Paramétrer des modèles épidémiologiques

Vladimir GROSBOIS
Biologiste des populations et biostatisticiens
CIRAD, département ES,
Equipe Animal et Gestion Intégrée des Risques

Introduction

Les étapes de l'analyse des dynamiques épidémiologiques par la modélisation

- Choisir un type de modèle
  - Modèle compartimenté (SIR, SEIR, SIS,etc……)
  - Modèle individu centré
  - Modèle stochastique
  - Modèle spatialisé
  - Modèle de réseau
- Définir la structure de ce modèle
  - Définir les paramètres du modèle
  - Décrire les équations qui gouvernent le comportement du modèle
    - Equations de changement d'état
    - Equation de changement d' effectifs des compartiments
    - Equations qui gouvernent le comportement spatial des unités épidémiologiques
    - ……..
- Etablir les valeurs des paramètres du modèle
- Etudier le comportement du modèle
  - Méthodes analytiques
  - Simulations
  - Analyses de sensibilité

Les différents types de paramètres d'un modèle épidémiologique

- Les équations qui définissent le modèle

Susceptibles: $\frac{dS}{dt} = \kappa \log(1-c)\, SI = -\beta SI$

Infectieux: $\frac{dI}{dt} = \beta SI - \gamma I$

Rétablis: $\frac{dR}{dt} = \gamma I$

- Les paramètres du modèle
  - c : probabilité de transmission lors d'un contact entre un infectieux et un susceptible

    κ : nombre de contact d'un individu avec d'autres individus de la population par unité de temps

    γ: probabilité de guérison au cours d'une unité de temps

    $\beta = -\kappa \log(1-c)$: paramètre de transmission

    $\lambda = \beta$ I: force d'infection = probabilité de transition de l'état susceptible à l'état infectieux par unité de temps

Les différents types de paramètres d'un modèle de dynamique épidémiologique. Exemple du modèle SIR

- Interaction hôte-pathogène
- c : probabilité de transmission lors d'un contact entre un infectieux et un susceptible
- γ : taux de guérison = probabilité de guérison au cours d'une unité de temps
- dans d'autres modèles: temps de latence, taux de réversion sérologique, taux de mortalité due à l'infection, etc…
- Comportement des hôtes
- κ: nombre de contacts d'un individu avec d'autres individus de la population par unité de temps
- dans d'autres modèles: démographie de la population d'hôte, structure des contacts dans la population d'hôtes, etc…
- Ecologie du pathogène
  - Par exemple survie dans l'environnement
- Interactions entre pathogènes (modèle multi-pathogènes)
  - Paramètres décrivant la compétition entre pathogènes ou entre souches

Principe de la paramètrisation

- Etablir des valeurs plausibles pour les paramètres du modèle:
  - Valeur unique
  - Une distribution de probabilité (pour prendre en considération l'incertitude ou la variabilité)
  - Scénarios

Principe de la paramétrisation

- Sans s'appuyer sur des données
  - Littérature
  - Opinion d'expert
- En s'appuyant sur des données
  - Estimation
  - Modélisation

Inférer les paramètres épidémiologiques par l'analyse de données. Nature des données

- Les données peuvent être des observations directes du paramètre à estimer
  - Des taux de contact peuvent être mesurés à l'aide de capteurs de proximité
  - Drewe et al. 2013: mesure des contacts entre blaireaux et bétail

Inférer les paramètres épidémiologiques par l'analyse de données. Nature des données

- Les données peuvent être des observations directes du paramètre à estimer
  - Des expérience de transmission peuvent fournir des observations directes d'évènements de transmission et être utilisées pour estimer des probabilités de transmission
  - Dekker et al. 2008: estimation de taux de transmission de la FMD à partir de données expérimentales

Inférer les paramètres épidémiologiques par l'analyse de données. Nature des données

- Les données peuvent être des observations directes du paramètre à estimer
  - Des infections expérimentales peuvent fournir des observations directes de temps de latence, de durée infectieuse ou de mortalité induite par la maladie
  - voir par exemple Hulse-Post et al. 2005 pour l'estimation des durée d'excrétion de virus d'influenza aviaire chez les canards
  - Mardones et al 2010 pour l'estimation des durées des différents stades infectieux pour les virus de FMD

Inférer les paramètres épidémiologiques par l'analyse de données. Nature des données

- Les données peuvent être des observations directes du paramètre à estimer
  - Des changements d' états (susceptibles -> infectieux; infectieux->rétablis) peuvent être observés lors de suivis longitudinaux à l'échelle individuelle
  - Faustino et al. 2004: estimations de taux d'infections de moineaux par un mycoplasme provoquant une conjonctivite à partir d'observation d'individus marqués individuellement

Inférer les paramètres épidémiologiques par l'analyse de données. Nature des données

- Les données peuvent être des observations des phénomènes engendrés par les processus décrits par les paramètres à estimer
  - Analyse rétrospective d'une épidémie
  - On dispose de données de type série temporelle de nombre de cas
  - Une structure de modèle est définie pour décrire les processus qui ont engendré cette série temporelle
  - On cherche à estimer les paramètres du modèle par la modélisation statistique de la série temporelle du nombre de cas
- Par exemple Haydon et al. 1997: estimation des paramètres de transmission de la fièvre aphteuse à partir d'une série temporelle du nombre de troupeaux infectés lors d'une épidémie

Inférer les paramètres épidémiologiques par l'analyse de données. Nature des données

- Les données peuvent être entachées de biais de perception
  - Par exemple les recensements de nombres de cas peuvent être biaisés par la sous-déclaration
  - Les observation de transitions entre états peuvent être biaisées par l'utilisation de tests imparfaits pour détecter le pathogène ou les anticorps

Type d'inférence: estimer des paramètres des modèles épidémiologiques

- Durées d'excrétion
- Temps de latence
- Temps de séroconversion
- Taux de contact
- Taux d'infection
- etc……

Type d'inférence: décrire des variations des paramètres des modèles épidémiologiques

- Le temps d'excrétion de virus d'influenza aviaire chez les canards varient-ils en fonction de l'âge des individus ?
- Les taux de contact entre bétail et blaireaux dépendent ils du type d'habitat autour des près utilisés par le bétail ?

Type d'inférence: tester des hypothèses alternative portant sur les processus épidémiologiques

- Par exemple Camacho et al. 2011 cherchent à identifier le modèle épidémiologique le plus compatible avec l'observation d'une courbe épidémique bimodale lors d'une épidémie de grippe sur l'île de Tristan da Cunha en 1971
  - Les données dont ils disposent consistent en une série temporelle de nombre de cas
  - Des modèles alternatifs (correspondant à différentes hypothèses portant sur les processus épidémiologiques) sont élaboré
  - On cherche à identifier l'hypothèse la plus plausible en termes de compatibilité des modèles avec les données

Méthodes utilisées pour inférer les paramètres épidémiologiques

- Dans le cas où les données sont des observations directes des paramètres, des méthodes statistiques classiques ou même pas de statistiques

- Dans le cas où les données sont des observations entachées de biais de perception ou des observations des phénomènes générés par les processus épidémiologiques, des modèles statistiques sophistiqués, voire très sophistiqués…..

