

## 1 Premier problème :

### 1.1 Décrivez la problématique soulevée par l'auteur et qui justifie cet article ? En quoi l'approche classique de l'analyse de telles données peut conduire à des biais ?

La problématique traitée par l'auteur est qu'il est difficile de quantifier l'effet de la durée d'un traitement sur différents patients. Pour cela, il présente une méthode qui va au delà de la comparaison naïve (qui introduit un "biais de temps immortel" ) entre les patients qui ont bénéficiés du traitement pour une longue durée et ceux pour une courte durée.

L'approche classique de comparaison mélange dans la comparaison tous les patients ayant bénéficié du traitement. Or certains ont en bénéficié pour une plus longue durée que d'autres car ils sont restés en vie (ou dans l'étude) pour une plus longue durée. C'est exactement ce qui introduit le "biais de temps immortel" : Par définition, la moyenne du temps de survie de ces patients là sera plus élevée.

### 1.2 Décrivez la proposition de l'auteur pour contourner cette difficulté.

La stratégie adoptée se base sur 3 étapes :

- **Le clonage** : On observe l'état du patient à l'instant initial puis on l'assigne à tous les traitements étudiés qui sont compatibles avec cette observation à l'état initial. Cela a pour effet de faire une sorte de sur-échantillonnage bien contrôlé des données observées.
- **La censure** : Quand un sujet dans notre dataset augmenté dévie de la stratégie où il a été assigné, on le censure. En faisant cela, un autre biais est introduit : le biais de sélection dû justement à cette opération de censure.
- **Pondération** : En utilisant le "inverse probability weighting", le but est de donner un poids à chacun des individus de la base de données augmentée afin d'éliminer le biais de sélection introduit par l'étape de censure. Chaque individu reçoit un poids égale à l'inverse de sa probabilité d'être censuré. Les individus censurés ayant un poids égal à 0.

### 1.3 Quelle autre approche de modélisation pourrait être utilisée pour analyser correctement ce type de données

[1]Le recours à l'appariement est la meilleure façon d'éviter un biais de temps immortel. L'appariement consiste à faire correspondre individuellement chaque participant du bras de traitement avec un participant du bras placebo. Si un participant du bras de traitement doit attendre 30 jours pour débiter l'essai, on l'apparie avec un participant du bras placebo qui a également survécu ou qui a survécu sans symptômes pendant 30 jours avant le recrutement. Ce type de correspondance est également appelé l'établissement de paires appariées.

## 2 Deuxième problème :

### 2.1 Approche de modélisation :

Le modèle choisi pour décrire la propagation du virus est GLEAM (Global Epidemic and Mobility Model).

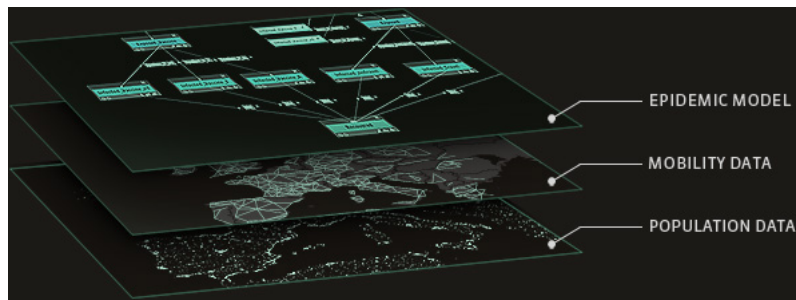


Figure 1: Modèle 3 couches

Ce modèle se base sur 3 couches de données principales :

### 2.1.1 Données de population :

Issue de sources de recensement, elles représentent la distribution de la population sur les différentes zones géographiques du globe terrestre. Cette distribution est faite suivant une subdivision par carrés de  $25Km \times 25Km$ , puis en regroupant ces carrés comme des cellules de Voronoï centrées autour d'un aéroport (en considérant la distance entre ces carrés précédemment définis ainsi que l'aéroport le plus proche, cela génère plus de 3300 cellules de Voronoï chacune centrée sur un aéroport)



Figure 2: Modèle de population

### 2.1.2 Données de mobilité de la population :

La mobilité des populations précédemment définies est supposée se tramer suivant deux mécanismes : le voyage entre aéroports et les voyages terrestres.

En se basant sur des données issues de OAG (Official Airline Guide), les gros tout comme les petits aéroports sont modélisés fidèlement avec les différentes paires origine-destination ainsi que le volume de voyageurs transportés propres à chaque aéroport.

D'une manière analogue, en utilisant des données statistiques nationales, les transports entre les cellules voisines ont été modélisées pour capturer les diffusions locales de l'infection entre elles.

