Surveillance analysis Tool for Outcome-based Comparison of the confidence of FREEdom

European Food Safety Authority

The STOC free model

Objectives, hypotheses and model overview

Aurélien Madouasse

&

the STOC free consortium

https://www.stocfree.eu/

June 18, 2021

Table of Contents

Table of Contents

Infectious diseases of cattle

Regulated diseases

- Public health threats

  e.g. Tuberculosis

- Economic impact

  e.g. Foot and mouth disease

- Legislation on how to perform surveillance in order to substantiate freedom from disease

- Every country performs surveillance in the same way $\rightarrow$ comparable output

$\Rightarrow$ **Input-based surveillance**

Infectious diseases of cattle

Non-regulated diseases

- A lot of important infectious diseases are not regulated but have regional / national control programmes in place

  e.g. BVD, paratuberculosis ...

  - No *legal* prescription on the way to perform surveillance
  - Important diversity in the design of surveillance programmes
  - The *free status* in one programme can have a different meaning than the *free status* in another programme

$\Rightarrow$ Creates difficulties when trading animals between herds enrolled in different programmes

$\Rightarrow$ **Output-based surveillance**: production of an output that is comparable regardless of the input surveillance data

A method for output-based surveillance

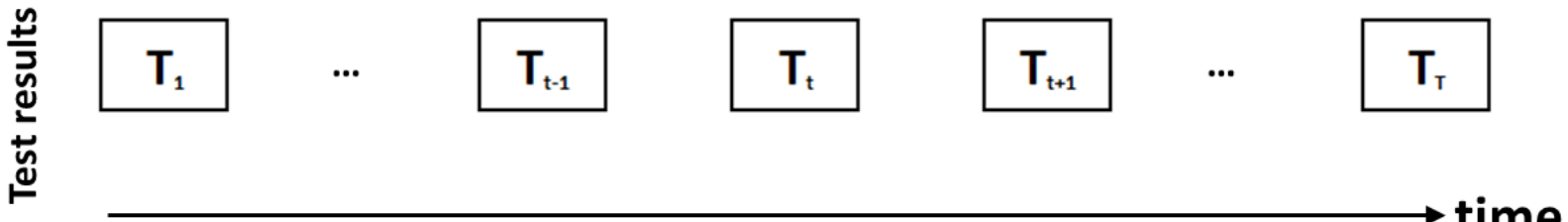
- Need for a method taking inputs from diverse surveillance programmes able to produce an output that is comparable
  - Structure of the method/model determined by what is common across surveillance programmes
  - Probability of freedom from infection estimated in a given programme from:
    - surveillance data available in the programme
    - relevant knowledge (e.g. test characteristics)

Control programmes

Common features

- *Control* programmes against cattle non-regulated diseases: prevalence > 0
- Objective: disease control or eradication
- Organised at regional or country level
- Rely on a surveillance component for the identification of infected herds or animals
- Detection of infection followed by control phase
- Surveillance performed in all participating herds
- Herd tested repeatedly over time

  $\Rightarrow$ Longitudinal data



Control programmes

Common features

- Tests are imperfect
  - Sensitivity: $Se = p(T^+|D^+) < 1$
  - Specificity: $Sp = p(T^-|D^-) < 1$

$\Rightarrow$ Uncertainty in the true status of tested animals / herds

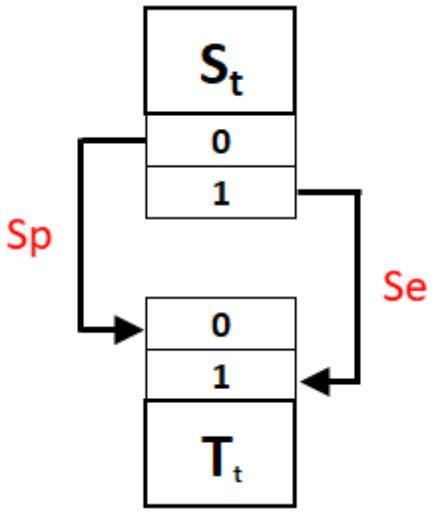


Table of Contents

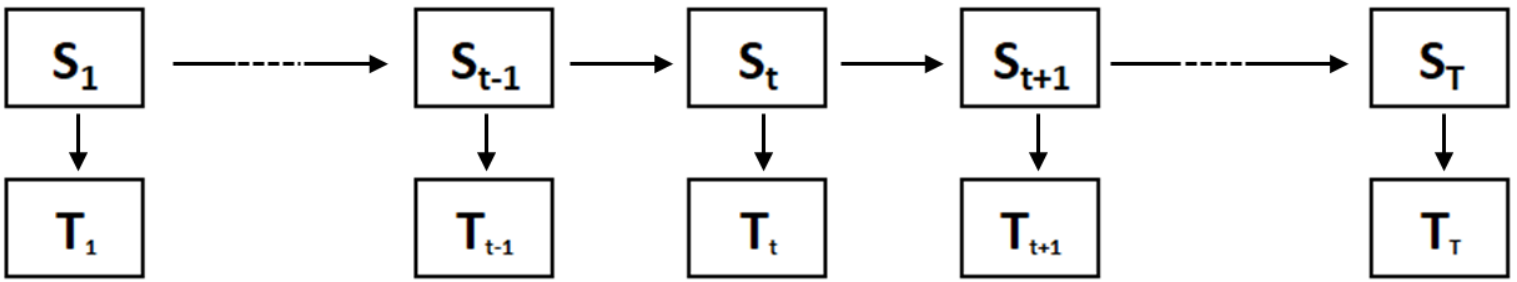
Modelling objectives

- Objective: Predict herd level probabilities of *(freedom from)* infection from longitudinal test data collected as part of surveillance programmes against endemic infectious diseases of cattle
- The modelling framework should:
  - Allow the use of longitudinal data i.e. account for the fact that sequences of test results are not random
  - Account for imperfect test information

Hidden Markov models

- Hidden Markov Models (**HMM**s) model a latent discrete variable with a Markovian dynamics, whose state at a given time determines the distribution of an observed variable
  - **discrete variable**: the variable of interest can be in 1 of k states. In the STOC free model, $k = 2$ (positive or negative status)
  - **latent variable**: this discrete variable is not directly observed. In the STOC free model, the **latent status** determines the probability of a negative or positive test result through test sensitivity and specificity
  - **Markovian dynamics**: the latent status is modelled in discrete time steps. The latent status at time t only depends on the status at time $t - 1$. The probabilities of transition between the k states between 2 time points are described by a k x k transition matrix.

Representation of surveillance programmes as HMMs



- S_t: latent status of interest at time t , from $t = 1$ to $t = T$
- T_t: test result at time t

Representation of surveillance programmes as HMMs

Status dynamics

- Time steps of equal duration

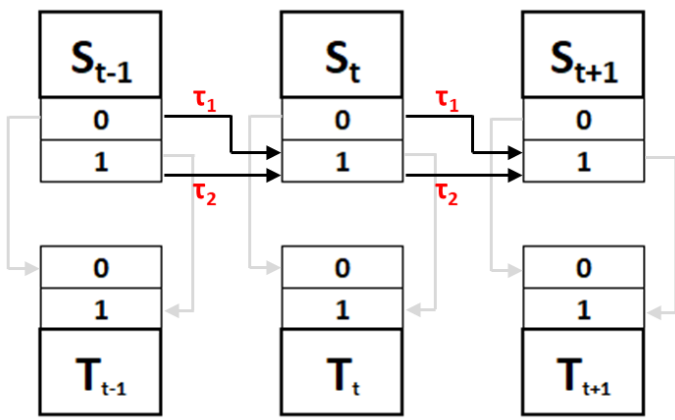
  (Discrete-time model)

- Herd status at time t depends on herd status at time t_1

  (Markovian property)

$$S_t \sim Bernoulli(\pi_t)$$

$$\pi_t = \begin{cases} \tau_1 & \text{if } S_{t-1} = 0 \\ \tau_2 & \text{if } S_{t-1} = 1 \end{cases}$$