- Camacho, A., Ballesteros, S., Graham, A.L., Carrat, F., Ratmann, O., Cazelles, B., 2011. Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. Proceedings of the Royal Society B: Biological Sciences 278, 3635-3643.
- Dekker, A., Vernooij, H., Bouma, A., Stegeman, A., 2008. Rate of Foot-and-Mouth Disease Virus Transmission by Carriers Quantified from Experimental Data. Risk Analysis 28, 303-309.
- Drewe, J., O'CONNOR, H., Weber, N., McDonald, R., Delahay, R., 2013. Patterns of direct and indirect contact between cattle and badgers naturally infected with tuberculosis. Epidemiology and infection 141, 1467-1475.
- Faustino, C.R., Jennelle, C.S., Connolly, V., Davis, A.K., Swarthout, E.C., Dhondt, A.A., Cooch, E.G., 2004. Mycoplasma gallisepticum infection dynamics in a house finch population: seasonal variation in survival, encounter and transmission rate. Journal of Animal Ecology 73, 651-669.
- Haydon, D., Woolhouse, M., Kitching, R., 1997. An analysis of foot-and-mouth-disease epidemics in the UK. Mathematical Medicine and Biology 14, 1-9.
- Hulse-Post, D., Sturm-Ramirez, K., Humberd, J., Seiler, P., Govorkova, E., Krauss, S., Scholtissek, C., Puthavathana, P., Buranathai, C., Nguyen, T., 2005. Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia. Proc. Natl. Acad. Sci. U. S. A. 102, 10682-10687.
- Mardones, F., Sanchez, J., Alkhamis, M., Carpenter, T., 2010. Parameterization of the duration of infection stages of serotype O foot-and-mouth disease virus: an analytical review and meta-analysis with application to simulation models. Veterinary research 41, 45.

Estimations de paramètres de durée de séjour dans un état épidémiologique à partir de données expérimentales

Analyse des données expérimentales de durée d'excrétion, temps de latence, temps de séroconversion

- Situation typique:
  - On inocule au temps t_0 le pathogène à un certain nombre d'animaux
  - On effectue à intervalle de temps régulier (chaque heure, chaque jour, chaque semaine) des mesures qui reflètent l'état épidémiologique de l'animal (excrétion, production d'anticorps, apparition de signes cliniques).
  - On obtient des dates de transition entre états épidémiologiques: transition latent-infectieux, transition infecté-rétablis, etc……)

Analyse des données expérimentales de durée d'excrétion, temps de latence, temps de séroconversion

- Données obtenues:
  - On obtient des dates de transition entre états :
    - transition latent-infectieux
    - transition infectieux-rétabli
    - transition transition infecté-mort
    - etc….
  - Et donc des temps écoulés jusqu'à la transition

Analyse des données expérimentales de durée d'excrétion, temps de latence, temps de séroconversion

- Modèles statistiques pour analyser ce type de données:
  - On utilise des modèles pour l'analyse de données de type « time to event »
  - Le modèle de référence pour ce type de données est le modèle de « proportional-hazards regression » de Cox

Cox proportional-hazards regression

- L'élément central du modèle est la fonction de risque instantané (instantaneous hazard function)
  - C'est la probabilité que l'évènement advienne sur un intervalle de temps (par exemple un jour). Cette fonction est noté h.

Cox proportional-hazards regression

- Il est possible de considérer différentes formes de la fonction de risque instantané
  - Risque constant: h=a
  - Notez que sous l'hypothèse d'une probabilité de transition d'état constante égale à a, la distribution du temps passé (d pour durée) dans cet état est une distribution exponentielle: prob (d>D) = exp(-aD)
  - Risque changeant linéairement avec le temps : h(t)=a+b(t)

Cox proportional-hazards regression

- Il est possible de décrire des variations de la fonction de hasard en fonction de covariables (par exemple l'âge des individus)
  - $\log(h_i)=a+ b\ x_i$
  - $h_i=\exp(a)*\exp(b\ x_i)$

Cox proportional-hazards regression

- Il est possible d'estimer des rapports de risque instantané
  - $\log(h_i)=a+ b\ x_i$
  - $h_i=\exp(a)*\exp(b\ x_i)$
  - $\log(h_j)=a+ b\ x_j$
  - $h_j=\exp(a)*\exp(b\ x_j)$
  - $h_i/h_j=\exp b(x_i - x_j)$

Intérêt de ce type de modèle

- Les estimations des risques instantanés sont des estimations des paramètres compartimentaux type SIR (par exemple le taux instantané de rétablissement (transition I -> R))

Logiciel

- package « survival » dans R

Bibliographie

Fox, J., 2002. Cox proportional-hazards regression for survival data. Appendix to « An R and S-PLUS Companion to Applied Regression »

Mardones, F., Sanchez, J., Alkhamis, M., Carpenter, T., 2010. Parameterization of the duration of infection stages of serotype O foot-and-mouth disease virus: an analytical review and meta-analysis with application to simulation models. Veterinary research 41, 45

Estimations de paramètres de transmission à partir de données expérimentales

Type d'expérience

- Un certain nombre d'animaux infectés mis en contact avec un certain nombre d'animaux susceptibles

- Le nombre d'individus susceptibles, d'individus infectés et d'individus rétablis est suivi à intervalle de temps régulier (par exemple quotidiennement)

Données générées par l'expérience

- Des séries temporelles de nombres d'individus dans les différentes classes (Susceptible, Infecté, Rétabli)

Estimation des paramètres épidémiologiques

- De nombreux modèles statistiques existent pour analyser ce type de données.
- Elle sont synthétisées dans Velthuis et al. 2007 pour les expériences portant sur des maladies animales
- Ces modèles sont généralement destinés à estimer les paramètres de transmission (β dans les modèles SIR)
- Ils sont relativement sophistiqués car les données dont on dispose ne sont pas des observations directes d'évènements de transmission mais les séries temporelles des effectifs des différentes classes épidémiologiques qui résultent des évènements de transmission

Exemple de modèle statistique. Modèles linéaires généralisé

- Décrit dans Velthuis et al. 2003
- Un modèle épidémiologique sous-jacent simple (modèle SI: pas de guérison)
- Terminologie:
  - β: paramètre de transmission
  - I: nombre d'infectieux,
  - N: nombre total d'individus

Exemple de modèle statistique. Modèles linéaires généralisé

- Pour un individu susceptible, le nombre de contacts infectieux lors d'un intervalle de temps Δt suit une loi de Poisson de paramètre $l = \beta (I/N) \Delta t$

- Loi de Poisson de paramètre l: $P(X = k) = \frac{l^k}{k!} e^{-l}$

- Ceci implique que la probabilité qu'un individu susceptible n'ait aucun contact infectieux dans l'intervalle Δt est : $P(C=0)=\exp(- \beta (I/N) \Delta t)$

- La probabilité d'avoir au moins un contact infectieux (et donc d'être infecté) est donc $1- \exp(- \beta (I/N) \Delta t)$

Exemple de modèle statistique. Modèles linéaires généralisé

- La probabilité qu'un individu susceptible soit infecté dans l'intervalle Δt est donc $1 - \exp(-\beta (I/N) \Delta t)$

- Le nombre d'individu passant de l'état susceptible à l'état infecté au cours d'un intervalle Δt peut alors être décrit comme un loi binomiale de paramètre:
  - probabilité: $1 - \exp(-\beta (I/N) \Delta t)$
  - nombre d'essais: S (nombre de susceptibles au début de l'intervalle)

- Il ne reste plus qu'à ajuster un modèle linéaire généralisé aux données

Exemple de modèle statistique. Modèles linéaires généralisé

- Il ne reste plus qu'à ajuster un modèle linéaire généralisé aux données
  - La variable réponse est le nombre de nouveaux cas
  - Une valeur observée de cette variable est disponible pour chaque intervalle de temps Δt
  - On utilise une distribution binomiale sous-jacente dont le paramètre « nombre d'essais » est le nombre de susceptibles au début de chaque intervalle de temps
  - Le paramètre de probabilité (p) de la distribution binomiale est le paramètre que l'on cherche à estimer
  - On utilise une fonction de lien complémentaire log-log
  - Et un terme d'offset égale à log(I/N)

Exemple de modèle statistique. Modèles linéaires généralisé

$$\log(-\log(1-p)) = a + \log(I/N)$$

exp(a) est un estimateur de β

$$\log(1-p) = -\exp(a) * I/N$$

$$\beta = \exp(a)$$

$$\log(1-p) = -\beta * I/N$$

$$1-p = \exp(-\beta * I/N)$$

$$p = 1 - \exp(-\beta * I/N)$$

Exemple de modèle statistique. Modèles linéaires généralisé

- Il suffit donc de construire le modèle linéaire généralisé

$$\log(-\log(1-p)) = a + \log(I/N)$$

dont le seul paramètre est a

- D'estimer a en ajustant ce modèle aux données (nombre de nouveaux cas, nombre d'infectieux, nombre total)

- Et de dériver une estimation de β en utilisant la relation

$$\beta = \exp(a)$$

- Le modèle fournira aussi un intervalle de confiance pour a dont les bornes pourront être utilisées pour obtenir un intervalle de confiance pour β

Logiciel

- Le modèle est un modèle statistique standard
- On peut par exemple utiliser R avec la fonction glm

- Velthuis, A., Bouma, A., Katsma, W., Nodelijk, G., De Jong, M., 2007. Design and analysis of small-scale transmission experiments with animals. Epidemiology and infection 135, 202-217.

