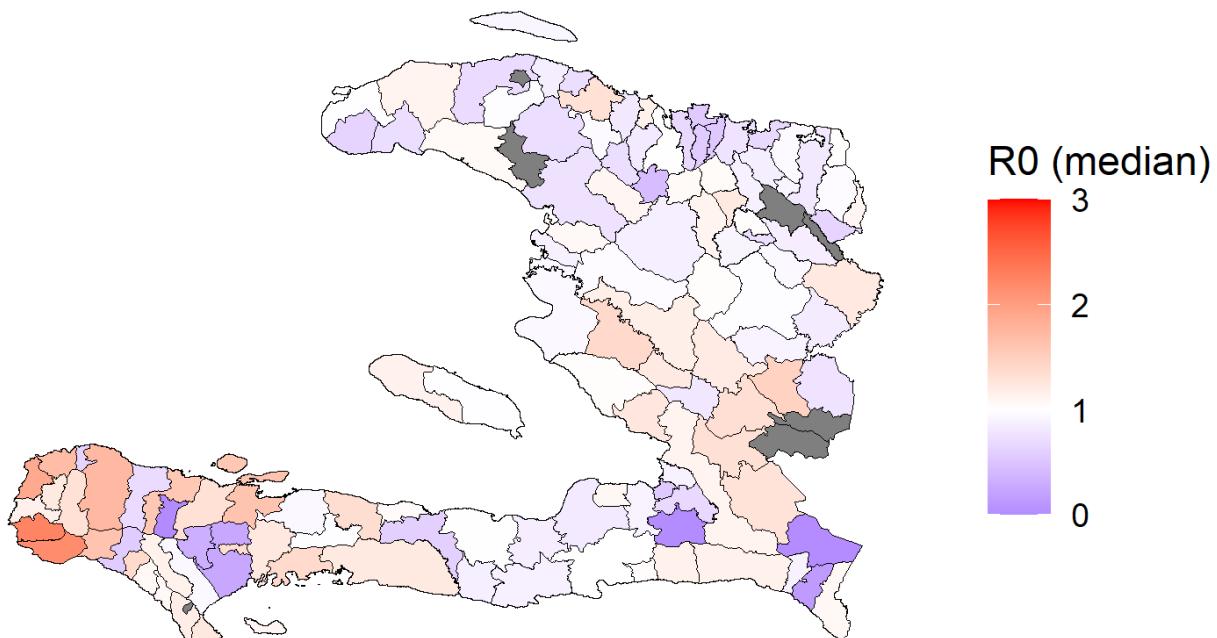
$$\frac{dX_i}{dt} = (1 - X_i)(1 - \alpha_i) \sum_j p_{i,j} \omega_j \beta_j \kappa_j - rX_i$$

$$\text{with } \kappa_j = \frac{\sum_l p_{l,j} X_l H_l}{\sum_l p_{l,j} H_l}$$

Cas simulés : $Z_i = \tilde{\alpha}(1 - X_i)H_i \sum_j p_{i,j} \omega_j \beta_j \kappa_j$



Estimations de R_0 par commune



Dans une commune:

- $R_0 > 1$: la maladie se propage dans la commune (**SOURCE**)
- $R_0 < 1$: la maladie s'éteindrait en l'absence d'importation (**SINK**)
- $R_0 = 1$: le credible interval à 50% de R_0 contient 1

Tester la validité du modèle en ne fittant que jusqu'en 2017

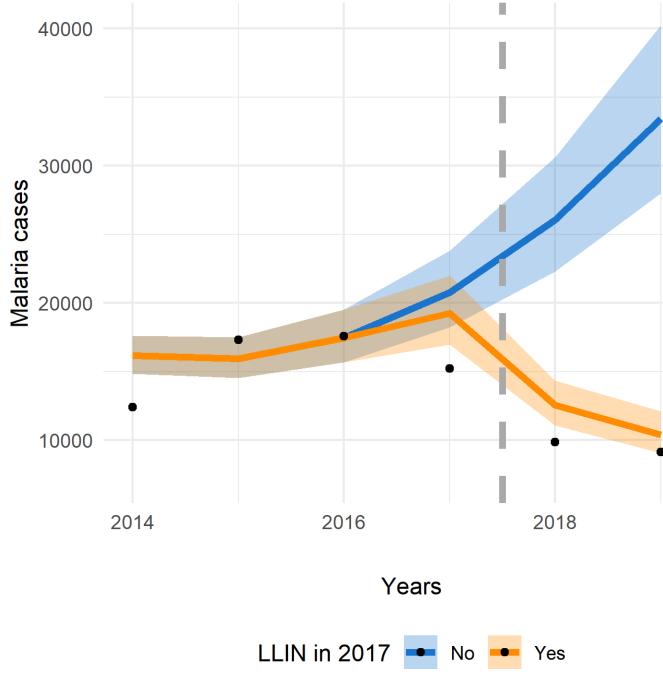
Posterior predictive simulations, summed at the national level

Le modèle est calibré au niveau des communes en utilisant les données jusqu'en 2017

On compare 2 scénarios:

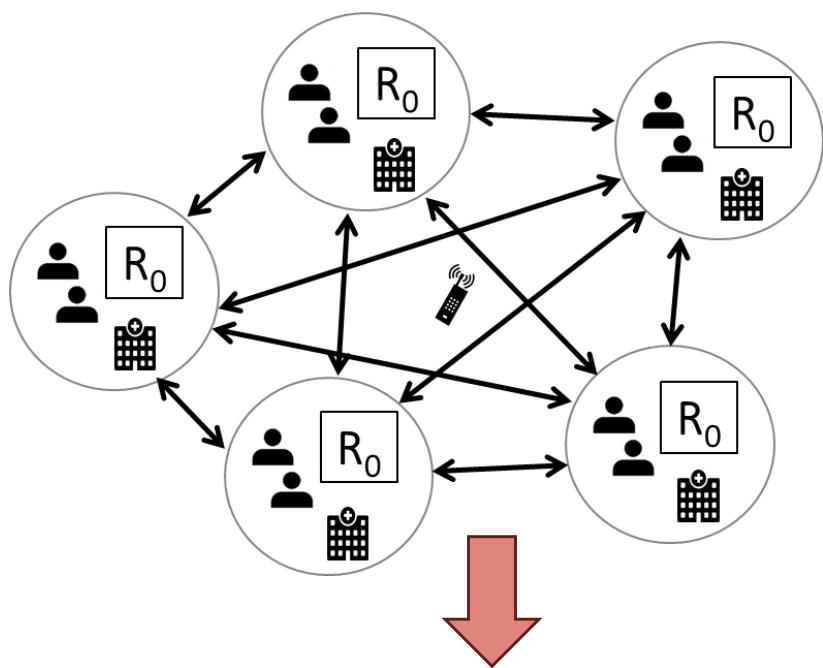
- En **ignorant** la distribution de moustiquaires en 2017
- En **incluant** la distribution de moustiquaires en 2017

Le modèle final utilise les données 2014-2019 et inclut la distribution de moustiquaires en 2017