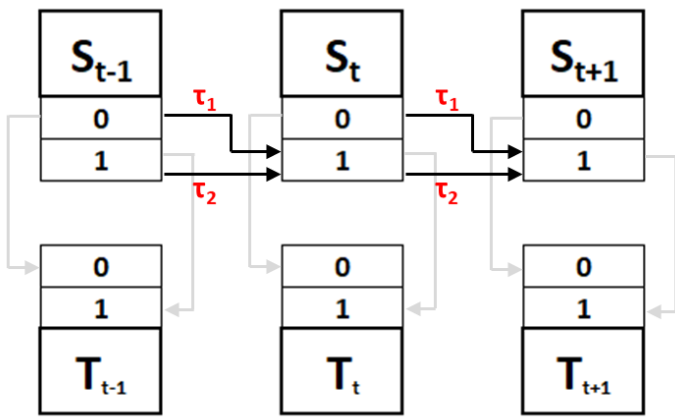


Representation of surveillance programmes as HMMs

Status dynamics

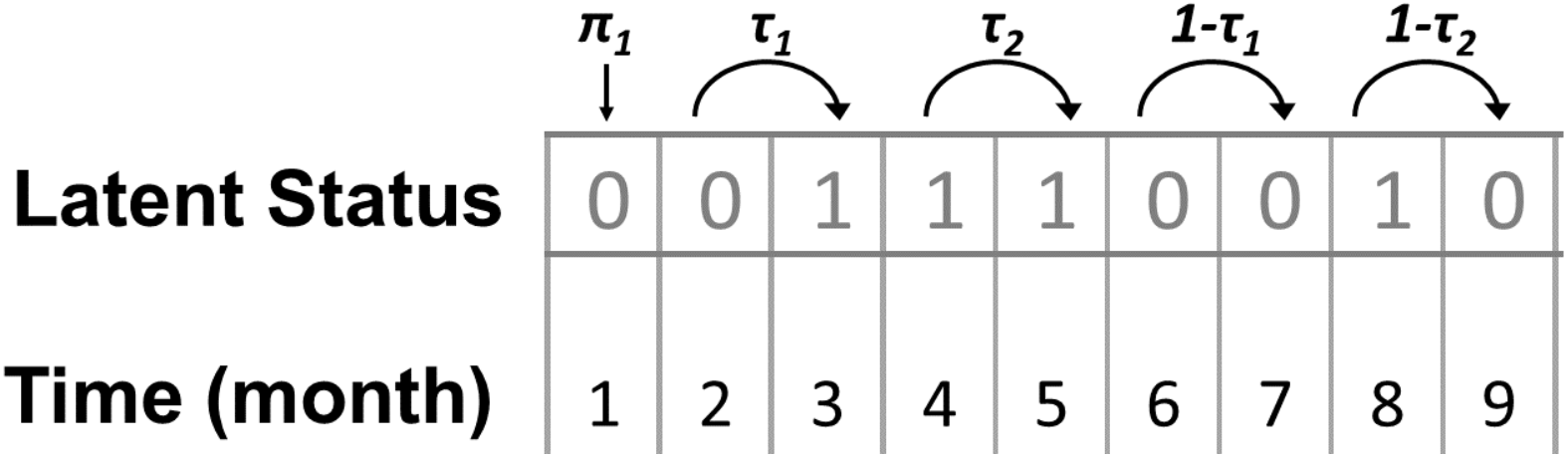
- Herd level probability of new infection (τ_{1t}) modelled as a function of one or several risk factors (X_t) using logistic regression

$$ln(\frac{\tau_{1t}}{1-\tau_{1t}}) = X_t\theta$$



Representation of surveillance programmes as HMMs

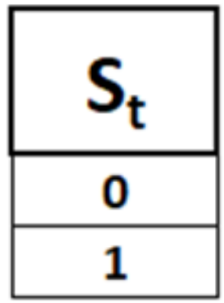
Status dynamics



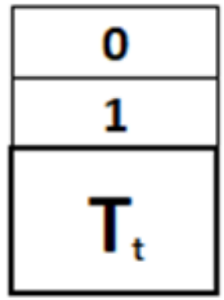
Representation of surveillance programmes as HMMs

Test results

- Herd status at time t is either negative (0) OR positive (1)



- Test result at time t is either negative (0) OR positive (1)

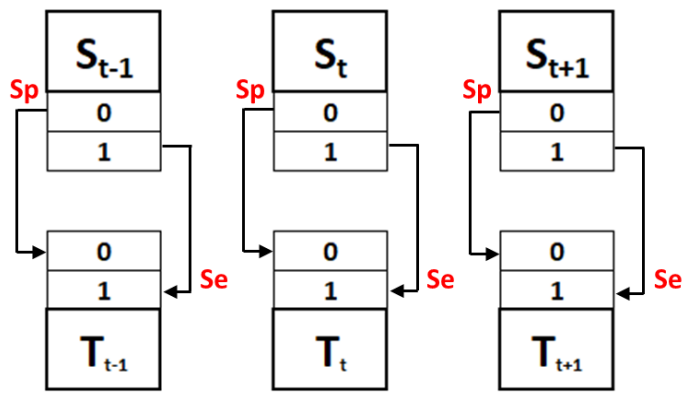


Representation of surveillance programmes as HMMs

Test results

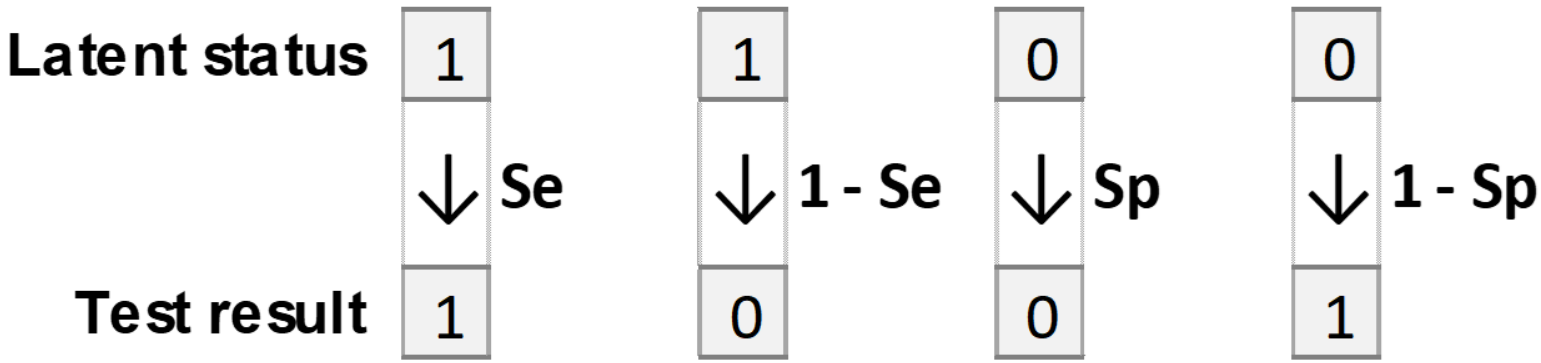
- Test results at the herd level
- Test results depend on the latent status through sensitivity and specificity

$$p(T_t = 1) = \begin{cases} 1 - Sp & \text{if } S_t = 0 \\ Se & \text{if } S_t = 1 \end{cases}$$



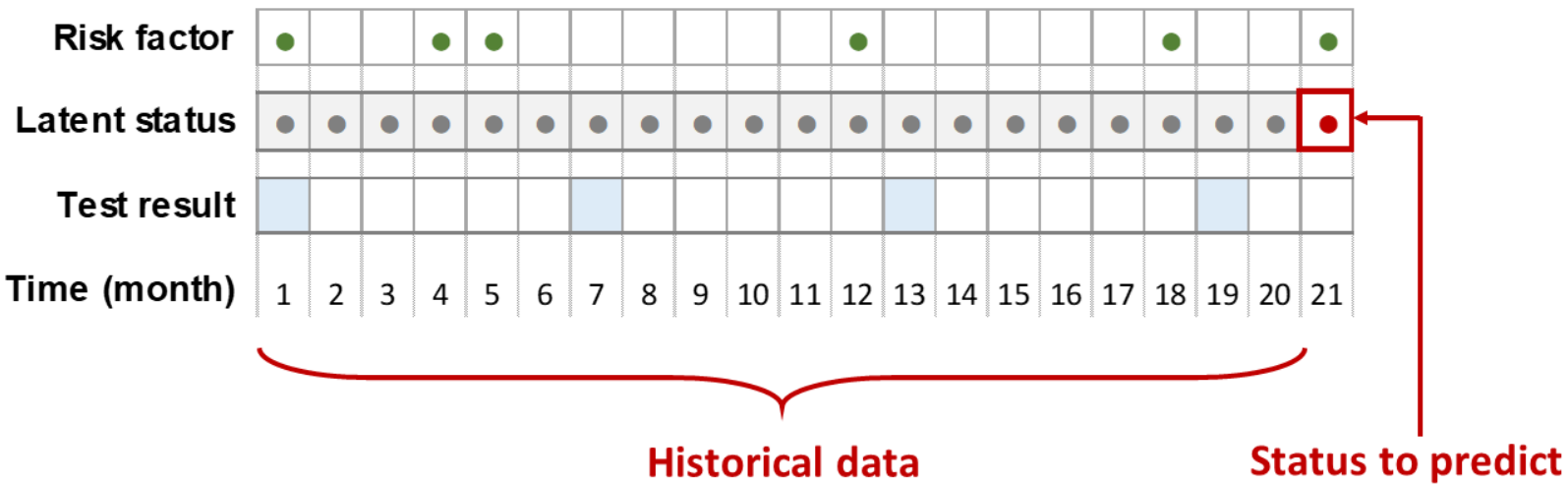
Representation of surveillance programmes as HMMs