- Velthuis, A., De Jong, M., Kamp, E., Stockhofe, N., Verheijden, J., 2003. Design and analysis of an *Actinobacillus pleuropneumoniae* transmission experiment. Preventive veterinary medicine 60, 53-68

Estimations de la force d'infection à partir de données de variation de la prévalence au cours de la vie d'une cohorte ou en fonction de l'âge

Ecology, 87(9), 2006, pp. 2356–2365
© 2006 by the Ecological Society of America

THE FITTING OF GENERAL FORCE-OF-INFECTION MODELS TO WILDLIFE DISEASE PREVALENCE DATA

DENNIS M. HEISEY,[1,4] DAMIEN O. JOLY,[2] AND FRANÇOIS MESSIER[3]

[1]*USGS-National Wildlife Health Center, Madison, Wisconsin 53711 USA*
[2]*Field Veterinary Program, Wildlife Conservation Society, Bronx, New York 10460 USA*
[3]*Department of Biology, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5E2 Canada*

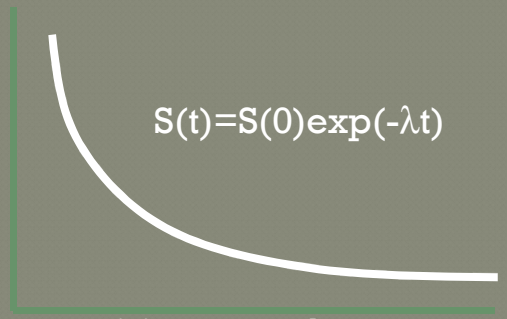
Conditions d'application

- Une situation endémique: la maladie a atteint un point d'équilibre dans la population d'hôtes: les paramètres qui gouvernent sa dynamique sont constants dans le temps

- Une mesure qui révèle une infection actuelle ou passée (par exemple la présence d'anticorps sous l'hypothèse que suite à une infection des anticorps peuvent être détectés tout au long de la vie d'un individu)

Logique sous-tendant le modèle

- Considérons une cohorte d'individus initialement susceptibles et suivons l'état de ces au cours du temps.
- Le nombre d'individus susceptibles (S) au sein de cette cohorte décroit de manière exponentielle avec le temps (t)



- λ est la force d'infection: c'est la probabilité instantanée de transition de l'état susceptible à l'état infecté
- Au début de la courbe, la décroissance est linéaire: au cours d'un intervalle de temps court δt le nombre de susceptible décroit de $\lambda\delta t$.
- Une proportion constante des susceptibles restants est infectée au cours d'un intervalle de temps d'une durée donnée.
- Cependant les susceptibles restants représentent un compartiment de plus en plus restreint à mesure que le temps passe, la proportion constante des susceptibles restants se traduit par une proportion décroissante de la cohorte.
- C'est ce qui explique la non linéarité de la décroissance du nombre de susceptibles

Sous forme d'équation différentielle

- Une proportion constante des susceptibles restants est infectée au cours d'un intervalle de temps d'une durée donnée.

$$\frac{dS(t)}{dt} = -\lambda S(t)$$

- Solution

$$S(t) = S(0)e^{-\lambda t}$$

- Cependant les susceptibles restants représentent un compartiment de plus en plus restreint à mesure que le temps passe, la proportion constante des susceptibles restants se traduit par une proportion décroissante de la cohorte.
- C'est ce qui explique la non linéarité de la décroissance du nombre de susceptibles

Ajustement d'un modèle statistique

- Décroissance du nombre de susceptible

$$S(t)=S(0)\exp(-\lambda t)$$

- En termes de pourcentage de la cohorte

$$S(t)/S(0)=\exp(-\lambda t)$$

- En termes d'incidence cumulée

$$I(t)/S(0)=1-\exp(-\lambda t)$$

- Si on dispose de données décrivant l'évolution du nombre de susceptibles dans le temps on peut facilement ajuster un modèle statistique à ces données pour estimer la force d'infection λ

$$\ln(S(t))-\ln(S(0)) = -\lambda t$$

$$\ln(S(t)) = -\lambda t + \ln(S(0))$$

Données

Offset

Paramètre à estimer

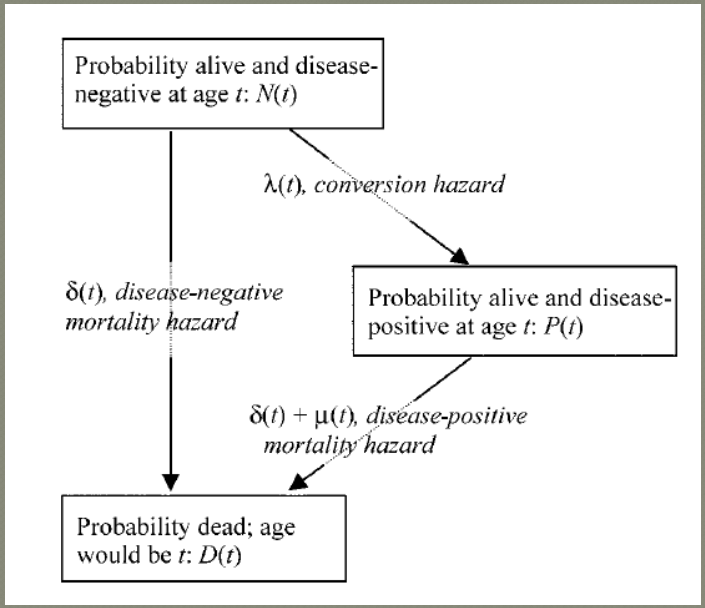
Prédicteur

Substitution du temps par l'âge

- Imaginons maintenant que nous menons une enquête transversale dans laquelle l'âge des individus se substitue au temps
- Au lieu de se focaliser sur l'évolution de l'incidence cumulée au cours du vieillissement d'une cohorte
- On s'intéresse aux variations de la prévalence (la plupart du temps sérologique) en fonction de l'âge d' individus prélevés au même moment

Problème pour l' application du modèle exponentiel

- Si la maladie entraîne une mortalité additionnelle, la prévalence sous-estime l'incidence cumulée et le biais augmente avec l'âge.



- D'autre part la force d'infection n'est pas forcément constante en fonction de l'âge
- Une fonction R associée à la publication permet d'ajuster des modèles qui tiennent compte de ces phénomènes

Estimations de paramètres de transmission à partir de données d'observation d'épidémies

Le problème

- Des données empiriques de dynamique épidémiologique
  - Typiquement, une série temporelle de nombre de cas
- Un (ou plusieurs) modèle(s) de dynamique épidémiologique décrit les processus qui sont supposés sous-tendre la dynamique épidémiologique
  - Typiquement un modèle compartimental de type SIR, SEIR
- Les valeurs d'un ou plusieurs paramètres de ce modèle sont inconnues
  - Par exemple paramètre de transmission, temps de latence, etc……..