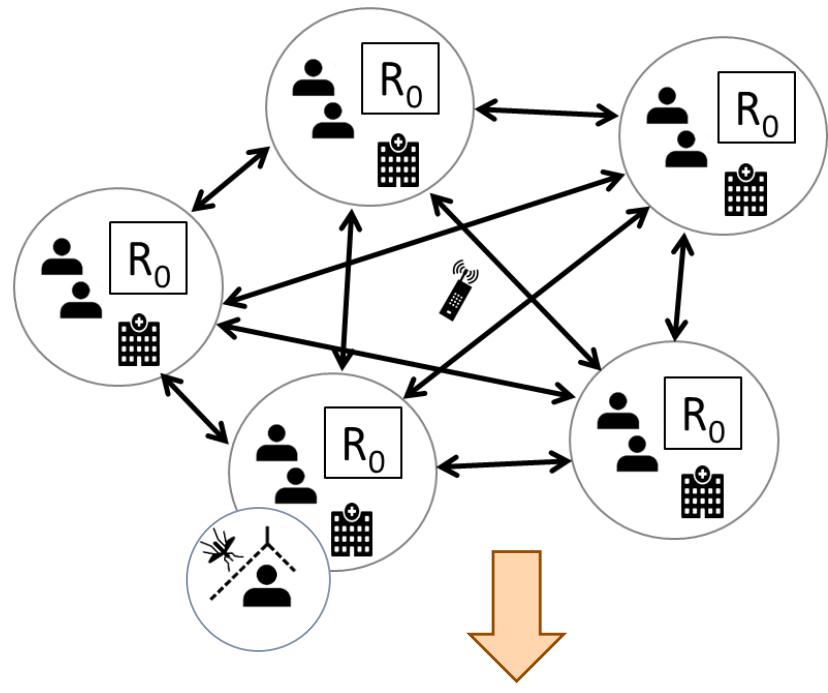


Comment utiliser le modèle pour répondre à la question posée?

Contrefactuel: pas de moustiquaires



Intervention: moustiquaires dans 1 commune

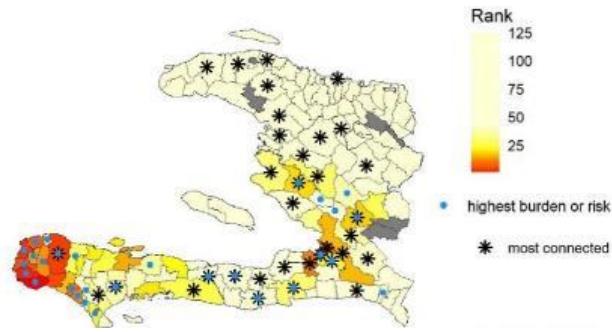


Priorisation des distributions de moustiquaires

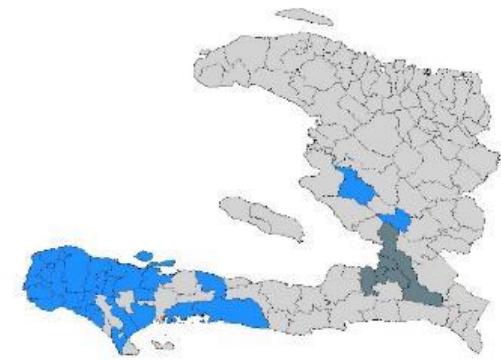
*Si on distribue des moustiquaires **dans une seule commune**, quell serait l'impact à l'échelle nationale en 2022?*



Dans quelles communes l'intervention aurait le plus d'impact au niveau national en 2022?

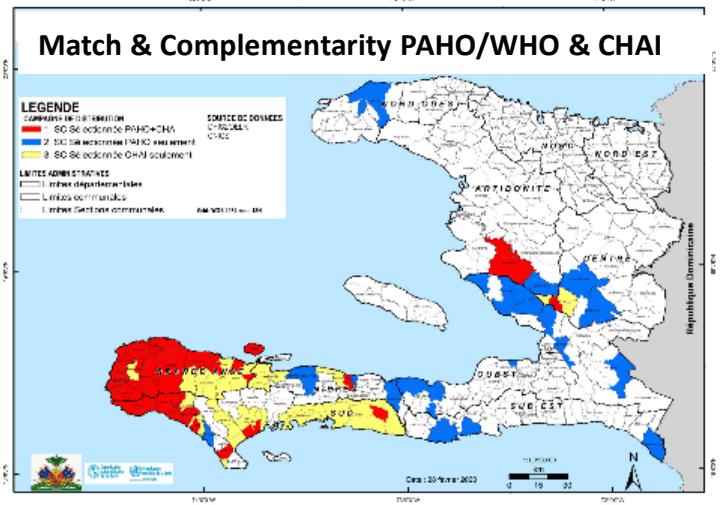


Quelles communes prioriser pour la campagne de distribution de 2020?

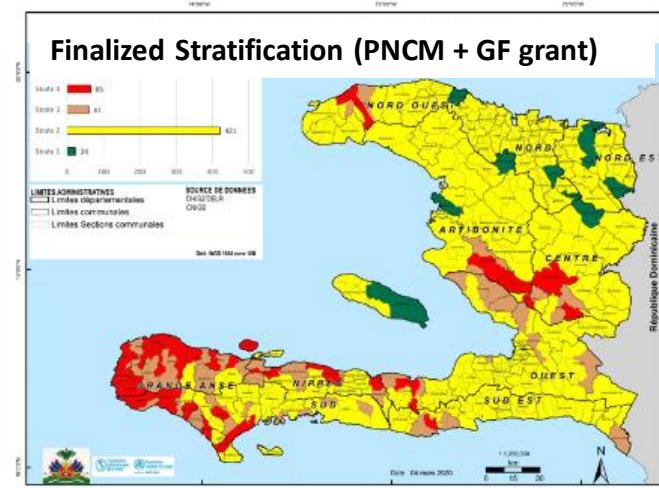


En combinant plusieurs versions du modèle, nous avons propose une liste de communes pour la distribution de moustiquaires

Utilisation des résultats dans l'application au Fonds Mondial



Simultanément, PAHO/WHO ont aussi réalisé une stratification du risque de paludisme

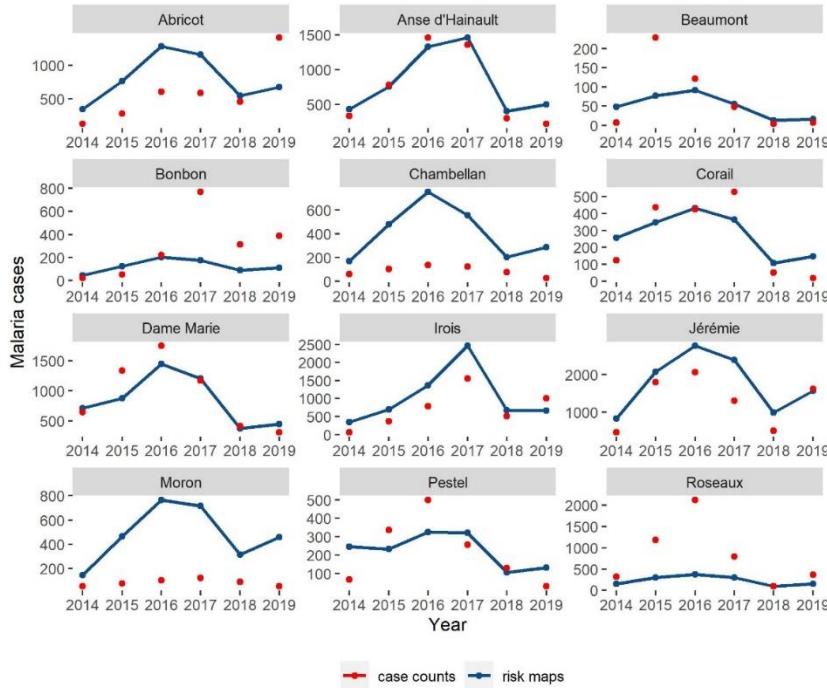


Les deux exercices de priorisation (PAHO et modèle) ont été combinés pour informer la stratégie nationale inclue dans le plan national stratégique.