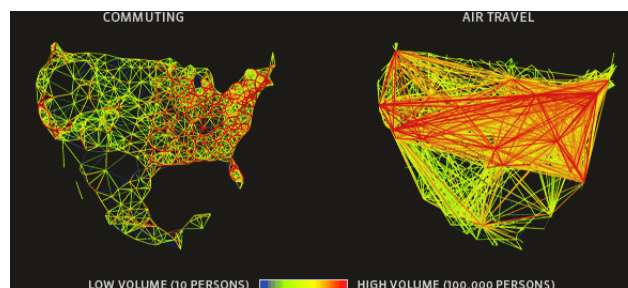


Figure 3: Modèle de mobilité

### 2.1.3 Dynamique d'infection :

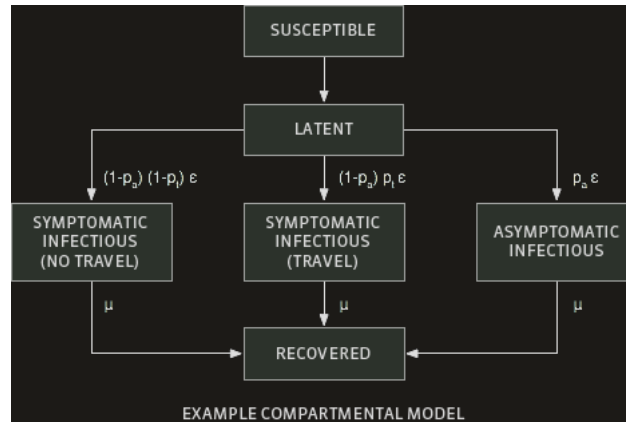


Figure 4: Modèle d'infection

Les dynamiques de propagation de l'infection dans une même cellule de Voronoï ainsi qu'entre les cellules sont définies par un modèle dit "compartmental" [2]. Les compartiments (états) que peuvent occuper les individus dans le modèle adopté sont :

- Susceptible (S) : C'est l'état dans lequel sont les individus qui ne sont pas infectés par le virus.
- Latent (L) : Lorsqu'un individu du compartiment  $S$  est exposé à la maladie, il ne devient pas nécessairement capable de la transmettre immédiatement. C'est donc dans cet état où il passe.
- Infectieux (I) : C'est l'état où figure un individu infecté et capable de transmettre le virus.
- Recovered (R) : C'est l'état où passe un individu qui n'est plus capable de transmettre le virus soit parce qu'il est : guéri, en quarantaine ou bien décédé.

Les probabilités de passage d'un état à un autre dépendent de la durée dans lequel un individu perdure dans son état actuel (probabilité de passer de L à I inversement proportionnelle au temps passé dans l'état L, probabilité de passer de I à R inversement proportionnelle au temps passé dans l'état I).

Le modèle est à temps discret où chaque pas de temps représente 1 jour afin que les dynamiques de transport et de transmission d'infection aient lieu.

## 2.2 Calibrage :

Le modèle est calibré en utilisant le calcul Bayésien approché (ABC : Approximate Bayesian computation) afin d'estimer  $P(R_0|D)$  : à partir d'informations et d'études antérieures sur le virus, des plages de paramètres plausibles ont été établies pour les conditions initiales de l'épidémie (durée de la période de latence et d'infection, intervalle de début de l'épidémie...). A partir de là, on échantillonne une combinaison de paramètres qui nous donne un temps de génération  $T_g$  afin de simuler un scénario épidémique. On compare le scénario épidémique obtenu avec le scénario observé en réalité (avec un certain seuil de tolérance de comparaison entre le scénario simulé et les vraies données car ces dernières sont issues de journaux d'informations et ne sont pas forcément exactes et complètes). Finalement, on estime la probabilité a postérieure de  $R_0|D$  voulue ce qui donne la distribution suivante (par exemple avec  $T_g = 7.5$  :

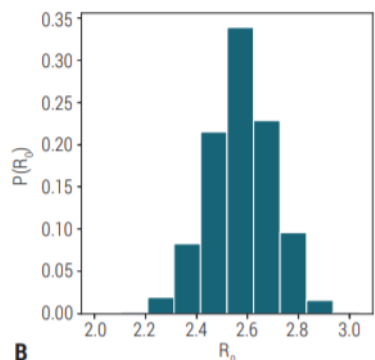


Figure 5: Distribution a postérieure du taux de reproduction de base  $R_0$

## 2.3 Résultats :

Le modèle ainsi construit a montré que l'épidémie s'était déjà propagée à d'autres grandes villes Chinoises quand la décision de restreindre les déplacements vers/ en provenance de Wuhan fut décrétée. En effet cela n'a eu pour effet que de retarder l'évolution du virus de quelques jours. Néanmoins cela a permis de ralentir la propagation du virus au niveau international par les sujets en provenance de Wuhan directement. Mais de la même manière, cela n'a pas empêché totalement sa propagation (dans les 2-3 semaines suivantes) vers d'autres pays (la transmission a persisté au travers les sujets en provenance des autres grandes villes chinoises comme le montre la **Fig3** du papier d'origine).

Cela mène au second résultat avancé : les restrictions de voyages seules ne sont pas la solution face au virus mais cela doit obligatoirement être couplé à des efforts comportementaux pour agir sur les probabilités d'infection et en limiter les transmissions.

## References

- [1] Académie européenne des patients. Biais de temps immortel.
- [2] Wikipedia. Modèles compartimentaux en épidémiologie.