Les objectifs de la modélisation statistique

- Obtenir des estimations de ces paramètres à partir de la modélisation statistiques des données empiriques

- Si plusieurs modèles de dynamiques épidémiologiques sont plausibles , déterminer lequel est plus compatible avec les données

Notations

- Y: les données (les observations)
- θ_M: l'ensemble des paramètres du modèle M (un terme par paramètre dans le modèle M)
- $P(Y|\theta_M)$: la probabilité de réalisation des observation Y pour le vecteur de valeurs des paramètres du modèle θ_M : vraissemblance
- $P(\theta_M)$: les distributions à priori des paramètres du modèle
- $P(\theta_M|Y)$: la distribution à postériori des paramètres du modèle M

Méthodes d'ajustement aux données

- Plusieurs méthodes (dont la mise en œuvre est souvent très technique) existent pour ajuster le modèle au données (estimer les valeurs inconnues des paramètres du modèle).
- Méthode du maximum de vraisemblance
- Méthode Bayésienne
- Méthode ABC (Approximate Bayesian Computation)

Méthodes du maximum de vraisemblance

- Ecrire la fonction de vraisemblance qui exprime la probabilité de réalisation de Y (les données) en fonction des paramètres du modèle de dynamique épidémiologique (θ_M)

- Cette fonction est généralement très complexe

- Pour un vecteur de valeurs des paramètres du modèle θ_M , cette fonction permet de calculer la vraisemblance $P(Y| \theta_M)$

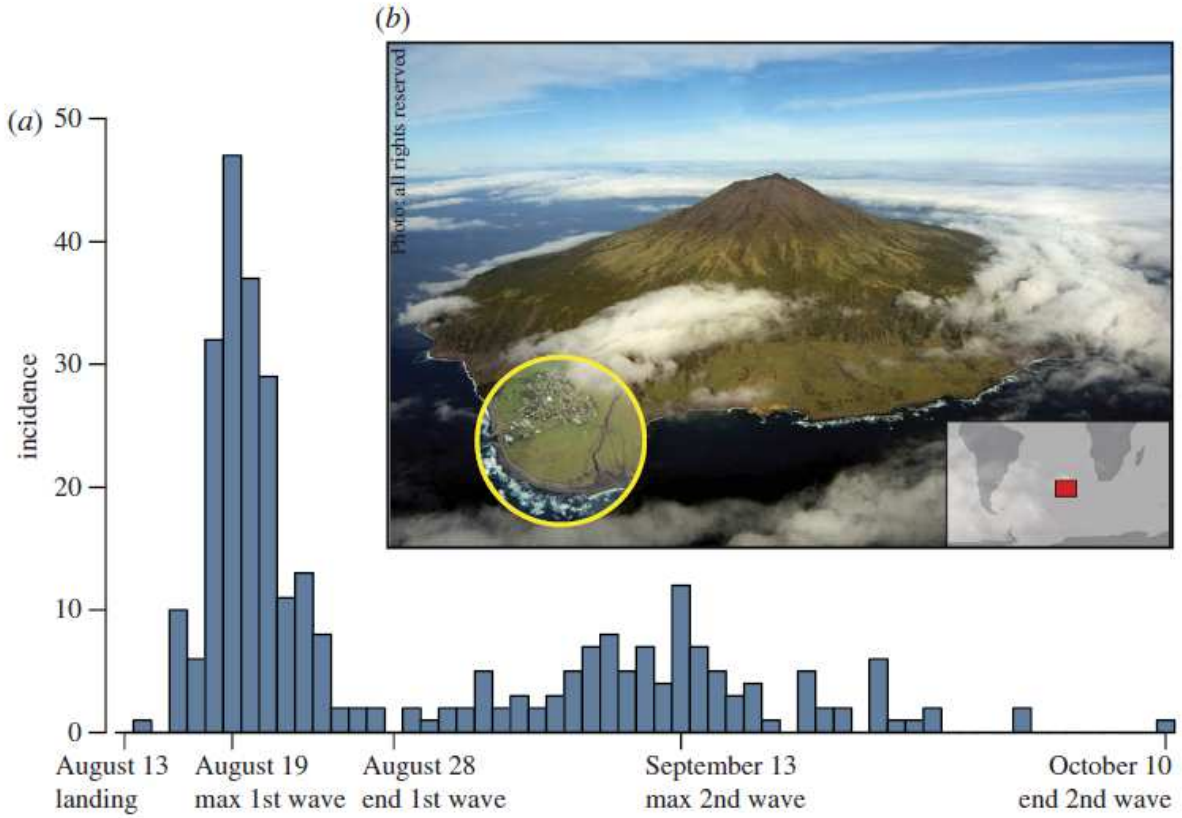
Méthodes du maximum de vraisemblance

- Objectif: déterminer les valeurs des paramètres du modèle qui maximisent la vraisemblance
- La fonction de vraisemblance étant généralement très complexe, cet objectif ne peut généralement pas être atteint avec des méthodes analytiques (dérivé première nulle et dérivée seconde négative)
- On utilise donc des algorithme d'exploration numérique de l'espace des paramètres
- A chaque itération, un vecteur de valeurs des paramètres est considéré et la vraisemblance est calculée pour ces valeurs des paramètres
- L'algorithme converge vers le point de maximum de vraisemblance. Les valeurs des paramètres en ce point sont les estimateurs du maximum de vraisemblance

Un exemple d'application de la méthode du maximum de vraisemblance

Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study

Anton Camacho[1,*], Sébastien Ballesteros[1,2], Andrea L. Graham[3], Fabrice Carrat[4,5,6], Oliver Ratmann[7,8] and Bernard Cazelles[1,9]