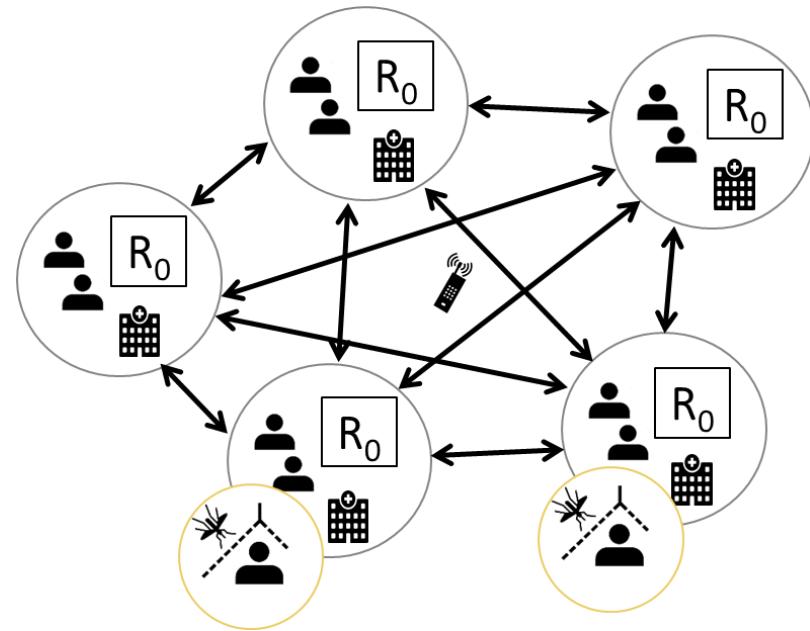
→ Les deux approches ont fourni des résultats similaires de manière générale

→ Le travail de modélisation a conduit à l'inclusion de 4 communes qui autrement auraient été exclues de la distribution de moustiquaires

Comment fait-on pour calibrer le modèle aux données?



Nombre de nouveaux cas par année
(pour plusieurs années d'observation, pour chaque commune)



Nombre d'individus par compartiment

Outline

- Présentation du domaine
- Un exemple d'application: prioritisation des distributions de moustiquaires en Haïti
- **Example d'application des statistiques: estimation par MCMC**

Faire le lien entre modèle et données: le processus d'observation

Quelques notations:

Données observées: X_{obs}

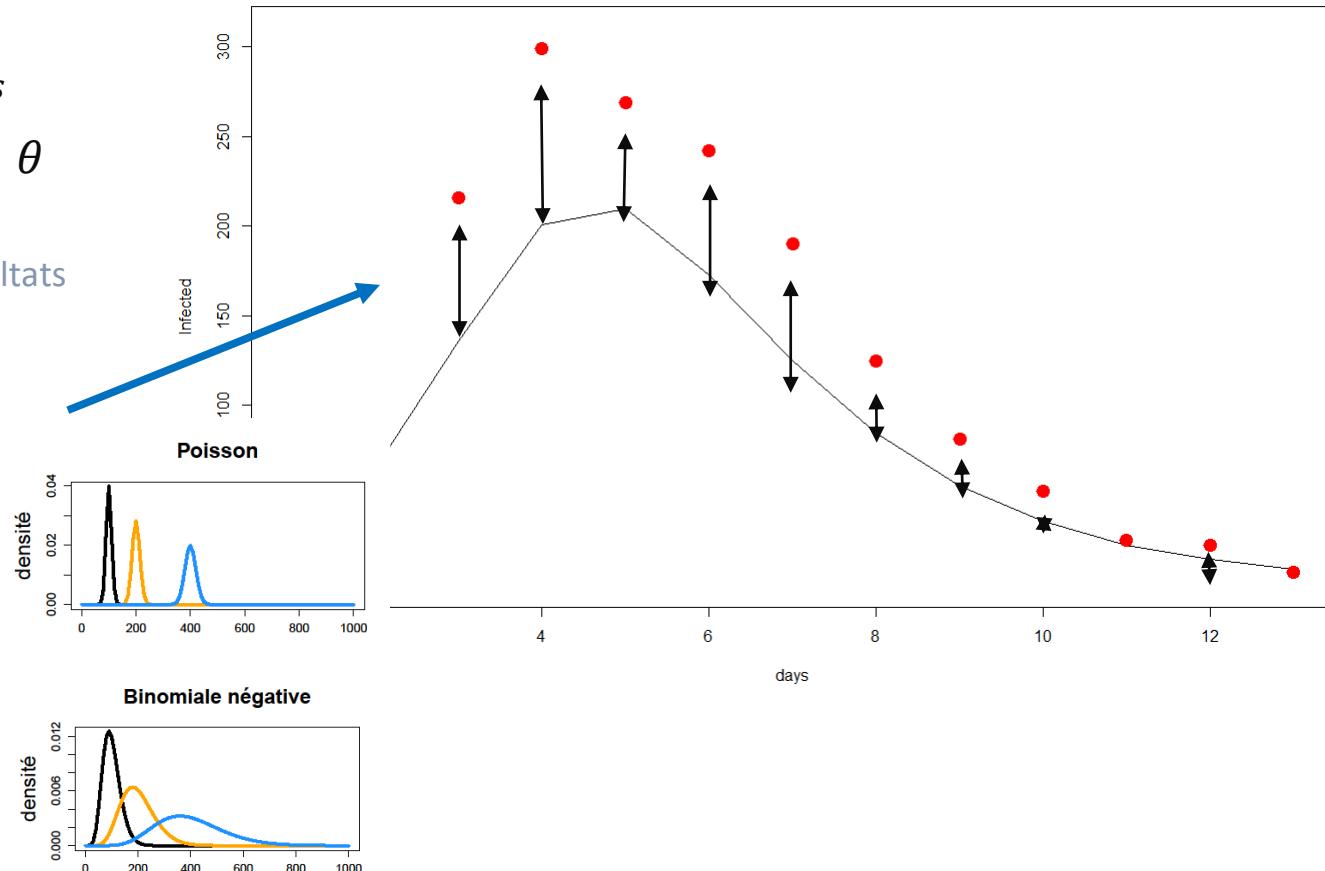
Vecteur de paramètres: θ

Métrique pour comparer les résultats
du modèle et les données

$$L_M(X_{obs}, \theta) = P_M(X_{obs}, \theta)$$

Peut être:

