

Soutenance du TER

Koen Aristote

Université Paris-Saclay

24 juillet 2020



- 1 Introduction
- 2 Contexte
- 3 Hypothèse
- 4 Méthodes, modélisation et Données
- 5 Résultats de l'étude :

Durant ce TER nous nous intéressons à l'article de Benjamin D.Dalziel : "Persistent chaos of measles epidemics in the pre-vaccination United-States caused by a small change in seasonal transmission patterns" (2016) ainsi qu'au modèle utilisé pour modéliser les épidémies de rougeole.

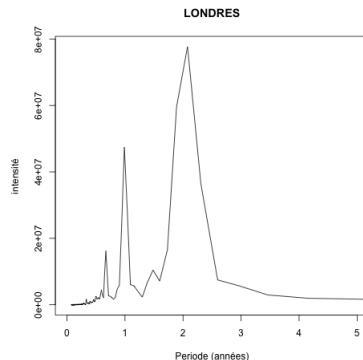
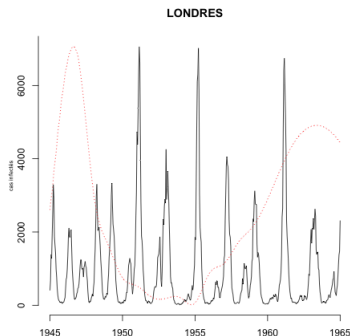
Travail effectué

- Obtention des données utilisées dans l'article
- Etude du modèle TSIR, estimation et prédiction
- Reproduction des résultats sur l'influence de la période de basse de transmission

CONTEXTE

Des cycles stables au Royaume-Uni

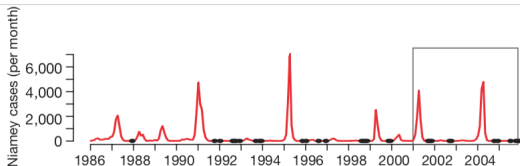
- Les cycles d'épidémies de rougeole au Royaume-Uni sont stables (Période de 2 ou 1 an)



Des cycles instables au Niger

Des fortes perturbations démographiques au Niger entraînent :

- Une divergence des cycles observés au Royaume-Uni vers des cycles irréguliers de haute périodicité moyenne (5 ans)
- Des pics d'amplitude variables
- Des périodes d'extinction de la maladie (Ferrari 2008)



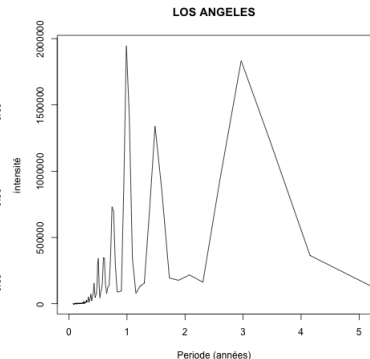
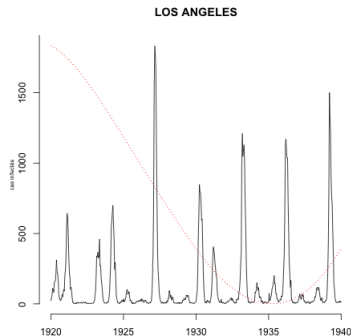
Opinion générale :

La seule manière de dévier des cycles stables d'épidémies est à travers de fortes perturbations démographiques

Des cycles instables persistants aux Etats-Unis

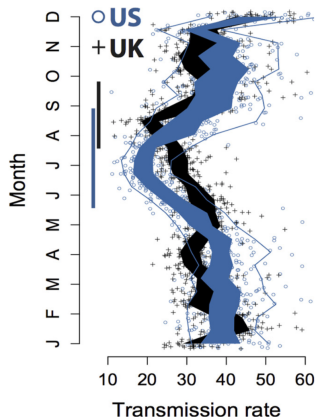
Aux Etats-Unis, on semble observer des cycles :

- De plus haute periodicité moyenne
- D'amplitude variable
- Avec une persistance de la maladie



Des différences sur le taux de transmission

On observe une différence entre les Etats-Unis et le Royaume-Uni sur la durée et l'amplitude de la période de basse transmission (Dalziel 2016).