Test results



Predictions

- The model predicts a herd level probability of being status positive at time T
- All the data available up to T are used for parameter estimation



Summary of the different model parameters

- π_1: probability of being status positive on the first test month
- τ_1: probability of becmoing status positive between 2 months
- $\theta_1, \theta_2 \ldots$: coefficients of the logistic regression for the probability of becoming status positive
- τ_2: probability of remaining status positive between 2 months
- Se: herd-level test sensitivity
- Sp: herd-level test specificity

The complete model - no risk factor

$$S_1 \sim Bernoulli(\pi_1)$$

$$S_t \sim Bernoulli(\pi_t) \quad \forall t > 1$$

$$\pi_t = \begin{cases} \tau_1 & \text{if } S_{t-1} = 0 \\ \tau_2 & \text{if } S_{t-1} = 1 \end{cases}$$

$$T_t \sim Bernoulli(p(T_t))$$

$$p(T_t) = \begin{cases} 1 - Sp & \text{if } S_t = 0 \\ Se & \text{if } S_t = 1 \end{cases}$$

The complete model - with risk factors

$$S_1 \sim Bernoulli(\pi_1)$$

$$S_t \sim Bernoulli(\pi_t) \quad \forall t > 1$$

$$\pi_t = \begin{cases} \tau_{1t} & \text{if } S_{t-1} = 0 \\ \tau_2 & \text{if } S_{t-1} = 1 \end{cases}$$

$$ln(\frac{\tau_{1t}}{1-\tau_{1t}}) = X_{ht}\theta$$

$$T_t \sim Bernoulli(p(T_t))$$

$$p(T_t) = \begin{cases} 1 - Sp & \text{if } S_t = 0 \\ Se & \text{if } S_t = 1 \end{cases}$$

Table of Contents

Model implementation

- Parameter estimation and prediction carried out in a Bayesian framework
  - The framework permits the incorporation of available knowledge using prior distributions
  - Two different programmes can be used to run the model: JAGS or Stan

The `STOCfree` R package

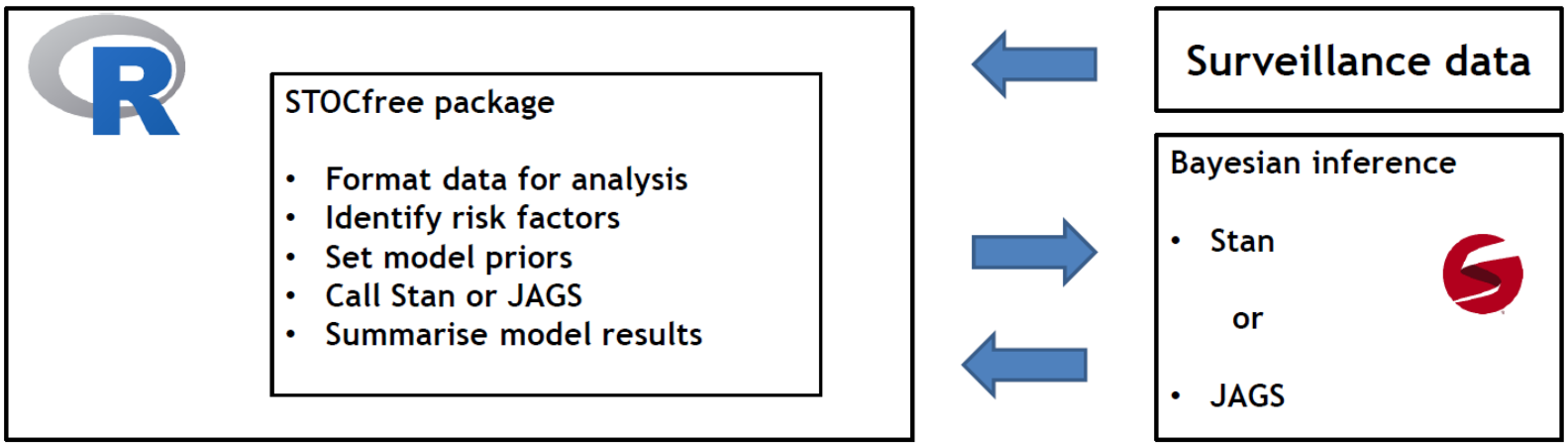
What is an R package?

- Programming environment for data manipulation and analysis
- Widely used
- Free

 package

- Set of functions gathered to perform specific tasks
- Users install a package and can use the functions they contain
- Packages are installed from the web (CRAN, GitHub...)