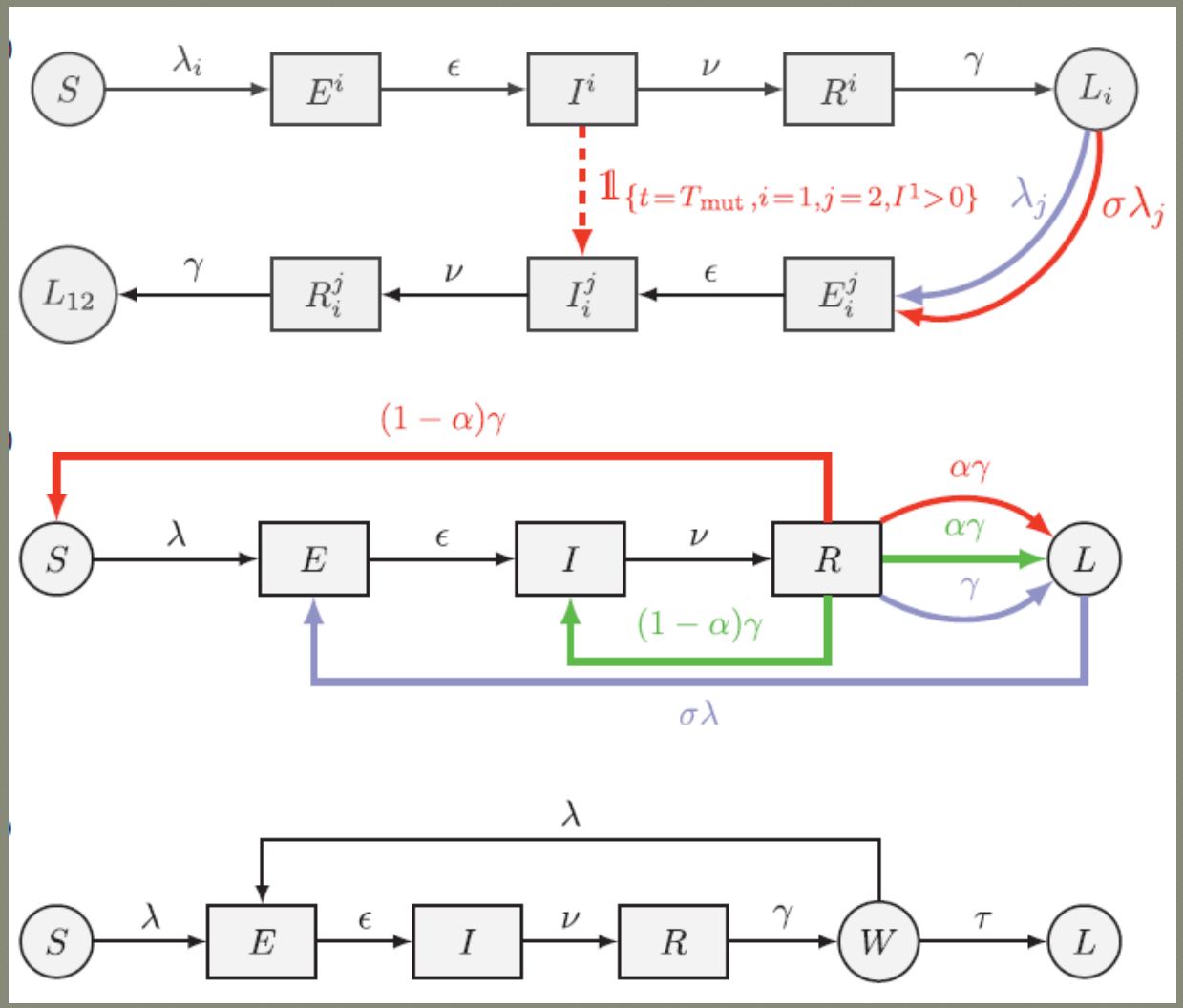


- Une épidémie de grippe sur l'île (très isolée) de Tristan Da Cuna présentant deux vagues épidémiques.
- Quels processus épidémiologiques peuvent expliquer ces deux vagues ?

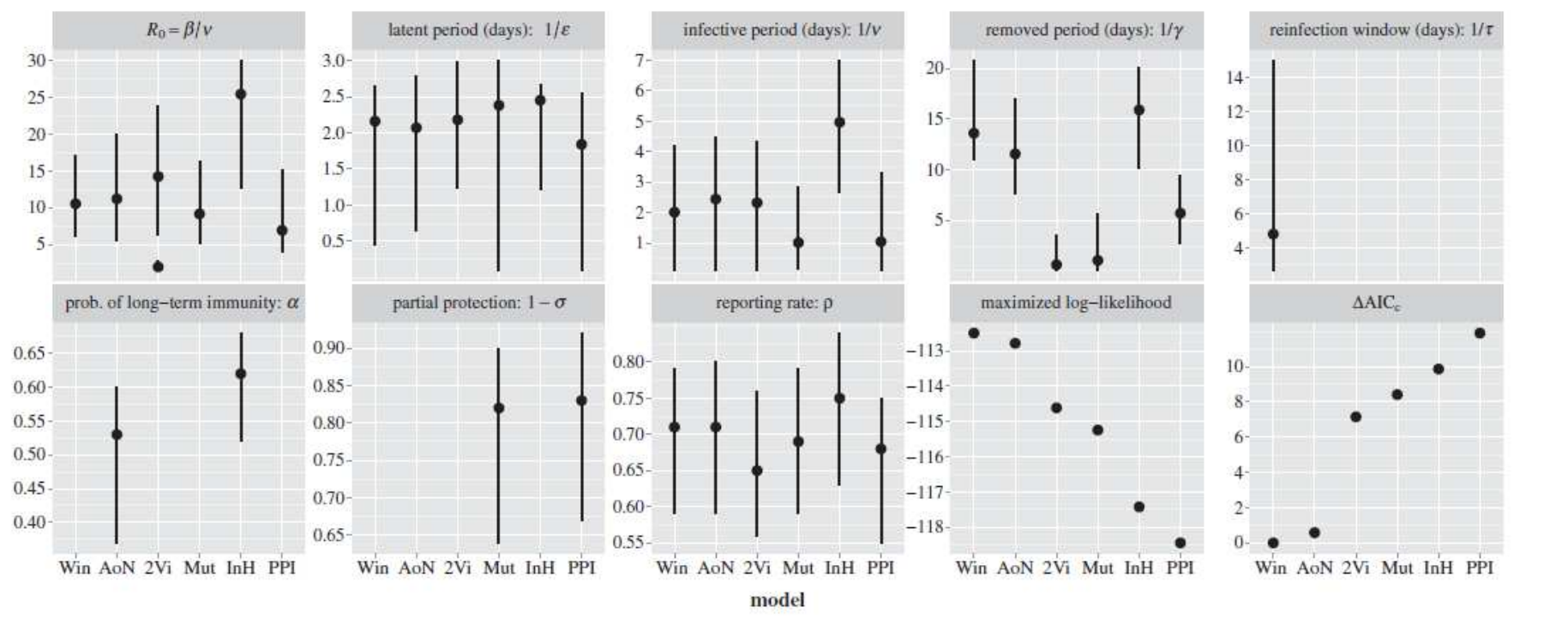
Les hypothèses susceptibles d'expliquer les deux vagues épidémiques

- 2 virus: deux souches distinctes ont été introduites
- Mutation: une seule source introduite qui a muté pour générer une nouvelle souche dotée de propriétés antigéniques différentes
- Tout ou rien: suite à la guérison des hôtes après la première vague, certains de ceux-ci n'ont pas développé une protection immunitaire durable et sont redevenus susceptibles à la même souche
- Protection partielle de tous les hôtes: suite à la première vague, tous les hôtes ont développé une protection partielle à la même souche
- Intra-hôte: certains hôte n'ont pas éliminé complètement la charge virale après la première infection. Les virus se sont à nouveau multiplié dans ces hôtes
- Fenêtre de réinfection: la protection à long terme ne se met pas en place dès la guérison: une réinfection par la même souche est possible

- Les modèles correspondants aux différentes hypothèses



- Les estimations de paramètres et les critères de sélection de modèle obtenus par maximum de vraisemblance



Logiciel

- Un logiciel performant pour le calcul numérique : par exemple MATLAB

Notations

- Y: les données (les observations)
- θ_M: l'ensemble des paramètres du modèle M (un terme par paramètre dans le modèle M)
- $P(Y|\theta_M)$: la probabilité de réalisation des observation Y pour le vecteur de valeurs des paramètres du modèle θ_M : vraissemblance
- $P(\theta_M)$: les distributions à priori des paramètres du modèle
- $P(\theta_M|Y)$: la distribution à postériori des paramètres du modèle M ou probabilité du modèle étant donné les observations

Les méthodes Bayésienne

- Reposent sur la formule de Bayes

$$P(\theta_M | Y) \propto P(Y | \theta_M)\, P(\theta_M)$$

- Tous les paramètres du modèle (θ_M) sont considérés comme des variables aléatoire

- $P(\theta_M)$: leurs distributions à priori représentent les connaissances que l'on possède sur les valeurs des paramètres avant d'analyser les données (issues d'autres études ou des connaissances générales de l'épidémiologie du pathogène)

- $P(Y | \theta_M)$ est la vraisemblance

Principe général

- Combiner les connaissances à priori à l'information contenue dans les données pour établir la distribution à posteriori des paramètres du modèle

$$P(\theta_M \mid Y) \propto P(Y \mid \theta_M)\, P(\theta_M)$$

- Procédure
  - Il faut malheureusement écrire la fonction de vraisemblance
  - Et déterminer les distributions à priori des paramètres
  - Puis appliquer la méthode MCMC (Monte Carlo Markov Chain) qui est une méthode d'échantillonnage des valeurs des paramètres dans les distributions à priori actualisées par la vraisemblance

MCMC: processus itératif

- Première itération:
  - Tirage d'une série de valeurs des paramètres suivant les densité de probabilités des distributions à priori
  - Calcul de la vraisemblance pour ces valeurs des paramètres
  - Actualisation des distributions à priori par la vraisemblance

- Seconde itération
  - Tirage d'une série de valeurs des paramètres suivant les densité de probabilités des distributions à priori actualisées
  - et ainsi de suite

- Le processus permet de converger vers les distributions à posteriori des paramètres (au bout de 100 000- 1 000 000 d'itérations)

MCMC

- Le processus permet de converger vers les distributions à posteriori des paramètres

$$P(\theta_M \mid Y) \propto P(Y \mid \theta_M)\, P(\theta_M)$$

- Les distributions à posteriori combinent les informations disponibles avant l'analyse des données et l'information contenue dans les données pour produire des estimations des paramètres du modèle sous forme de distribution de densité de probabilité.