Cette trajectoire est le résultat de la différence sur la période de basse transmission

On cherche donc à montrer :

- Possibilité de dévier vers des cycles instables de la rougeole en appliquant de faibles perturbations sur la durée de basse transmission en été.
- Les épidémies sont de plus persistantes.

Les méthodes que nous utiliserons sont :

- Comparaison entre 40 villes (EU) et 40 villes (RU) sur 20 ans
- Modélisation des cas infectés par le modèle TSIR
- Simulations des cas infectés 100 ans en avant
- Comparaison des spectrogrammes de simulations
- Comparaison des exposants de Lyapunov pour confirmer la nature chaotique des épidémies aux US
- Etude du comportement des épidémies en faisant varier l'amplitude ou la durée de la période de basse transmission sur les simulations
- Comparaison du nombre de semaines avec moins de 1 cas

Le modèle TSIR : Time-series Susceptible Infected Recovered

Notations :

- I_t le nombre de cas à la mesure t
- N_t la taille de la population à l'instant t
- β_t le taux de transmission à l'instant t , supposé saisonnier
- α un réel strictement positif
- S_t La taille de la population susceptible d'être infectée
- B_t Le nombre de naissances à l'instant t

Le modèle TSIR est établi comme suit :

$$\mathbb{E}[I_{t+1}] = \beta_t I_t^\alpha S_t N_t^{-1}$$

$$\text{et } I_{t+1} \sim \mathcal{NB}(\mathbb{E}[I_{t+1}], I_t)$$

$$\text{Avec } S_{t+1} = S_t + B_t - I_{t+1} + u_t$$

Réécriture du modèle

Problème :

- On ne connaît généralement S_t , besoin de la reconstruire
- I_t est souvent sous estimée

Si on considère :

$$I_{t+1} = \beta_t I_t^\alpha S_t N_t^{-1} \varepsilon_t$$

En passant au log et en supposant $\log(\varepsilon_t)$ comme un bruit gaussien :

$$\log(I_{t+1}) = \log(\beta_t) + \alpha \log(I_t) + \log(S_t) - \log(N_t) + \log(\varepsilon_t)$$

Si I_t est le nombre de cas réel alors on modélise :

- $I_t = \rho_t C_t$
- C_t est le nombre de cas reportés (les données)
- $\rho_t > 0$ le taux de rapports.

En remplaçant dans la 3ème équation du modèle :

$$S_{t+1} = S_t + B_t - \rho_{t+1} C_{t+1} + u_t$$

Réécriture du modèle (suite)

Si on considère $\sigma_t = \bar{S}N_t$ où \bar{S} est la proportion moyenne de susceptibles on peut écrire $S_t = \sigma_t + D_t = \bar{S}N_t + D_t$ avec D_t les déviations de la population par rapport à σ_t . On a alors :

$$D_{t+1} = D_t + B_t - \rho_{t+1}C_{t+1} + u_t + \bar{S}(N_t - N_{t+1})$$

$$\iff D_{t+1} = D_0 + \sum_{i=0}^t B_i - \sum_{i=0}^t \rho_{i+1}C_{i+1} + \sum_{i=0}^t u_i + \bar{S}(N_0 - N_{t+1})$$

$$D_{t+1} = D_0 - \rho \sum_{i=0}^t C_{i+1} + \sum_{i=0}^t B_i - \sum_{i=0}^t (\rho_{i+1} - \rho)C_{i+1} + \sum_{i=0}^t u_i + \bar{S}(N_0 - N_{t+1})$$

$$\sum_{i=0}^t B_i = -D_0 + \rho \sum_{i=0}^t C_{i+1} + D_{t+1} + \sum_{i=0}^t (\rho_{i+1} - \rho)C_{i+1} - \sum_{i=0}^t u_i + \bar{S}(N_{t+1} - N_0)$$

Dans le cas où les 3 derniers termes sont nuls on peut estimer ρ comme la pente de regression entre le cumul des naissances et des cas rapportés et obtenir D_t comme les résidus de la régression.