- Gaussienne (moindre carrés)
- Poisson
- Binomiale négative



Le modèle en entier: modèle espace état

Exemple du modèle SIR

- Equation de transition:

$$\frac{dS}{dt} = -\frac{bSI}{N}$$

$$\frac{dI}{dt} = \frac{bSI}{N} - rI \quad \longrightarrow \quad X_{sim_{1,\dots,n}}$$

$$\frac{dR}{dt} = rI$$

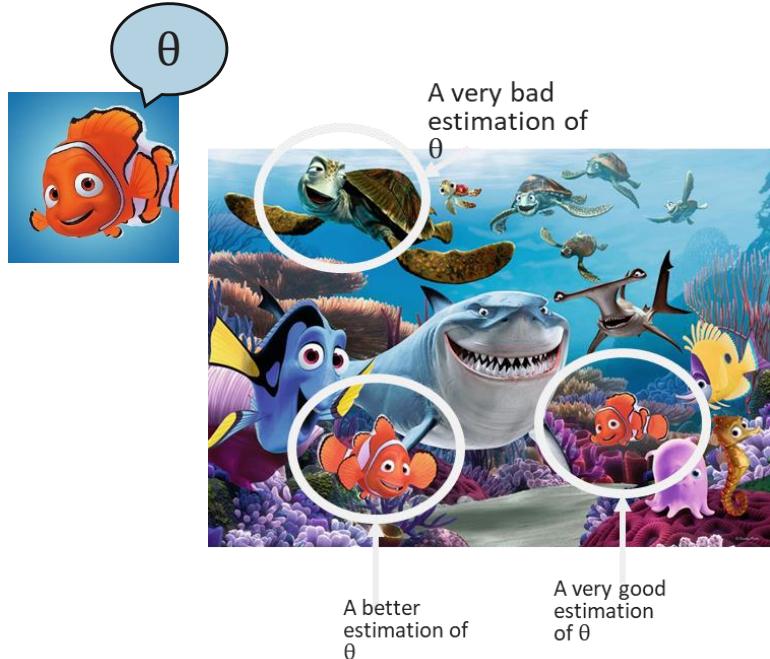
- Equation d'observation: $X_{obs_{1,\dots,n}} \sim Poisson(\rho X_{sim_{1,\dots,n}})$

Vraisemblance d'un jeu de paramètres $\theta = (b, r, \rho)$:

$$L_M(X_{obs_{1,\dots,n}}, \theta) = \prod_{t=1}^n P'_\theta(X_{obs_{1,\dots,n}} = x_{obs_{1,\dots,n}})$$

Estimation dans le contexte bayésien

Contexte fréquentiste



On cherche un estimateur de la vraie valeur (inconnue) de θ , i.e. une façon de calculer de θ à partir des données

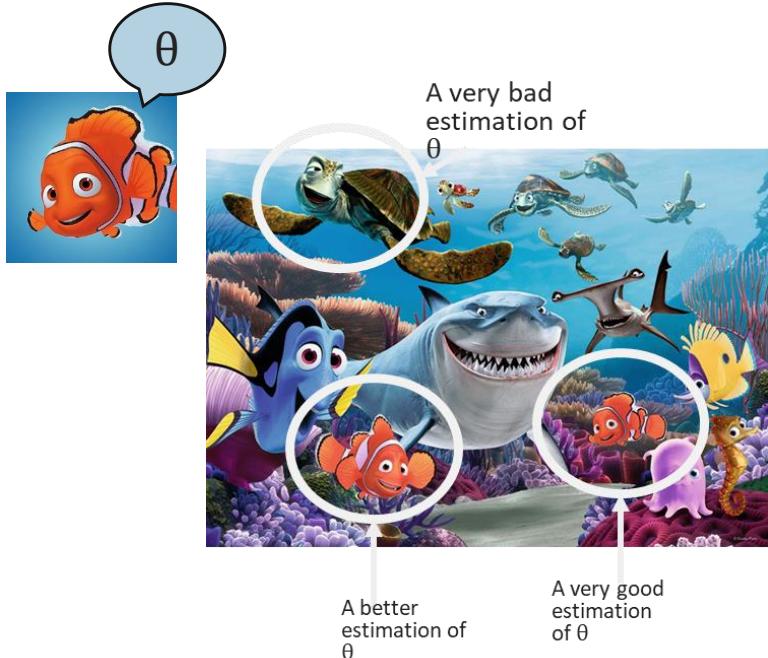
Contexte bayésien



On cherche une distribution de probabilité pour θ , sachant qu'on a observé les données, i.e. la distribution a posteriori $P(\theta|X_{obs}) \propto L_M(X_{obs}, \theta)\pi(\theta)$

Estimation dans le contexte bayésien

Contexte fréquentiste



On cherche un estimateur de la vraie valeur (inconnue) de θ , i.e. une façon de calculer de θ à partir des données

Contexte bayésien



Example of a non-informative prior

On cherche une distribution de probabilité pour θ , sachant qu'on a observé les données, i.e. la distribution a posteriori $P(\theta|X_{obs}) \propto L_M(X_{obs}, \theta)\pi(\theta)$

COMMENT LA CALCULER?

MCMC = Markov Chain Monte Carlo

- **Monte Carlo:** Répéter une expérience pour évaluer une quantité
- **Chaîne de Markov:** Collection de variables aléatoires $(X_i)_{i \leq 0}$ telles que $P(X_{i+1}|X_0, \dots, X_i) = P(X_{i+1}|X_i) = Q$

Sous certaines conditions, la chaîne admet une probabilité invariante Π telle que si $X_i \sim \Pi$ alors $X_{i+1} \sim \Pi$

Sous certaines conditions, si on fait tendre i vers l'infini, alors la distribution limite de X_i est Π

Idée du MCMC

- On veut simuler selon une loi cible f – ici la distribution a posteriori $f = P(\theta|X_{obs})$
- On génère une réalisation $(\theta_0, \theta_1, \dots)$ d'une chaîne de Markov pour laquelle $\Pi = f = P(\theta|X_{obs})$
- Quand la chaîne aura convergé, elle imitera un échantillon aléatoire tiré selon $\Pi = P(\theta|X_{obs})$

L'algorithme de Metropolis-Hastings

La loi cible est f . On a besoin d'un point d'initialisation θ_0 et d'une densité de transition $q(\cdot | \theta_i)$

1. Initialiser θ_0
2. **for** $i = 1 \dots N$
3. Echantillonner θ^* avec $q(\cdot | \theta_i)$
4. Accepter θ^* avec probabilité $1 \wedge \frac{f(\theta^*)q(\theta_i|\theta^*)}{f(\theta_i)q(\theta^*|\theta_i)}$
5. Si θ^* est accepté, $\theta_{i+1} = \theta^*$. Sinon $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \dots, \theta_N)$