Exemple d'application des méthodes bayésiennes

- Cauchemez et al. 2004
- Analyse de données de suivi longitudinal d'infection de foyers par la grippe saisonnière
- Modèle SIR hiérarchisé (intra-foyer, inter-foyer) et structuré en âge (enfants, adultes)
- Estimation des paramètres de transmission et des durées des périodes infectieuses

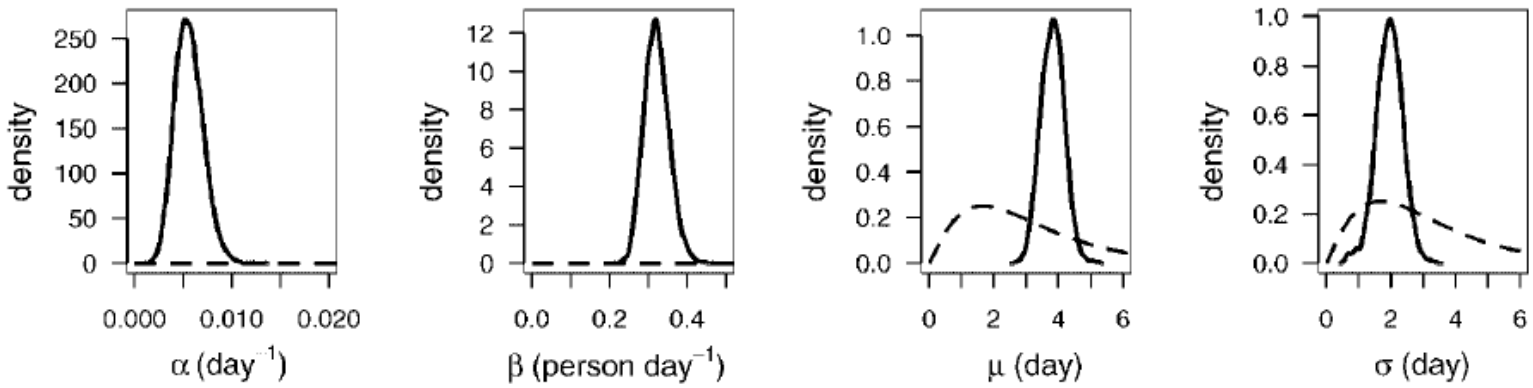


Figure 2. MCMC output, *prior* and *posterior* distributions for the community risk of infection (α), the household risk of infection (β), the mean duration of infectious period (μ) and the standard deviation of infectious period (σ). Dotted lines represent *priors*.

Logiciel

- Un logiciel dédié aux analyses Bayésiennes : par exemple WINBUGS, OPENBUGS ou JAGS

- Blum, M.G., Tran, V.C., 2010. HIV with contact tracing: a case study in approximate Bayesian computation. Biostatistics 11, 644-660.

- Camacho, A., Ballesteros, S., Graham, A.L., Carrat, F., Ratmann, O., Cazelles, B., 2011. Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. Proceedings of the Royal Society B: Biological Sciences 278, 3635-3643.

- Cauchemez, S., Carrat, F., Viboud, C., Valleron, A., Boelle, P., 2004. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. Statistics in medicine 23, 3469-3487.

- LaDeau, S.L., Glass, G.E., Hobbs, N.T., Latimer, A., Ostfeld, R.S., 2011. Data-model fusion to better understand emerging pathogens and improve infectious disease forecasting. Ecological Applications 21, 1443-1460.

Un exemple d'application de la méthode ABC

Preventive Veterinary Medicine 116 (2014) 380–390

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Inferences about the transmission of Schmallenberg virus within and between farms

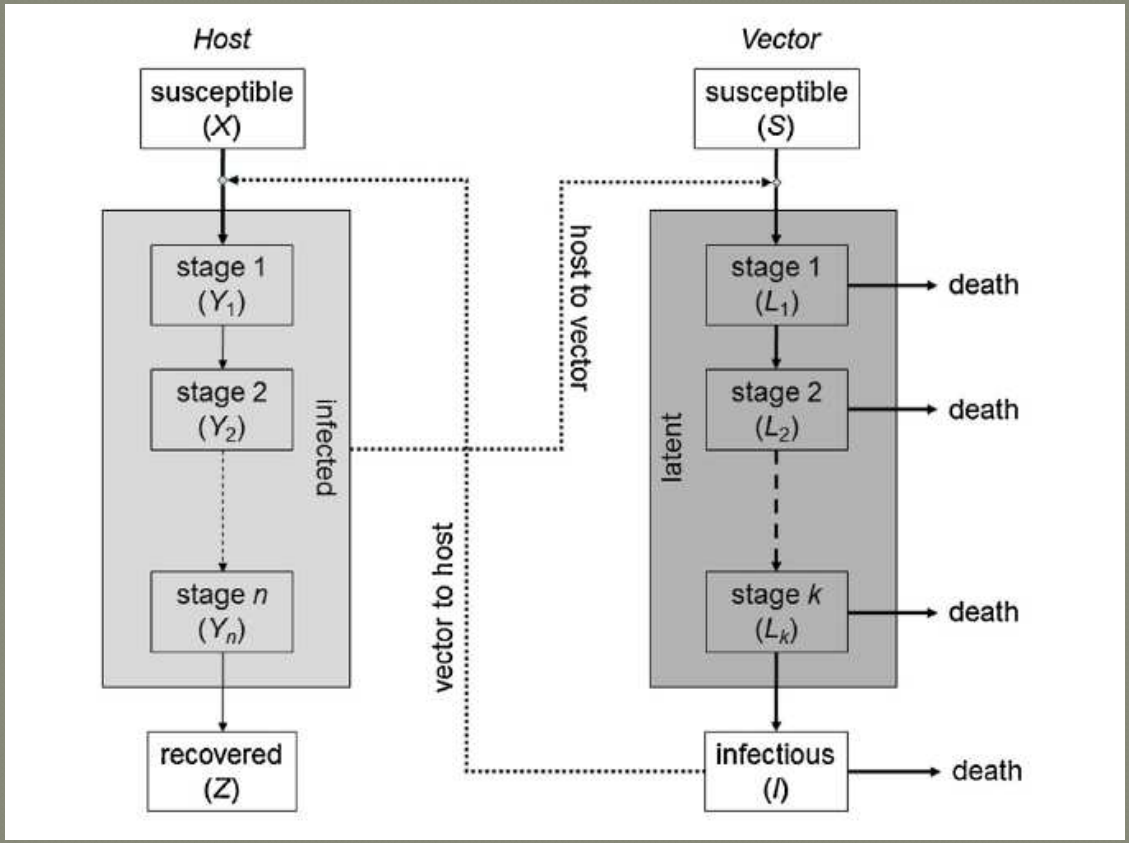
Simon Gubbins[a,*], Joanne Turner[b], Matthew Baylis[b], Yves van der Stede[c,d], Gerdien van Schaik[e], José Cortiñas Abrahantes[f], Anthony J. Wilson[a]

Objectif

- Estimer les paramètres qui gouvernent la dynamique épidémiologique intra-troupeau du virus Schmallenberg
- A partir de données de prévalence sérologique intra-troupeau provenant d'élevages bovins et ovins en Belgique (422 et 82) et aux Pays-bas (543, 342).
- Pour chaque élevage les informations suivantes sont considérées:
  - Nombre total d'animaux dans l'élevage
  - Nombre d'animaux échantillonnés
  - Région
  - Date de prélèvement

Modèle épidémiologique

- Maladie virale à transmission vectorielle



Modèle épidémiologique

Table 1
Transitions, probabilities and population sizes in the transmission model for SBV.