Estimation des paramètres

En remplaçant S_t par sa nouvelle expression on a :

$$\log(I_{t+1}) = \log(\beta_t) + \alpha \log(I_t) + \log(\bar{S}N_t + D_t) - \log(N_t) + \log(\varepsilon_t)$$

En connaissant \bar{S} et ayant extrait les résidus de la régression, on a un modèle linéaire

Pour estimer les paramètres :

- Calcul de la vraisemblance pour plusieurs valeurs candidates de \bar{S} avec α et β_t estimés pour chaque valeur de \bar{S}
- Choix de \bar{S} maximisant la vraisemblance
- Estimation de beta et alpha pour la valeur choisie de \bar{S}

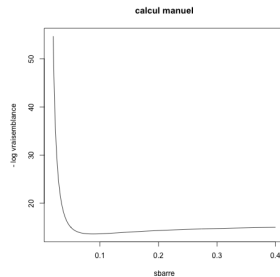
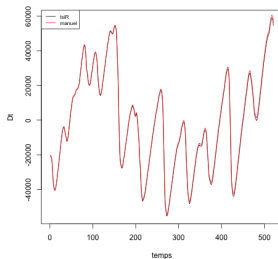
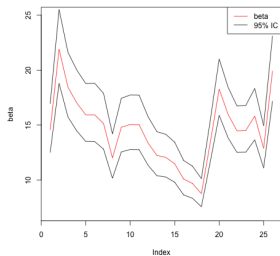
On estime donc 28 paramètres au total.

Une fois les paramètres estimés on peut donc simuler les séries à travers la loi négative binomiale, ou selon le modèle déterministe :

$$I_t = \mathbb{E}[I_t]$$

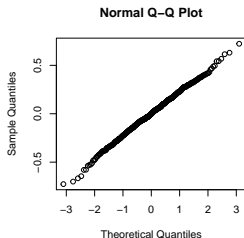
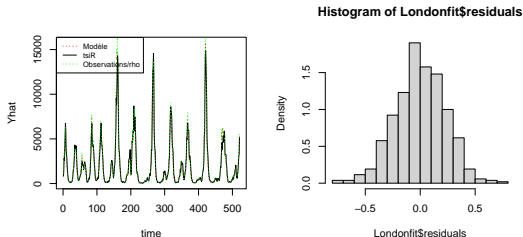
- **Cas de rougeole et population** : travaux plus anciens pour le RU et Tycho project pour les EU
- **Population** : Travaux anciens pour RU, recensement pour EU
- **Nombre de naissances** : Estimés en différenciant la série bimensuelle des enfants âgés de moins d'un an, ajustée du taux d'enfants sortant de cette catégorie
- **Villes choisies US** : les 40 villes aux EU avec le plus de cas de rougeole
- **Villes choisies RU** : Les 39 plus grandes villes à plus de 50km de Londres
- **Periode étudiée** : 1920-1940 US, 1944-1964 UK.

Estimation des paramètres



Ajustement du modèle

On observe un bon ajustement du modèle avec des résidus qui semblent bien normalement distribués autour de 0.