Cet algorithme génère une chaîne de Markov $(\theta_0, \theta_1, \dots, \theta_N)$ dont la densité invariante est f

L'algorithme de Metropolis-Hastings

La loi cible est f . On a besoin d'un point d'initialisation θ_0 et d'une densité de transition $q(\cdot | \theta_i)$

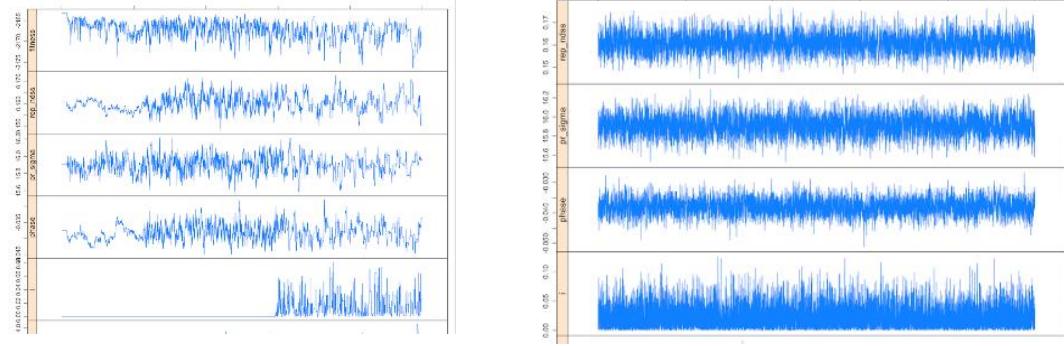
1. Initialiser θ_0
2. **for** $i = 1 \dots N$
 3. Echantillonner θ^* avec $q(\cdot | \theta_i)$
 4. Simuler le modèle épidémiologique avec θ^*
 5. Calculer le numérateur du posterior pour θ^* i.e. $L(\theta^*, X_{obs})\pi(\theta^*)$
 6. Accepter θ^* avec probabilité $1 \wedge \frac{L(\theta^*, X_{obs})\pi(\theta^*)}{L(\theta_i, X_{obs})\pi(\theta_i)} q(\theta_i | \theta^*) / q(\theta^* | \theta_i)$
 7. Si θ^* est accepté, $\theta_{i+1} = \theta^*$. Sinon $\theta_{i+1} = \theta_i$
8. **end for**
9. **return** $(\theta_0, \theta_1, \dots, \theta_N)$

L'algorithme de Metropolis-Hastings

Il est important de vérifier que la chaîne a convergé !

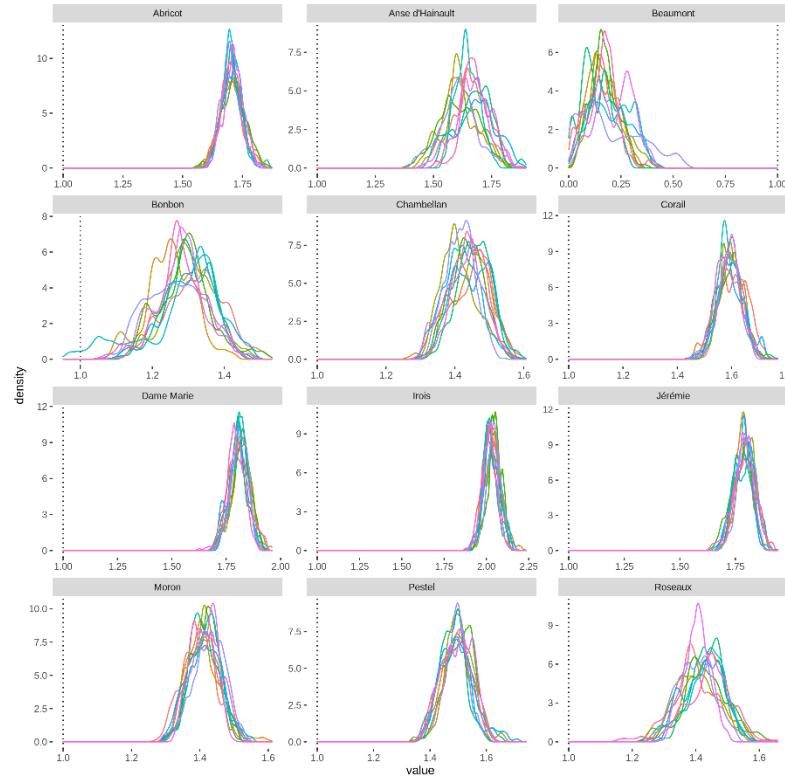
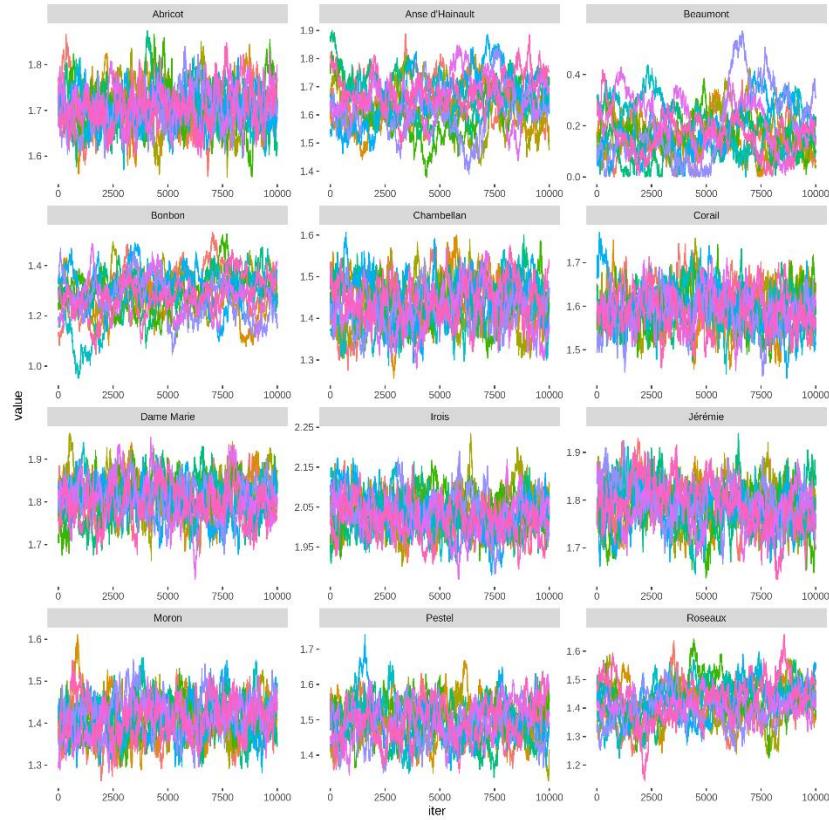
Cela peut se faire visuellement ou avec des tests de convergence.

Il est généralement mieux de simuler plusieurs chaînes indépendantes et s'assurer qu'elles convergent vers la même distribution.