Description	Transition	Probability	Population size
Hosts			
Infection	$\begin{cases} X^{(i)} \to X^{(i)} - 1 \\ Y_1^{(i)} \to Y_1^{(i)} + 1 \end{cases}$	$\lambda_i \delta t$	$X^{(i)}$
Completion of infection stage j ($j = 1,\ldots,n_i - 1$)	$\begin{cases} Y_j^{(i)} \to Y_j^{(i)} - 1 \\ Y_{j+1}^{(i)} \to Y_{j+1}^{(i)} + 1 \end{cases}$	$n_i r_i \delta t$	$Y_j^{(i)}$
Recovery	$\begin{cases} Y_{n_i}^{(i)} \to Y_{n_i}^{(i)} - 1 \\ Z^{(i)} \to Z^{(i)} + 1 \end{cases}$	$n_i r_i \delta t$	$Y_{n_i}^{(i)}$
Vectors			
Infection	$\begin{cases} S \to S - 1 \\ L_1 \to L_1 + 1 \end{cases}$	$\lambda_v \delta t$	S
Completion of extrinsic incubation period (EIP), stage j ($j = 1,\ldots,k - 1$)	$\begin{cases} L_j \to L_j - 1 \\ L_{j+1} \to L_{j+1} + 1 \end{cases}$	$k\nu\delta t$	L_j
Vector mortality (during EIP) ($g = 1,\ldots,k$) (and compensatory recruitment)	$\begin{cases} L_j \to L_j - 1 \\ S \to S + 1 \end{cases}$	$\mu\delta t$	L_j
Completion of EIP	$\begin{cases} L_k \to L_k - 1 \\ I \to I + 1 \end{cases}$	$k\nu\delta t$	L_k
Vector mortality (infectious vectors) (and compensatory recruitment)	$\begin{cases} I \to I - 1 \\ S \to S + 1 \end{cases}$	$\mu\delta t$	I

Table 2
Parameters in the model for the within-farm transmission of Schmallenberg virus (SBV).

Description	Symbol	Prior distribution, estimate or function	Comments
Inferred using seroprevalence data			
Probability of transmission from vector to host	b	Beta(7.38,2.13)	–
Probability of transmission from host to vector	β	Beta(6.60,28.75)	–
Vector to host ratio for species i	m_i	Triangular(0,1000,5000)[a]	Based on a maximum host biting rate ($m_C a$) of 2500 bites per host per day (Gerry et al., 2001); cf. light trap catches of up to 10,000 midges per trap night (Meiswinkel et al., 2008)
Number of animals of species i	H_i	–	Obtained from the serosurveillance data-sets
Duration of viraemia (cattle)			
Mean (days)	$1/r_C$	Log normal(1.40,0.40)	–
No. stages	n_C	Uniform(1,20)	Constrained to take only integer values
Duration of viraemia (sheep)			
Mean (days)	$1/r_S$	Log normal(1.40,0.40)	–
No. stages	n_S	Uniform(1,20)	Constrained to take only integer values
Extrinsic incubation period (EIP)			
Mean (days)	$1/\nu$	$\nu(T) = \max(0, \alpha(T - T_{\min}))$	Reciprocal of mean EIP depends on temperature (T) (cf. Carpenter et al., 2011)
No. stages	k	log normal(2.66,0.77)	Constrained to take only integer values
Virus replication rate above threshold	α	Normal(0.019,0.010)	Used to compute reciprocal of mean EIP ($1/\nu$)
Threshold temperature for virus replication	$T_{\min}$	Normal(13.34,1.09)	
Fixed (not inferred from seroprevalence data)			
Reciprocal of the time interval between blood meals	a	$a(T) = 0.0002T(T - 3.7)(41.9 - T)^{1/2.7}$	Depends on temperature (T) (Mullens et al., 2004)
Vector mortality rate	μ	$\mu(T) = 0.009\exp(0.16T)$	Depends on temperature (T) (Gerry and Mullens, 2000)
Vector recruitment rate	ρ	–	For simplicity, assumed to be equal to the vector mortality rate
Vector population size	N	–	For simplicity, assumed to be constant; given by $N = m_i H_i$
Vector activity			
Intercept	b_0	−1.71	
Sin, 12 month period	b_{11}	−1.56	Based on analysis of data from a network of 12 suction traps in England (Sanders et al., 2011)
Cos, 12 month period	b_{21}	−3.74	
Sin, 6 month period	b_{12}	−1.49	
Cos, 6 month period	b_{22}	−1.00	

[a] Parameters for the triangular distribution are minimum, mode and maximum.

Modèle d'observation

- Echantillonnage
  - Pour chaque élevage, le nombre de séropositifs dans l'échantillon est supposé suivre une loi hypergéométrique de paramètres
    - Nombre d'individus infectés ou rétablis
    - Nombre total d'individus dans l'élevage
    - Taille de l'échantillon
- Test
  - Le test utilisé est jugé très performant: sensibilité et spécificité =1

Démarche d'ajustement du modèle aux données

- Echantillonner un jeu de paramètres dans les distributions à priori
- Simuler des données D en utilisant le modèle de dynamique épidémiologique et d'observation
  - Une date d'introduction du virus est tirée au hasard pour chaque élevage dans une distribution uniforme dont les bornes sont
    - deux semaines avant les premiers cas enregistrés en Belgique et aux Pays
    - la date d'échantillonnage
- Evaluer la similarité entre les données simulées (D) et les données observées D_{obs} à l'aide d'une métrique $M(D, D_{obs})$ mesurant la distance entre les données simulées et les données observées
- Si $M(D,D_{obs}) <$ seuil, le jeu de paramètres utilisé pour générer les données est retenu, sinon il est rejeté
- On recommence

La métrique de distance

- La comparaison porte sur la distribution des prévalences intra-troupeaux (histogramme de fréquence des prévalences exprimées en catégories de 5%: 0-5%; 5-10%,….,95-100%)

$$M = \sum_{i,c} \sum_{j} p_j^{(i,c)} \log \left(\frac{p_j^{(i,c)}}{q_j^{(i,c)}} \right) + q_j^{(i,c)} \log \left(\frac{q_j^{(i,c)}}{p_j^{(i,c)}} \right),$$

  - $p_j^{(i,c)}$: fréquence observée de la catégorie de prévalence j pour l'espèce i dans le pays c
  - $q_j^{(i,c)}$: fréquence simulée de la catégorie de prévalence j pour l'espèce i dans le pays c
  - Seuil pour M déterminé de manière à ce que le taux d'acceptation soit dans l'intervalle 0.5-1%