Nous avons ensuite effectué des simulations :

Sur la période étudiée à partir du modèle déterministe

Sur la période étudiée à partir du modèle stochastique

100 ans en avant à partir du modèle déterministe

Dans le but de :

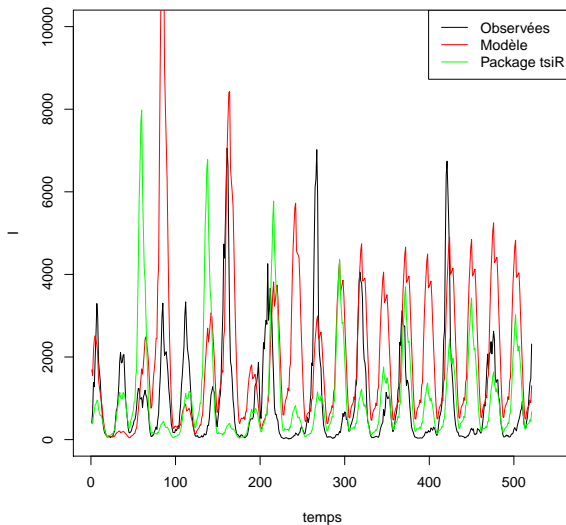
Vérifier la qualité des modèles

Etudier le comportement du modèle stochastique

Pouvoir avoir assez de données pour calculer les coefficients de Lyapunov et tracer les bifurcations

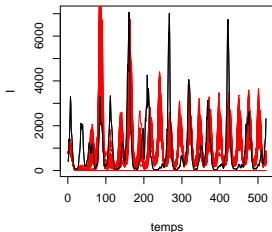
Qualité de la simulation déterministe

Qualité de la simulation déterministe

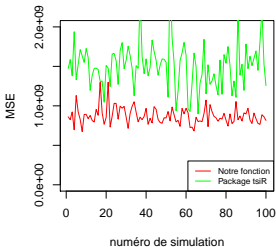
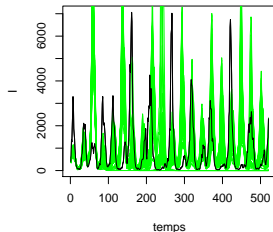


Qualité de la simulation stochastique

100 simulations stochastiques

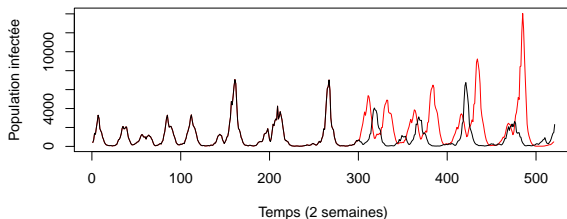


100 simulations stochastiques avec package

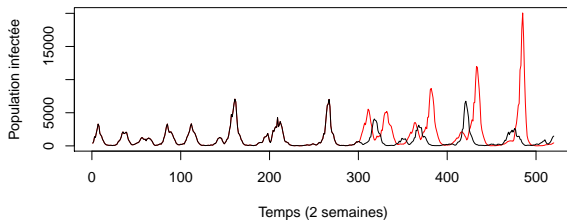


Qualité des prédictions

Déterministe

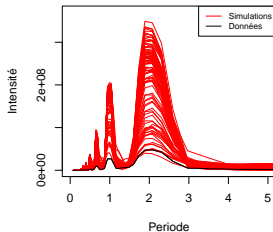


Stochastique

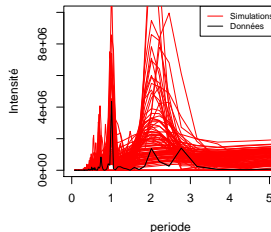


Periodogrammes des simulations stochastiques

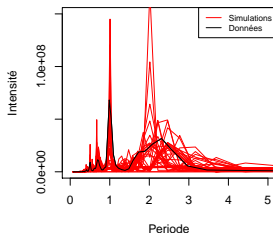
LONDON



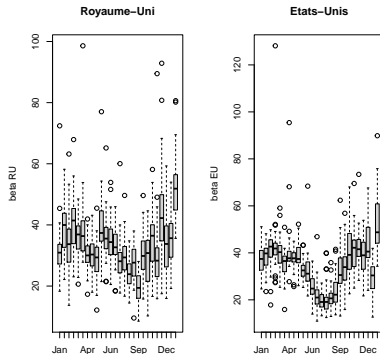
BOSTON



NEW YORK



Boxplots des valeurs de β_t



Exposants de Lyapunov

Soit $u_{t+1} = f(u_t)$, l'exposant de Lyapunov mesure le logarithme moyen de l'accroissement, c'est une mesure de l'amplification de l'erreur. Dans le cas d'une fonction $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \log(\|J_t U_0\|)$$

avec

$U_0 = (1, 0, \dots, 0)$ et J_t la jacobienne à l'instant t

Dans notre cas on a :