The STOCfree R package

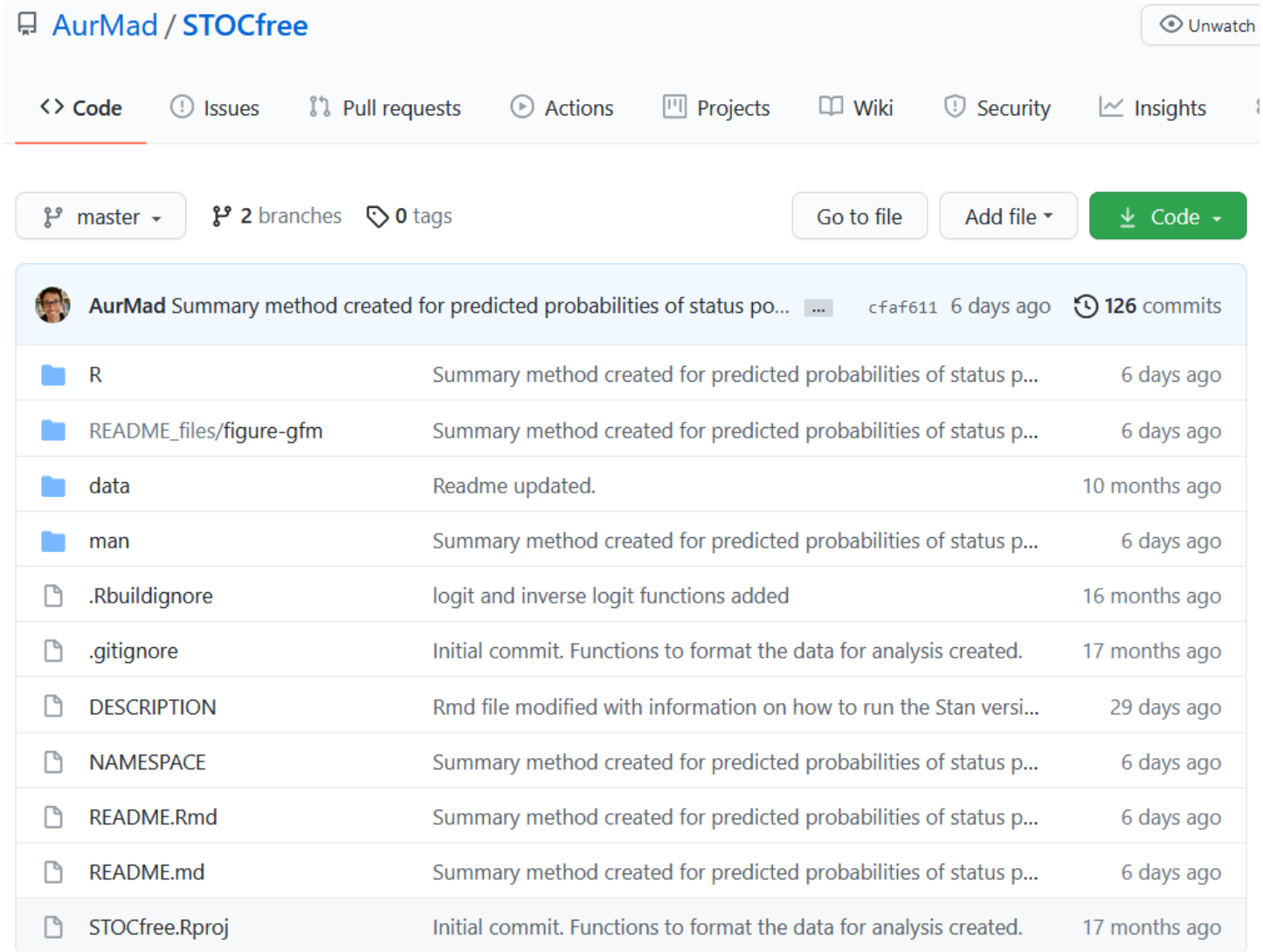


The STOCfree R package on Github

- The package is hosted on Github
  https://github.com/AurMad/STOCfree
- Github is a server hosting:
  - The package code
  - The package documentation
  - The history of development and different package versions, using the Git versioning programme

The STOCfree R package on Github

- All the code is in the `R` folder



The STOCfree R package on Github

- The documentation is at the bottom of the page

README.md

STOCfree: prediction of probabilities of freedom from infection from longitudinal data

Overview

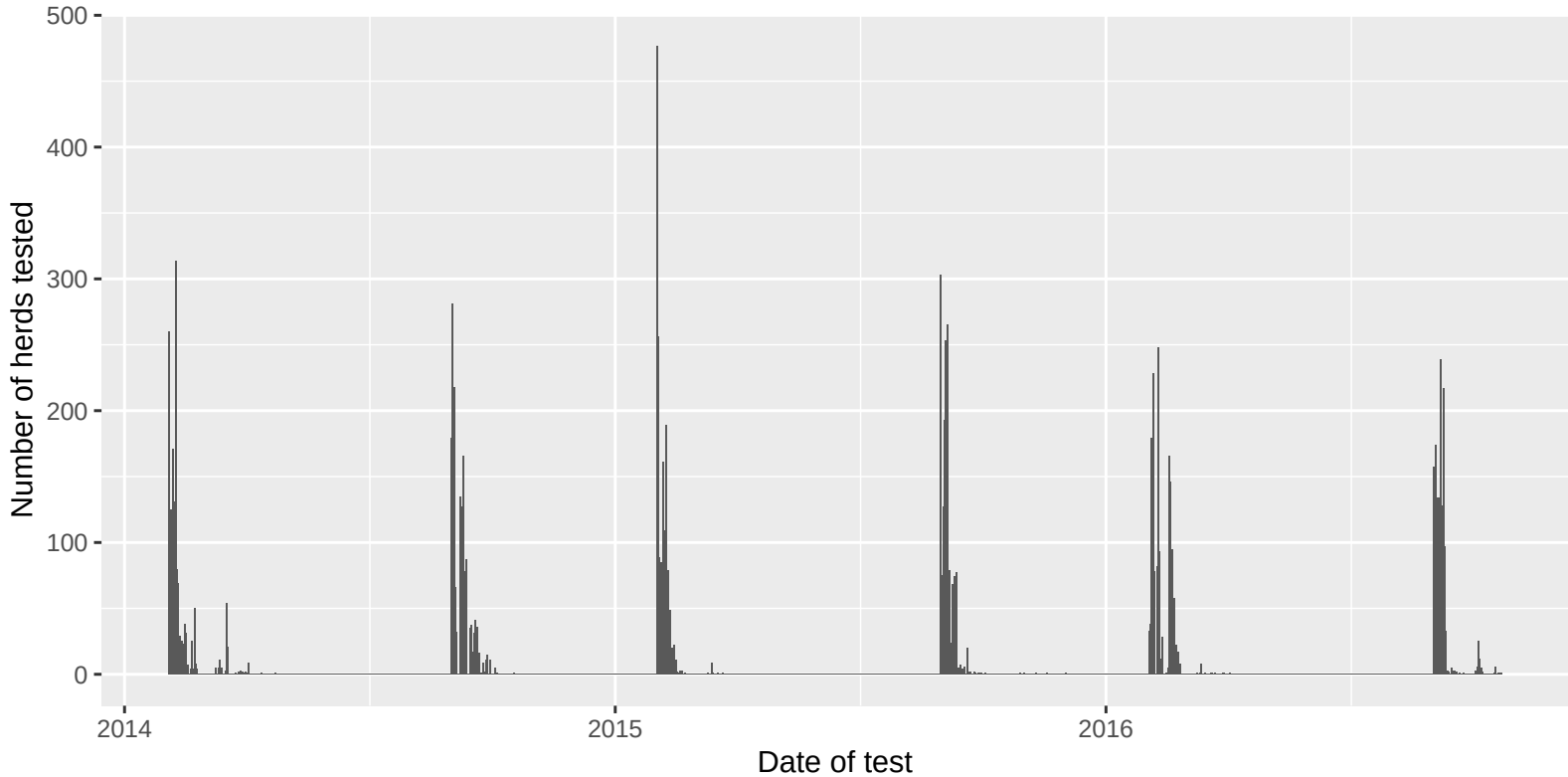
The aim of the `STOCfree` package is to predict herd level probabilities of freedom from infection from

Table of Contents

Data

- Surveillance data collected as part of a dairy cattle BVD control programme in Loire-Atlantique (France)
  - All herds tested every 6 months
  - Antibody ELISA on bulk tank milk
  - Data from 1 687 herds between 2014 and 2017

Data


500
400
300
200
100
0
Number of herds tested
2014
2015
2016
Date of test

Models

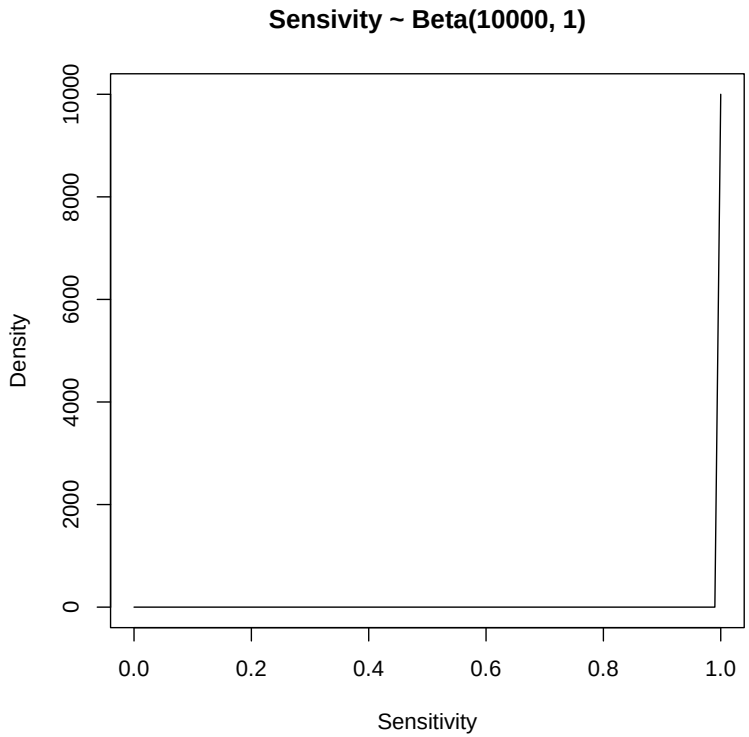
- Four different models incorporating different hypotheses were run in both Stan and JAGS and the results compared
  - **Model 1:** Perfect test, no risk factors
  - **Model 2:** Imperfect test, no risk factors
  - **Model 3:** Perfect test, risk factors
  - **Model 4:** Imperfect test, risk factors