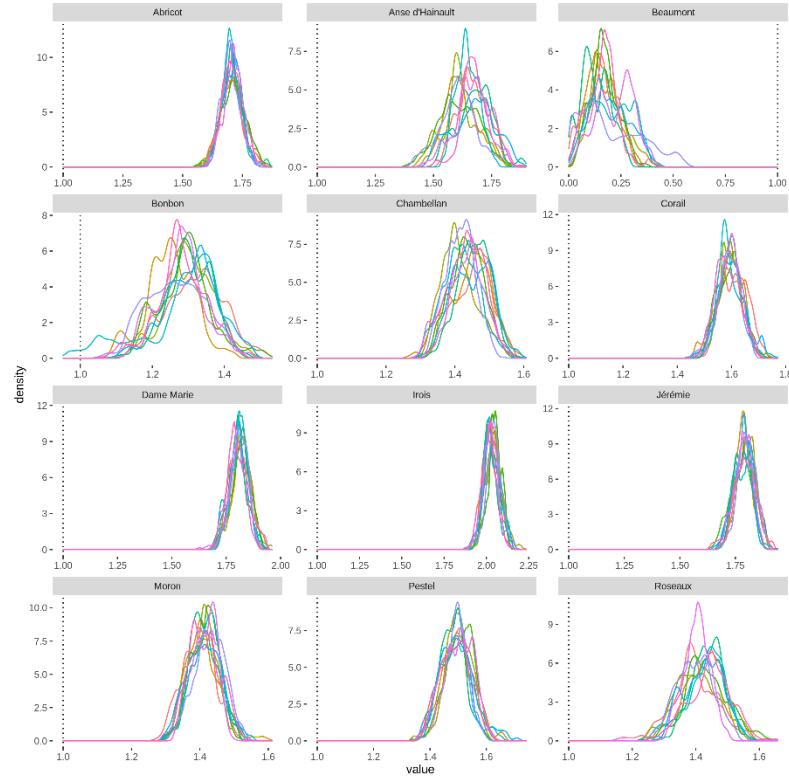
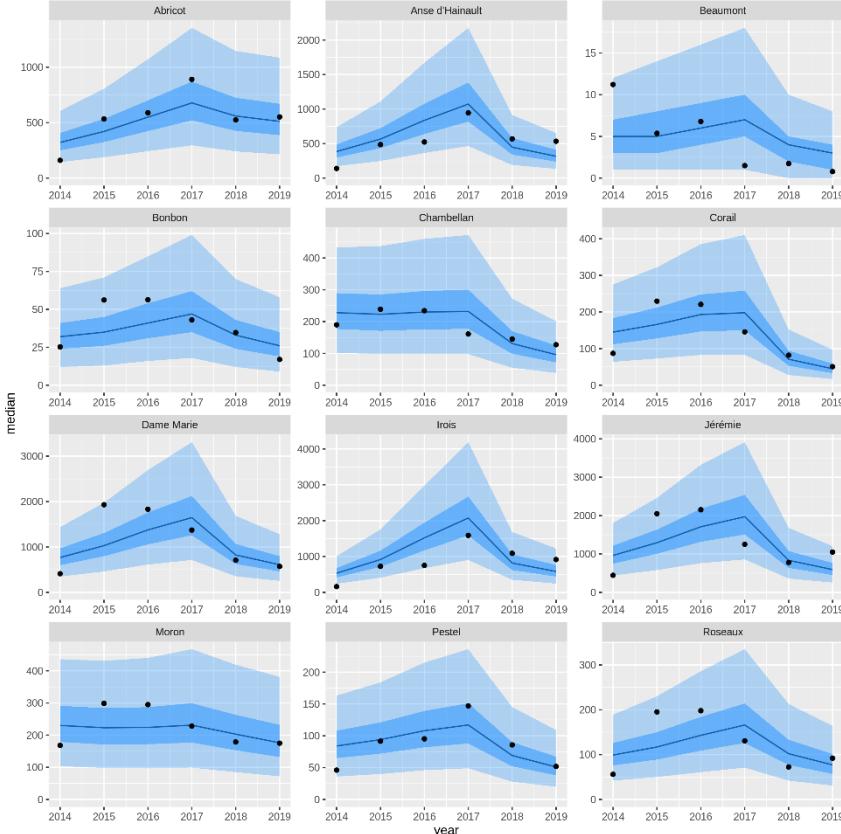


Il existe d'autres algorithmes de MCMC (e.g. Gibbs sampling, HMC, etc.) dont on ne parlera pas aujourd'hui.

MCMC: l'exemple du paludisme en Haïti



MCMC: l'exemple du paludisme en Haïti





Quelques mots de conclusion

La modélisation en épidémiologie a bien sûr ses limites!



- **Un modèle mathématique ne prédit pas le futur**
 - Le modèle correspond à une réalité simplifiée
 - Généralement, une analyse de modélisation est conçue pour répondre à une question spécifique



- **Le modèle presuppose des conditions parfaites et ne tient pas compte de tous les aspects non épidémiologiques**



- **Le modèle repose sur des hypothèses**
 - Certain éléments ne sont pas inclus, d'autres sont très simplifiés, d'autres moins simplifiés
 - Le meilleur modèle n'est pas toujours le plus complexe
 - Un modèle conçu pour répondre à une question peut ne pas être adapté pour répondre à une autre question



- **Le modèle repose sur la disponibilité et la qualités des données**



- **La modélisation prend du temps**
 - Pour développer, adapter et utiliser les modèles
 - Pour échanger avec les utilisateurs du modèle pour s'assurer que le modèle répond à la bonne question avec des hypothèses raisonnables



- **Un modèle est toujours incertain... Et toutes les incertitudes ne sont pas toujours tractables**
 - Un bon modèle peut aider à naviguer et explorer cette incertitude

Take home message

- La modélisation mathématique apporte un éclairage utile sur la compréhension des épidémies et peut s'avérer un outil utile pour la prise de décisions de santé publique
- Comme tout exercice de modélisation, les modèles épidémiologiques ont des limites
- Confronter les modèles avec des données peut nécessiter l'usage de méthodes statistiques plus ou moins complexes. L'exemple du MCMC a été présenté

Merci pour votre attention !

clara.champagne@swisstph.ch

Questions ?