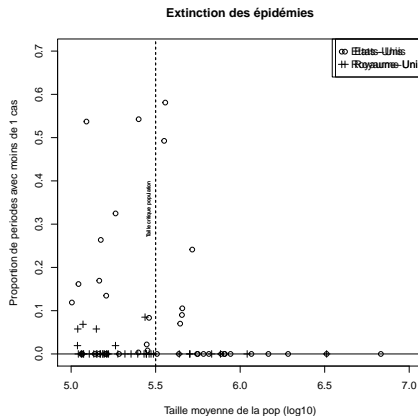
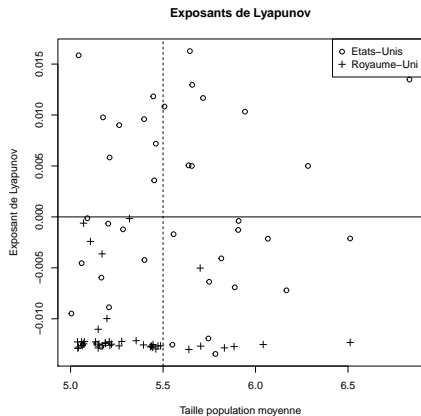
$$(S_{t+1}, I_{t+1}) = (S_t + B_t - \beta_s I_t^\alpha S_t N_t^{-1}, \beta_s I_t^\alpha S_t N_t^{-1}) = f(S_t, I_t)$$

et donc $U_0 = (1, 0)$ la jacobienne vaut :

$$\begin{pmatrix} 1 - \beta_s I_t^\alpha / N_t & -\beta_s S_t (I_t^{\alpha-1} \alpha / N_t) \\ \beta_s I_t^\alpha / N_t & \beta_s S_t (I_t^{\alpha-1} \alpha) / N_t \end{pmatrix}$$

Exposants de Lyapunov, Persistence

Exposants de Lyapunov de chaque ville et proportion de périodes avec moins de deux cas en fonction de la taille de la population moyenne pour des simulations du modèle déterministe 100 ans en avant.



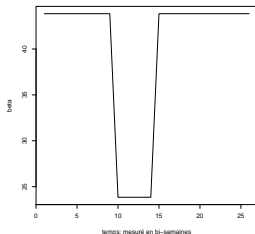
Bifurcations

Pour l'étude de l'influence des perturbations sur la période de basse transmission :

- Création de plusieurs valeurs de la fonction du taux de transmission à travers une fonction synthétique
- Tracé des bifurcations sur des simulations 100 ans en avant

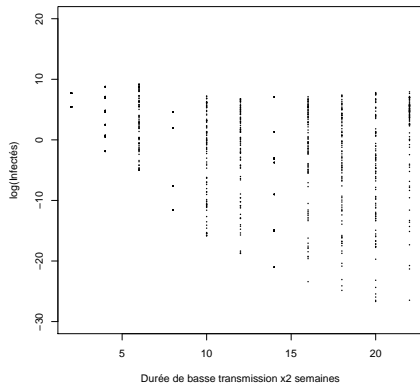
La fonction du taux de transmission synthétique est définie par :

$$\beta_t = \begin{cases} \beta_- & \text{si } a \leq t \leq b \\ \beta_+ & \text{sinon.} \end{cases}$$



Bifurcations

Bifurcation en fonction de la durée de basse transmission



bifurcation en fonction de l'amplitude