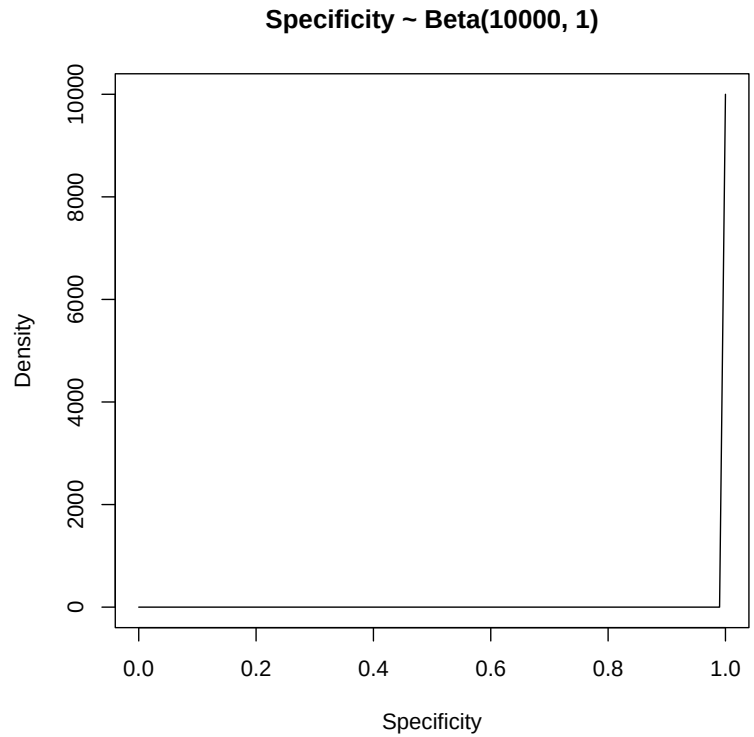
Models

- Four different models incorporating different hypotheses were run in both Stan and JAGS and the results compared
  - **Model 1:** Perfect test, no risk factors
  - **Model 2:** Imperfect test, no risk factors
  - **Model 3:** Perfect test, risk factors
  - **Model 4:** Imperfect test, risk factors
- Risk factors
  - Two considered: number of cattle introduced, local seroprevalence
  - Only number of cattle introduced retained in the final model