Remerciements:

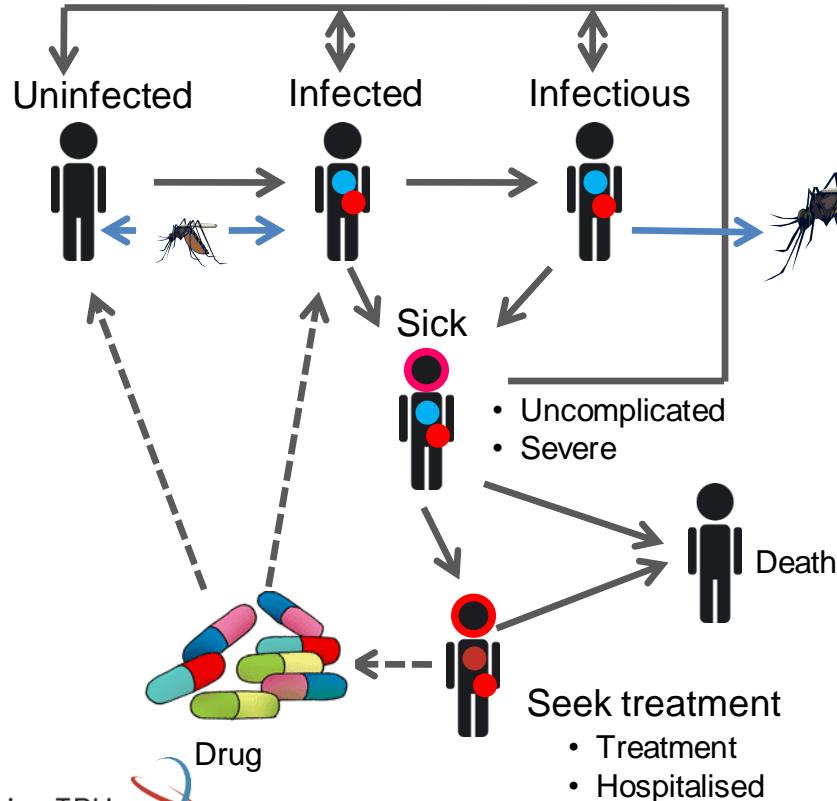


BILL & MELINDA
GATES foundation

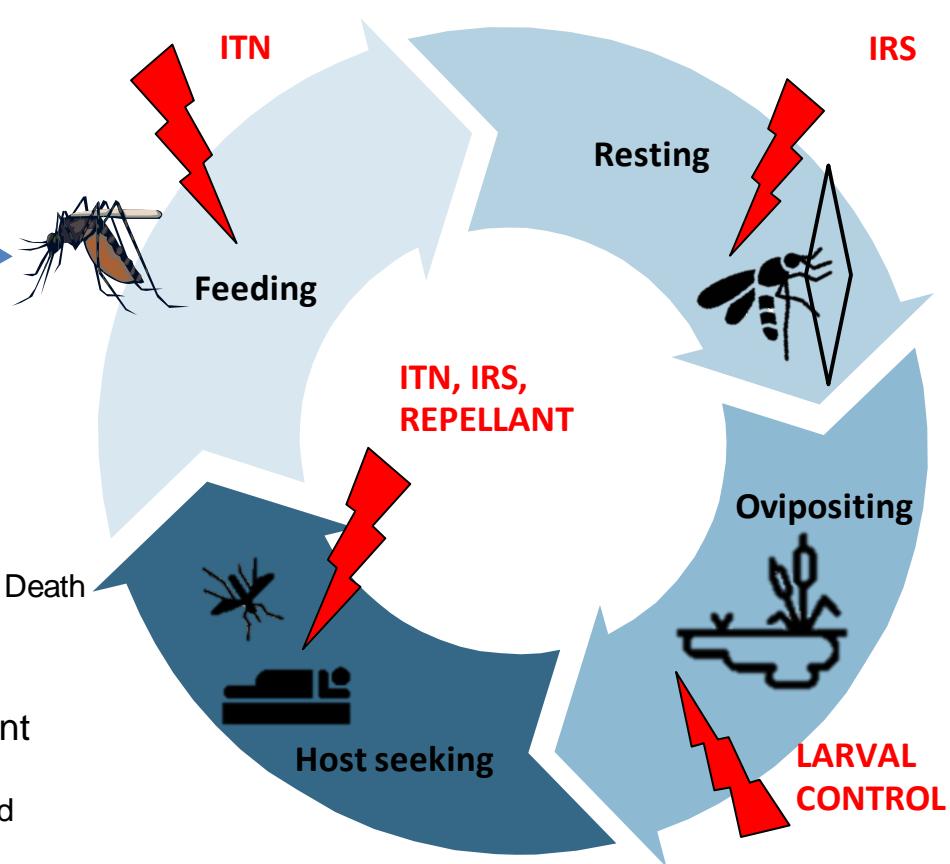


OpenMalaria : an individual-based model for malaria

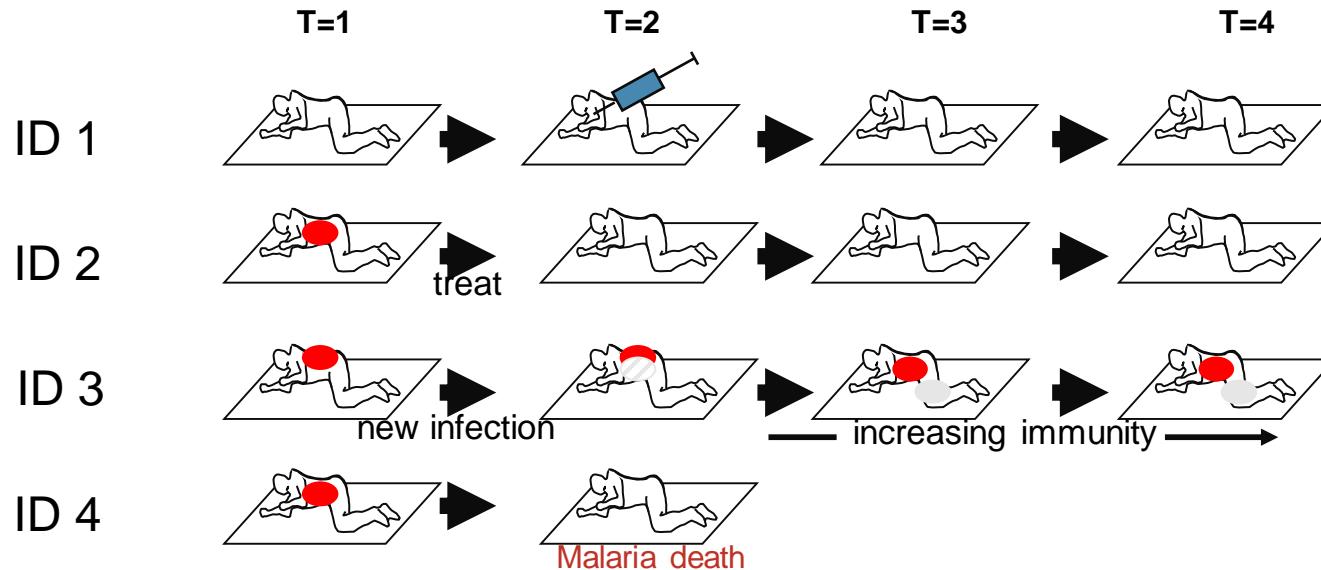
Simulation of **the dynamics of transmission** for each individual.