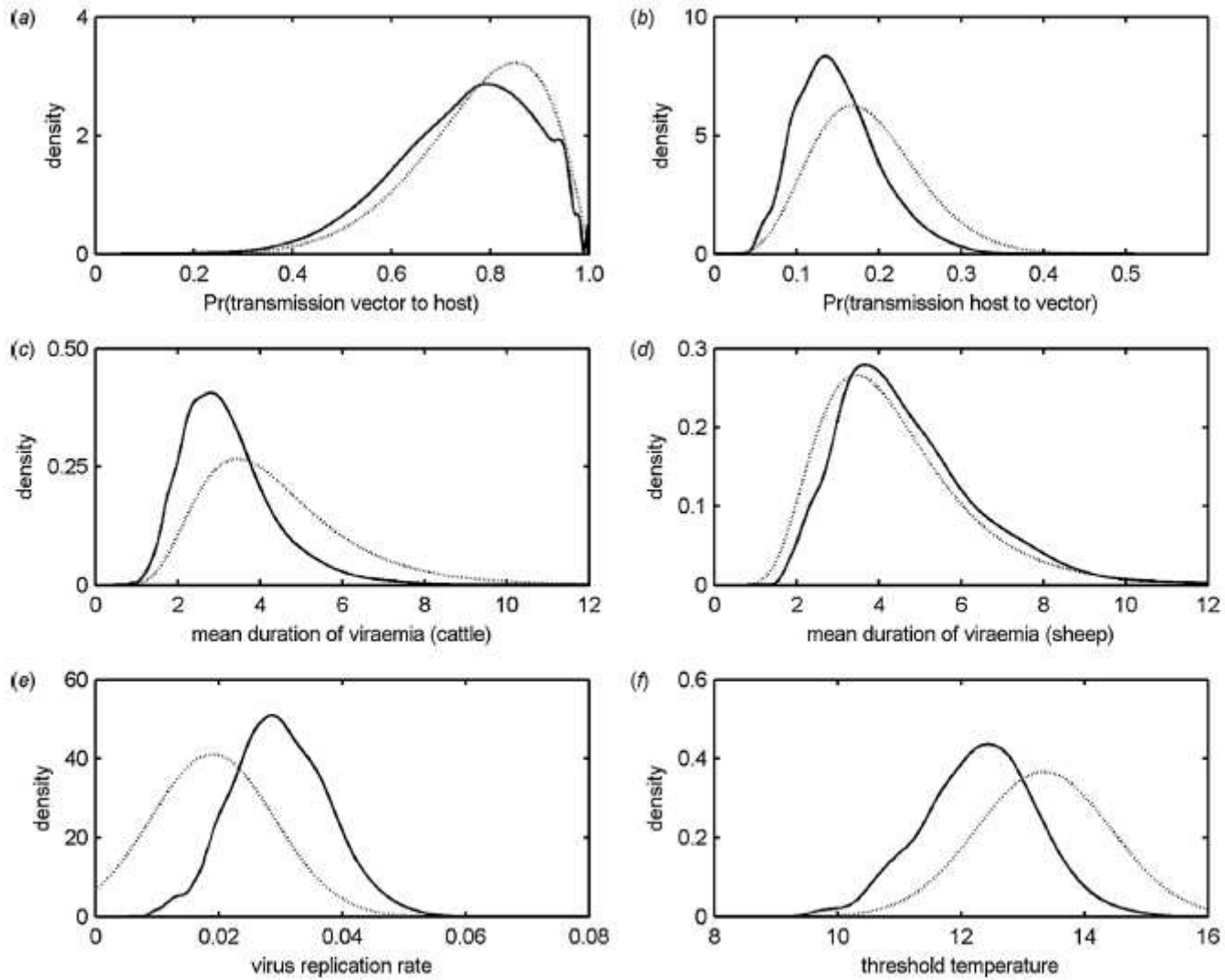


Fig. 3. Marginal posterior distributions for epidemiological parameters for Schmallenberg virus (SBV): (a) probability of transmission from vector to host; (b) probability of transmission from host to vector; (c, d) mean duration of viraemia (days) in (c) cattle or (d) sheep; (e) virus replication rate; and (f) threshold temperature (°C) for virus replication. Each figure shows the prior (dotted black line; Table 2) and posterior (solid black line) densities when the model for the within-farm transmission of SBV was fitted to seroprevalence data for cattle and sheep from Belgium and The Netherlands.

Table 3
Posterior median and 95% credible intervals (CI) for parameters in the model for the within-farm transmission of Schmallenberg virus (SBV).

Parameter	Median	95% CI
Probability of transmission		
Vector to host	0.76	(0.46, 0.95)
Host to vector	0.14	(0.07, 0.26)
Duration of viraemia (cattle)		
Mean (days)	3.04	(1.63, 5.91)
No. stages	11	(1, 20)
Duration of viraemia (sheep)		
Mean (days)	4.37	(2.24, 9.02)
No. stages	11	(1, 20)
Extrinsic incubation period		
Virus replication rate	0.030	(0.016, 0.045)
Threshold temperature	12.35	(10.52, 14.02)
No. stages	6	(2, 35)

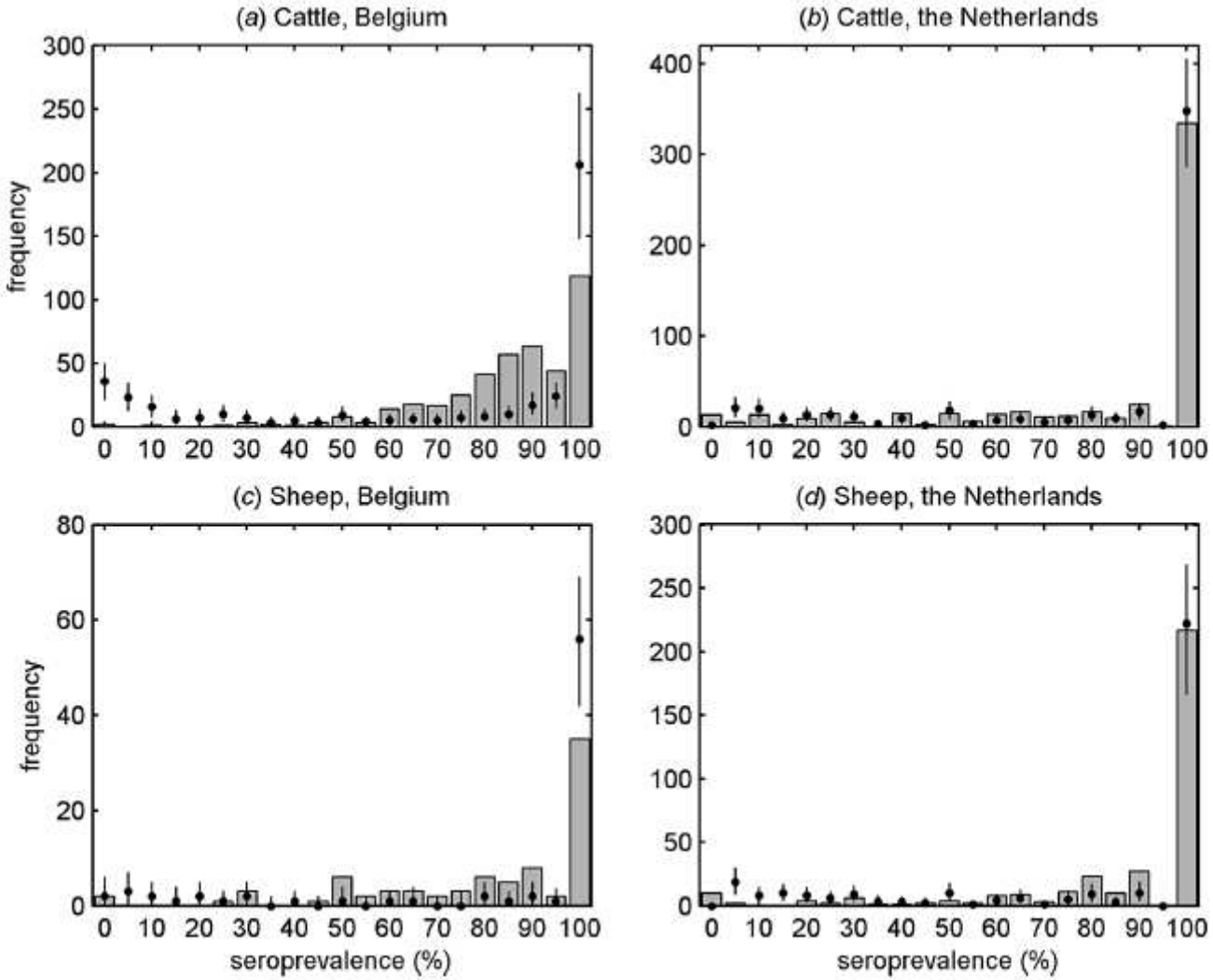


Fig. 2. Observed (bars) and expected (median (circles) and 95% prediction intervals (error bars)) distribution of within-farm seroprevalence of Schmallenberg virus in (a, b) cattle and (c, d) sheep from (a, c) Belgium and (b, d) The Netherlands.

Estimations de paramètres de transition entre états à partir de données de suivi longitudinal à l'échelle individuelle