Prior distributions

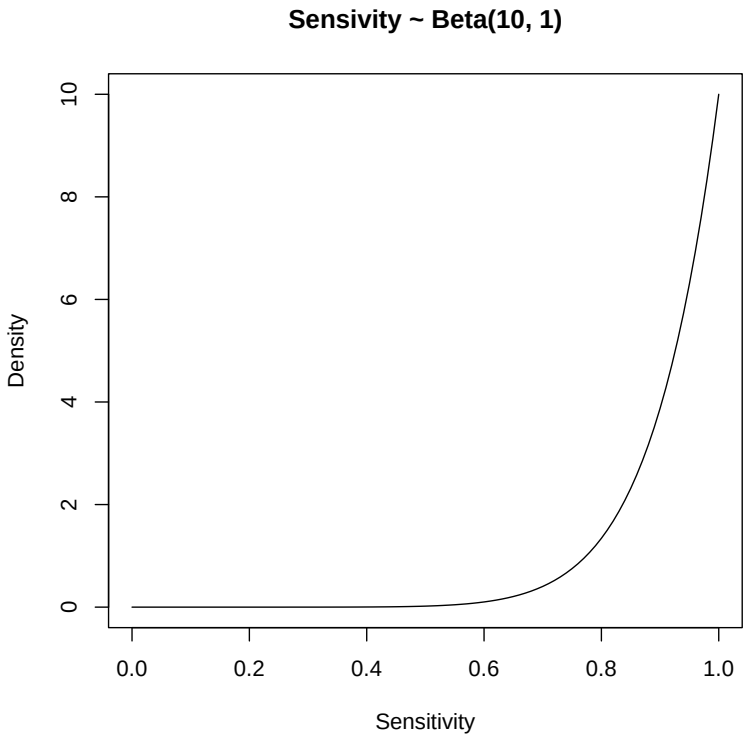
Perfect test

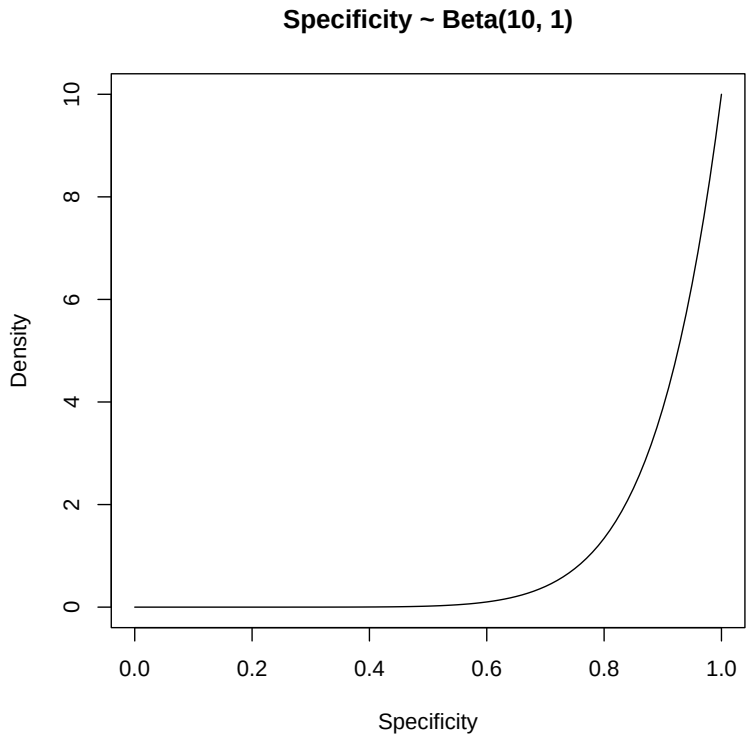




Prior distributions

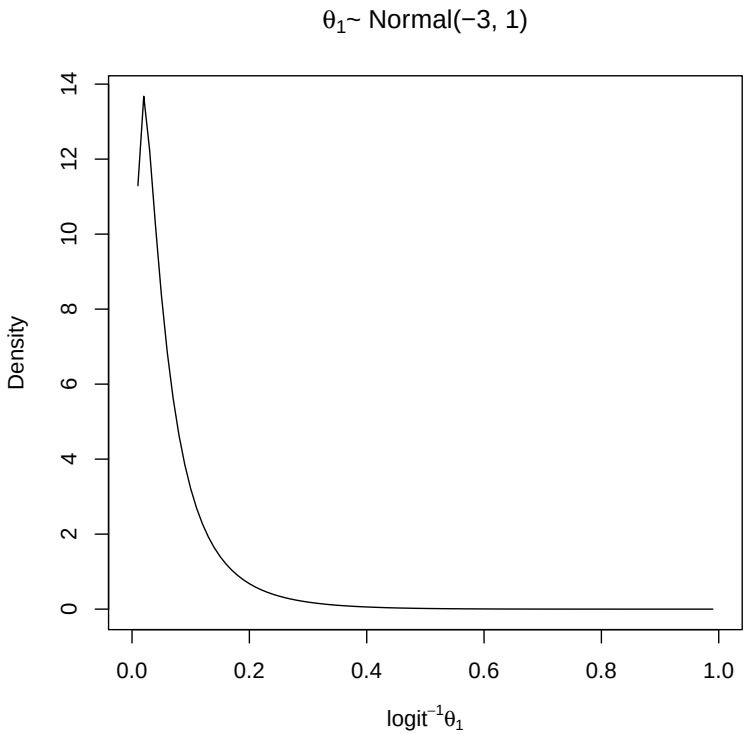
Imperfect test

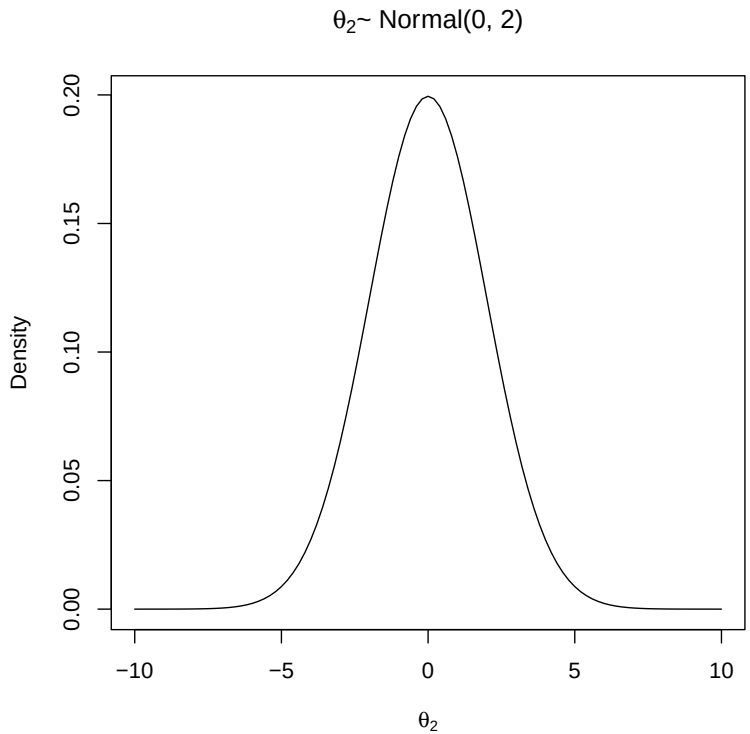




Prior distributions

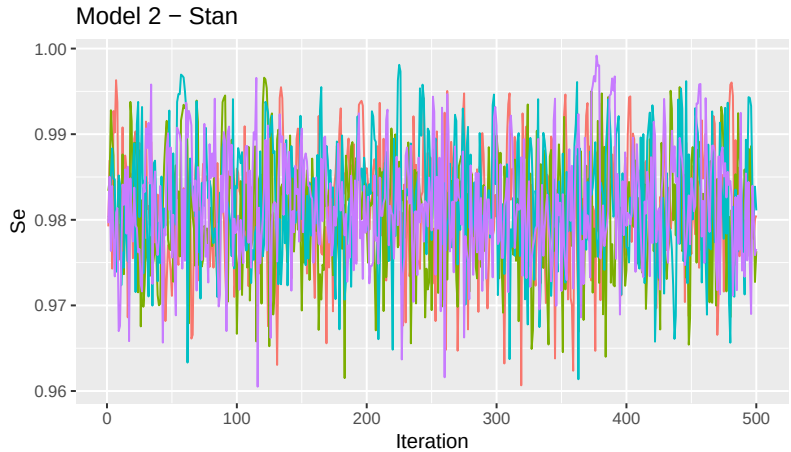
Risk factors: intercept and coefficient

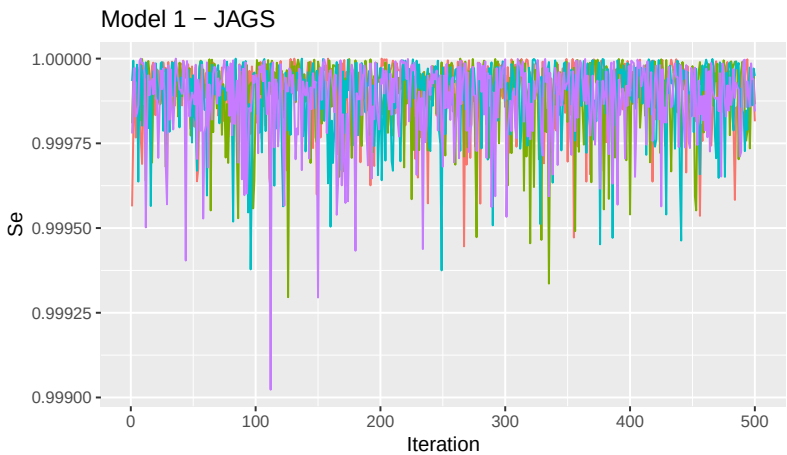


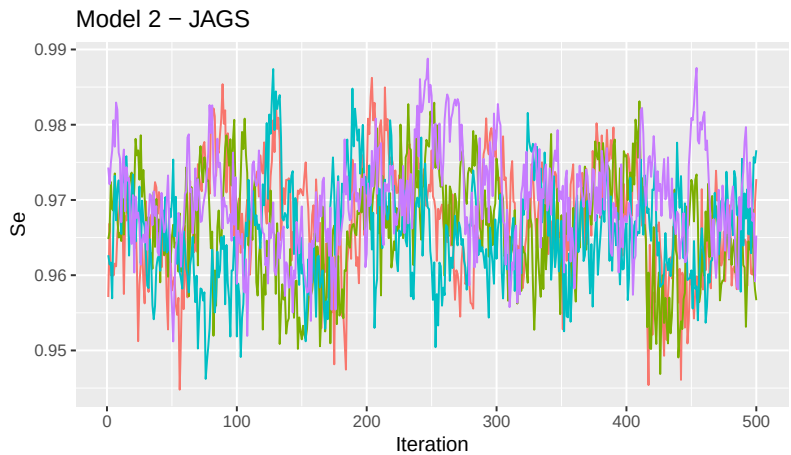


Model convergence

- Convergence much better with the Stan version of the model, especially with imperfect test