Mosquito behaviour

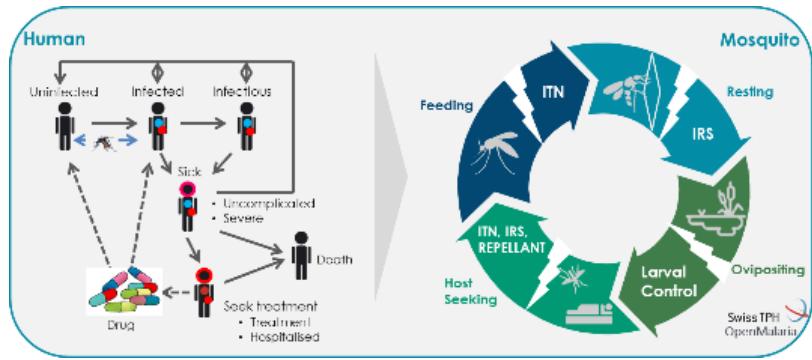


Each individual has a different history of malaria infection

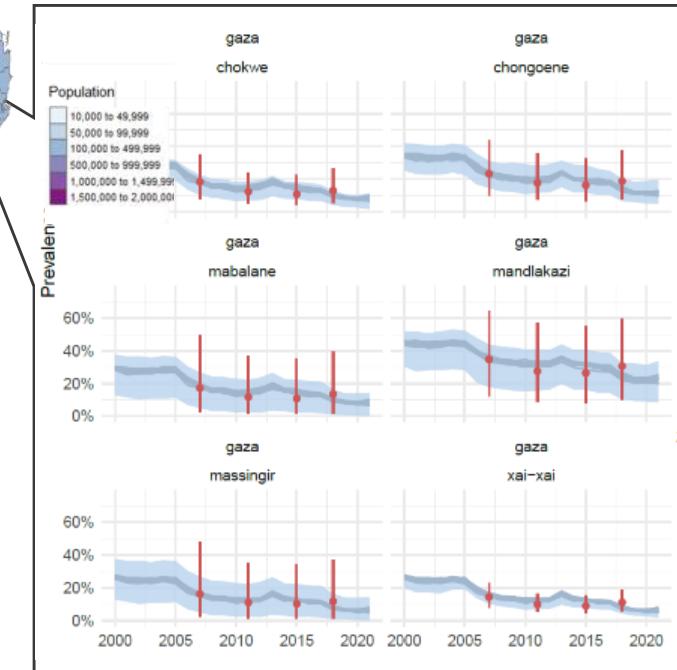
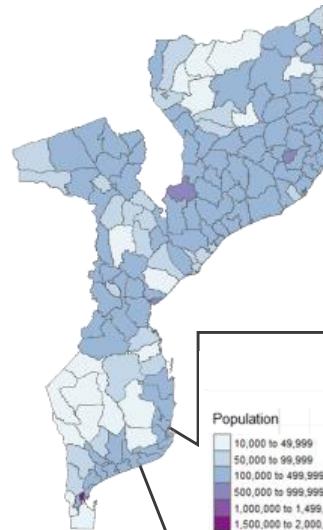
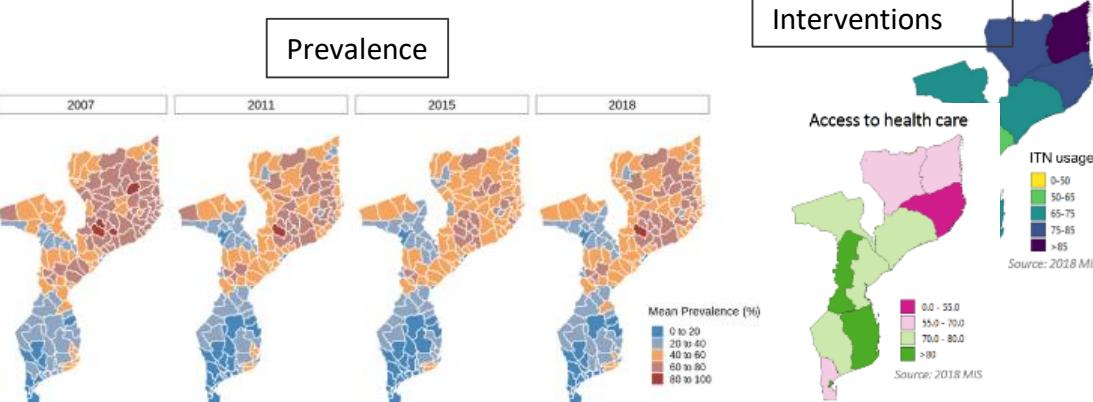


3. Modelling support in Mozambique

OpenMalaria

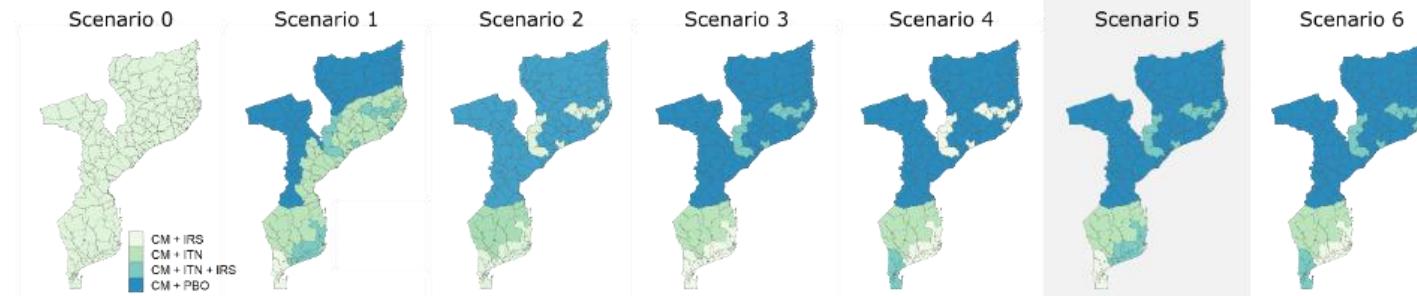


Country specific data from Mozambique

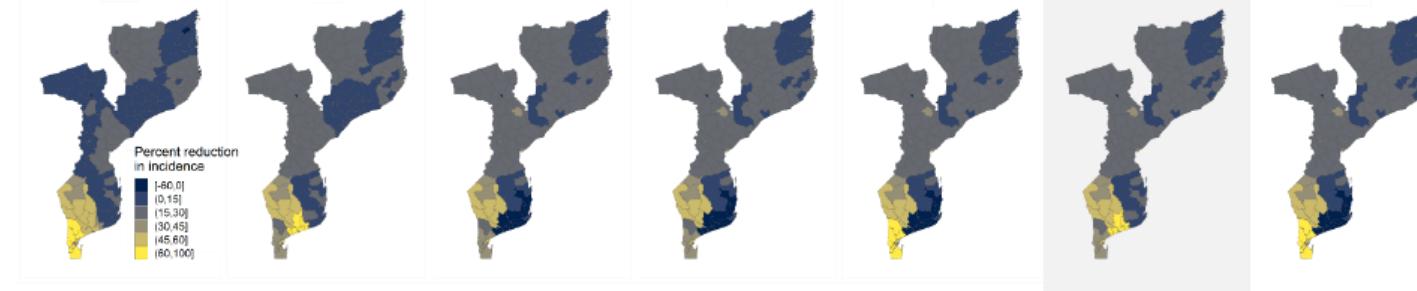


3. Modelling support in Mozambique

Intervention packages by scenario



Percent Reduction in incidence by scenario
2024 vs 2018



Analysis used as evidence to support the Mid-term Review (MTR) of the Strategic Plan, stratification update, and subsequent 2020 Global Fund application (which was approved and it now being implemented)