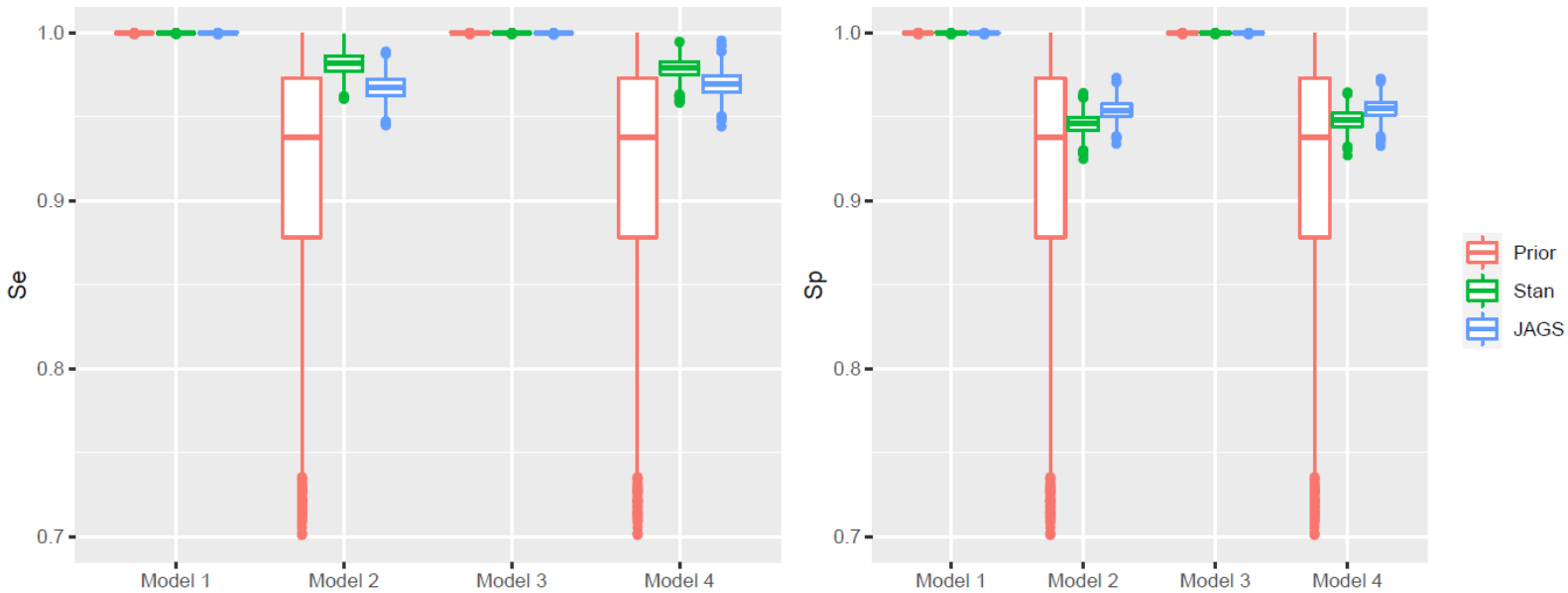






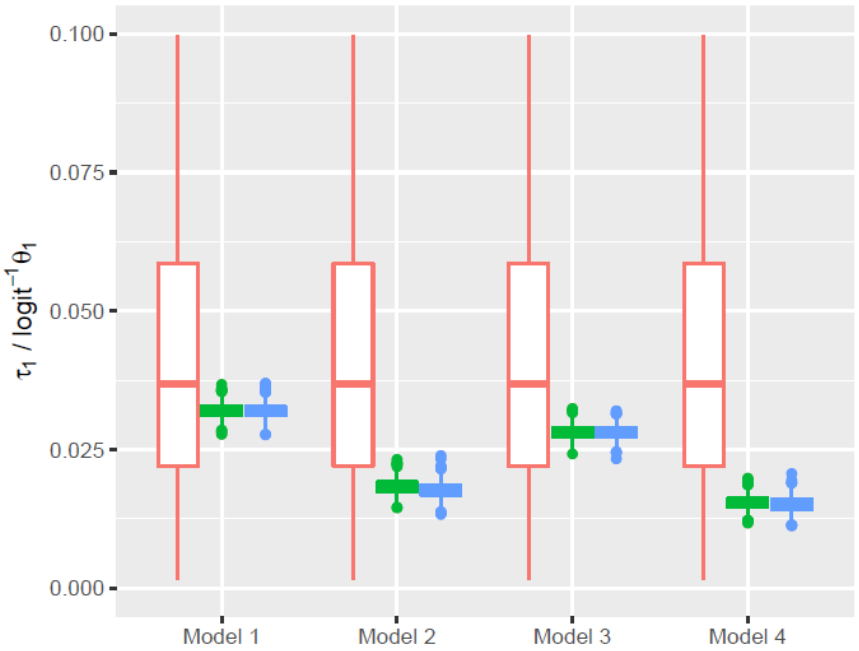
Parameter estimates

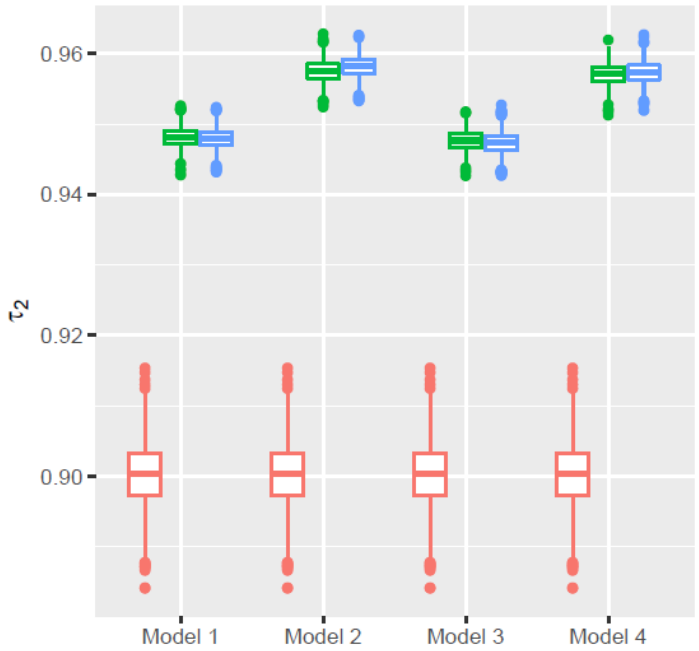
Test characteristics



Parameter estimates

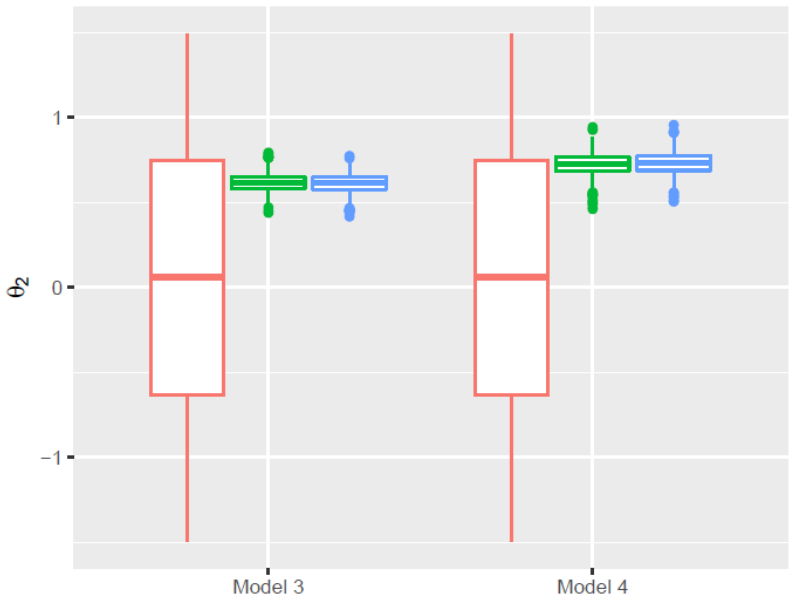
Infection dynamics

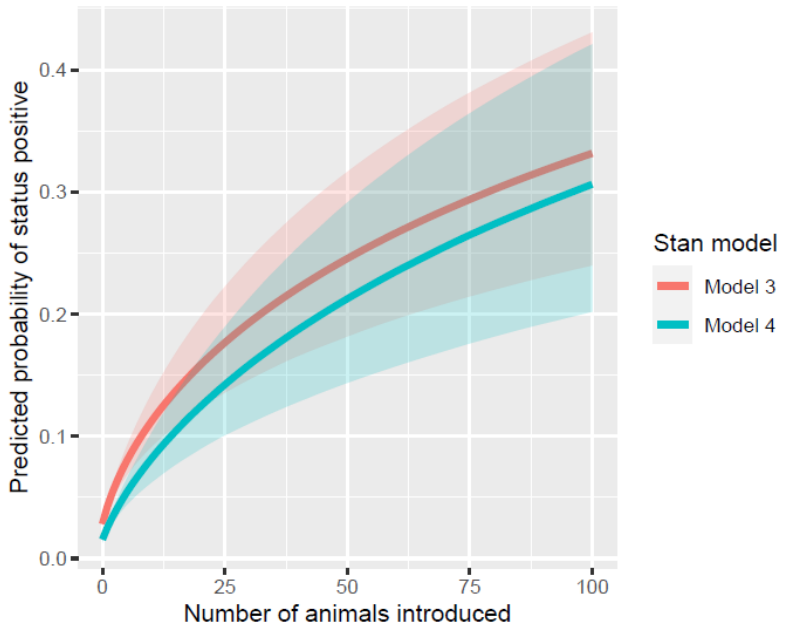




Parameter estimates

Risk factors of new infection





Parameter estimates

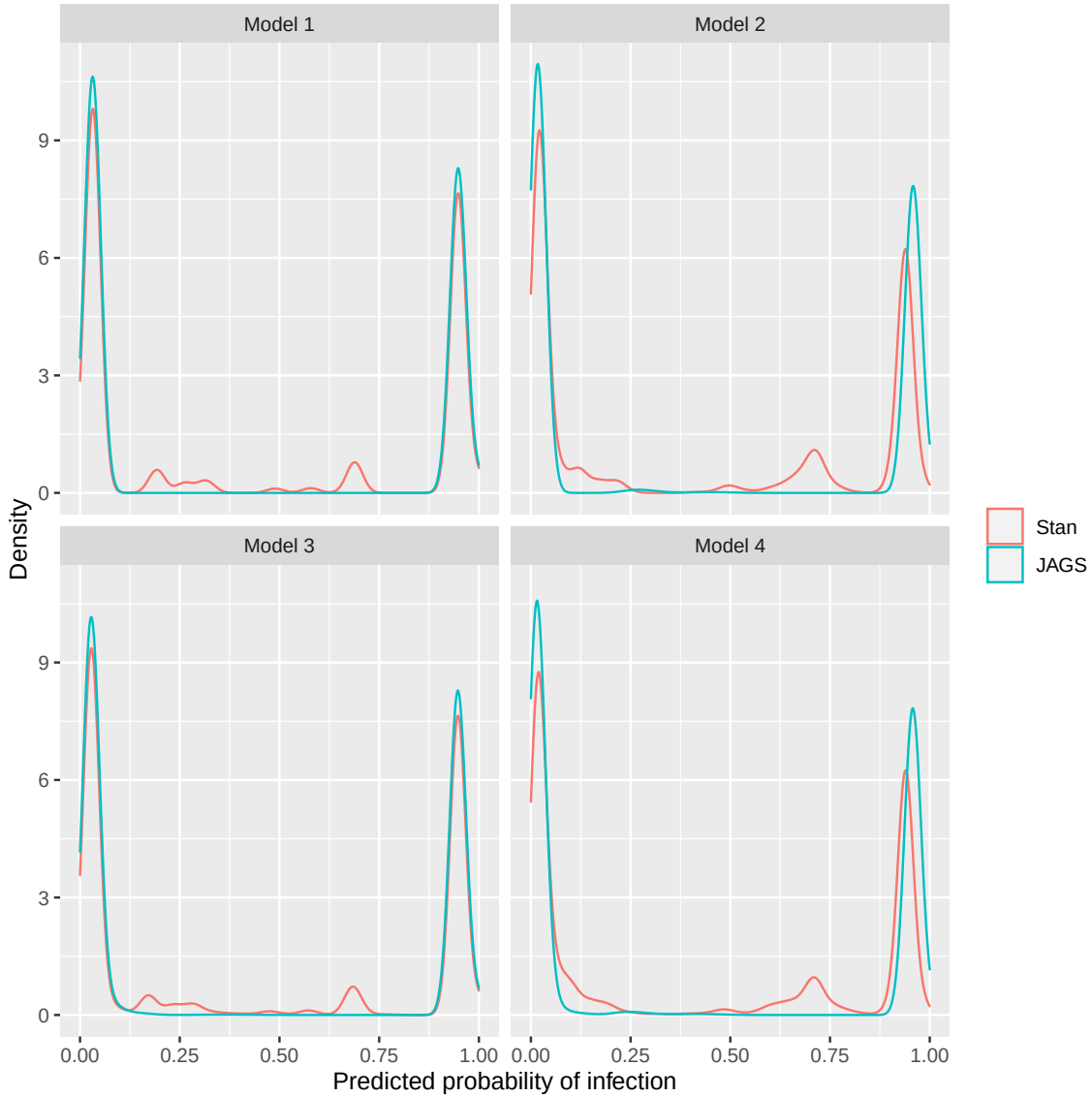
- Estimates consistent between JAGS and Stan
- Posterior distributions much narrower and sometimes far from the prior distributions
- The number of cattle introduced increases the probability of infection $\Rightarrow$ increase in the sensitivity of detection

Predicted probabilities of infection

- Predicted for:
  - all herds
  - the last month in which a test result was available (October 2016)

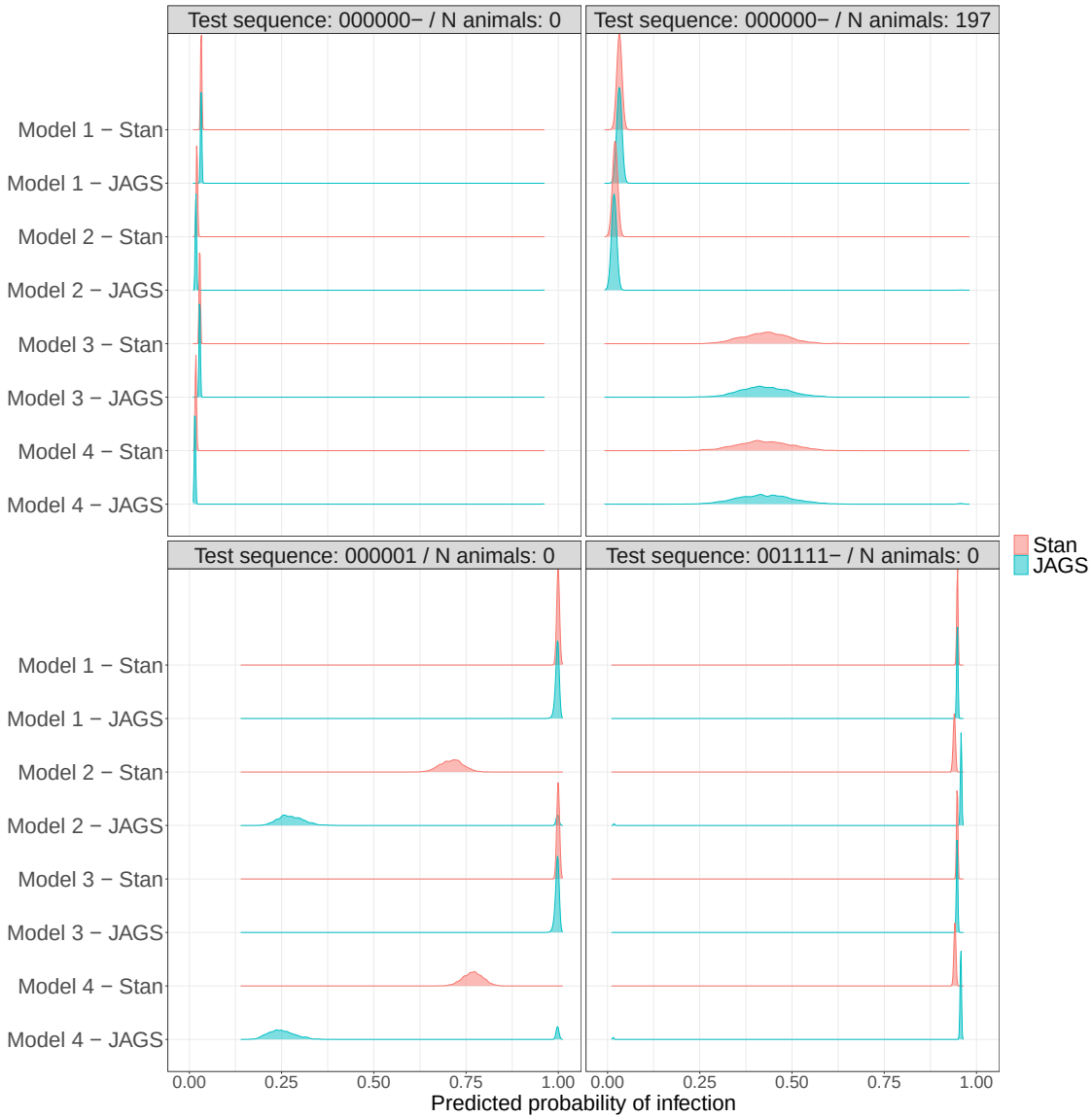
Predicted probabilities of infection

Overall posterior distributions



Predicted probabilities of infection

Posterior distributions for 4 herds



Take home

- The `STOC free` model predicts herd-level probabilities of infection from longitudinal surveillance data

Take home

- The `STOC free` model predicts herd-level probabilities of infection from longitudinal surveillance data
- Estimation / prediction in a Bayesian framework:
  - allows the incorporation of available knowledge on test characteristics and disease epidemiology in a straightforward way using prior distributions
  - the Stan implementation converges much better than the JAGS implementation

Take home

- The `STOC free` model predicts herd-level probabilities of infection from longitudinal surveillance data
- Estimation / prediction in a Bayesian framework:
  - allows the incorporation of available knowledge on test characteristics and disease epidemiology in a straightforward way using prior distributions
  - the Stan implementation converges much better than the JAGS implementation
- Further work:
  - Categorise herds into infection free / not free from the predicted posterior distributions of infection
  - How to define prior distributions for herd-level test characteristics?

Take home

- The `STOC free` model predicts herd-level probabilities of infection from longitudinal surveillance data
- Estimation / prediction in a Bayesian framework:
  - allows the incorporation of available knowledge on test characteristics and disease epidemiology in a straightforward way using prior distributions
  - the Stan implementation converges much better than the JAGS implementation
- Further work:
  - Categorise herds into infection free / not free from the predicted posterior distributions of infection
  - How to define prior distributions for herd-level test characteristics?
- All code and documentation available on Github
  - R package, submitted paper, this presentation and a tutorial

  $\Rightarrow$ please use and improve ...

Take home

- All code and documentation available on Github
  - R package: https://github.com/AurMad/STOCfree
  - Submitted paper: https://www.biorxiv.org/content/10.1101/2020.07.10.197426v4
  - This presentation and a tutorial: https://github.com/AurMad/STOCfree_model_tutorial

$\Rightarrow$ please use and improve !

STOC free

efsa
European Food Safety Authority

Thank you for your attention

http://www.stocfree.eu/

This study was awarded a grant by EFSA and was co-financed by public organisations in the countries participating in the study.

SRUC · UCD DUBLIN · SVA NATIONAL VETERINARY INSTITUTE · Oniris Nantes-Atlantic National College of Veterinary Medicine, Food Science and Engineering · Universiteit Utrecht · FRIEDRICH-LOEFFLER-INSTITUT FLI Bundesforschungsinstitut für Tiergesundheit Federal Research Institute for Animal Health · GD · ahi Animal Health Ireland