Capacity of a Bayesian model to detect infected herds using disease dynamics and risk factor information from surveillance programmes: A simulation study

M. Mercat, A.M. van Roon, I. Santman-Berends, G. van Schaik, M. Nielen, D. Graham, S.J. More, M. Guelbenzu-Gonzalo, C. Fourichon, A. Madouasse

Version of Record: https://www.sciencedirect.com/science/article/pii/S0167587722000150
Manuscript_5eb65b18b4a1a342f4d7511a1b91d88b

Capacity of a Bayesian model to detect infected herds using disease dynamics and risk factor information from surveillance programmes: A simulation study

M. Mercat[1], A.M. van Roon[2], I. Santman-Berends[2,3], G. van Schaik[2,3], M. Nielen[2], D. Graham[4], S.J. More[5], M. Guelbenzu-Gonzalo[4], C. Fourichon[1], A. Madouasse[1]

[1]INRAE, Oniris, BIOEPAR, Nantes 44300, France

[2]Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, PO Box 80151, 3508, TD Utrecht, the Netherlands

[3]Royal GD, PO Box 9, 7400 AA, Deventer, the Netherlands

[4]Animal Health Ireland, Unit 4/5, The Archways, Bridge St., Carrick-on-Shannon, Co. Leitrim N41 WN27, Ireland

[5]Centre for Veterinary Epidemiology and Risk Analysis, UCD School of Veterinary Medicine, University College Dublin, Belfield, Dublin D04 W6F6, Ireland

*** Correspondence:**

Aurélien Madouasse

aurelien.madouasse@oniris-nantes.fr

Keywords: output-based surveillance, Hidden Markov Model, herd-level probability of infection, freedom from infection, longitudinal data, repeated testing.

1 Abstract-

Control programmes against non-regulated infectious diseases of farm animals are widely implemented. Different control programmes have different definitions of “freedom from infection” which can lead to difficulties when trading animals between countries. When a disease is still present, in order to identify herds that are safe to trade with, estimating herd-level probabilities of being infected when classified “free from infection” using field data is of major interest. Our objective was to evaluate the capacity of a Bayesian Hidden Markov Model, which computes a herd-level probability of being infected, to detect infected herds compared to using test results only. Herd-level risk factors, infection dynamics and associated test results were simulated in a population of herds, for a wide range of realistic infection contexts and test characteristics. The model was used to predict the infection status of each herd from longitudinal data: a simulated risk factor and a simulated test result. Two different indexes were used to categorize herds from the probability of being infected into a herd predicted status. The model predictive performances were evaluated using the simulated herd status as the gold standard. The model detected more infected herds than a single final test in 85% of the scenarios which converged. The proportion of infected herds additionally detected by the model, compared to test results alone, varied depending on the context. It was higher in a context of a low herd test sensitivity. On average, around 20%, for high test sensitivity scenarios, and 40%, for low test sensitivity scenarios, of infected herds that were undetected by the test were accurately classified as infected by the model. Model convergence did not occur for 39% of the scenarios, mainly in association with low herd test sensitivity. Detection of additional newly infected herds was always associated with an increased number of false positive herds (except for one scenario). The number of false positive herds was lower for scenarios with low herd test sensitivity and moderate to high incidence and prevalence. These results highlight the benefit of the model, in particular for

control programmes with infection present at an endemic level in a population and reliance on test(s) of low sensitivity.

2 Introduction

Various control programmes (CPs) against infectious diseases of farm animals are implemented in Europe. In order to control or eradicate these diseases, CPs typically focus on the identification of infected units (animals or herds) using diagnostic tests performed at regular time intervals. CPs may be deployed across a territory, from regional to national scale. Testing schemes can vary in terms of type and performance of the test used, the cohorts and numbers of animals tested, and the time interval between tests. These differences have been documented for some endemic cattle diseases, including infections by bovine viral diarrhoea virus (van Roon et al., 2020b), *Mycobacterium avium* subspecies *paratuberculosis* (Whittington et al., 2019), and bovine herpesvirus 1 (Raaperi et al., 2014).

Heterogeneity in CPs may lead to difficulties when trading animals between different regions or countries, as each CP has its own definition of "freedom from infection" which cannot be directly compared. These definitions of a "free status" are usually based on one, or a combination of several, diagnostic test result(s). Limitations in test performance, and time interval between tests lead to uncertainty around these statuses. Imperfections in the testing schemes lead to two types of error. Firstly, a lack of specificity means that some uninfected herds are wrongly categorized as infected, i.e. false positives. Secondly, a lack of sensitivity leads to some infected herds being wrongly categorized as free from infection, i.e. false negatives. The time interval between tests may result in a delay between the times of infection and detection. For herds classified as "free from infection", those that become infected between two consecutive tests will remain classified as "free from infection" until a next test event. Hence, as each CP has its own surveillance strategy, the confidence

63 and associated uncertainty in the true status of a herd classified as “free from infection” may vary
64 depending on the CP. Currently when purchasing an animal from a herd classified as “free from
65 infection” under different CPs, it is not possible to assess the probability of infection for that animal.
66 As trade can be an opportunity for infectious diseases to spread, confidence in “free status” is a key
67 point to support international trade.

68 There is a need for the development of methods that enable a CP-level comparison of confidence of
69 herd-level “freedom from infection”. The traditional solution to obtain a comparable surveillance
70 output in different regions or countries is to use input-based surveillance. This type of surveillance
71 consists in prescribing how surveillance should be performed in terms sampling design, sample size
72 and tests used. However, input-based surveillance does not take into account the diversity of contexts
73 in which CPs are applied (van Roon et al., 2020b) and can be expensive to run, while not being
74 adapted to the specific context of each CP (Cameron, 2012). Alternatively, output-based surveillance
75 may be used, which is not prescriptive in terms of the elements of the programme, but rather in the
76 degree of confidence associated with a free status that must be achieved.

77 Imperfect testing regimes lead to misclassification errors, as highlighted above. To account for this,
78 known risk factors (RFs) for the introduction of infection could be included in the calculation of
79 probability of freedom, as predictors of either current or new infection. Data on such disease-specific
80 RFs should be available for many CPs, given that action on these RFs is used as a way to prevent the
81 introduction of infection. Disease-specific RFs for introduction depend on the pathogen as well as the
82 route of transmission (direct or indirect transmission). For many diseases, animal purchase is a
83 common RF for introduction of infection into herds (Rangel et al., 2015; van Roon et al., 2020a). In
84 the European Union, where cattle identification and the recording of cattle movements between
85 holdings are mandatory, these data could be used to predict (new) infections through purchase, thus
86 contributing to improved estimation of the infection-free status of a herd.

In Madouasse et al. (2021), a modelling framework was described, called the STOC free (Surveillance analysis Tool for Output based Comparison of the confidence of FREEdom from infection) model, that estimates the herd-level probabilities of infection, using data from CP and taking RF occurrence into account. The model estimates the probability of infection at the last time-step for each herd (in a series of sequential test results). Model inputs include repeated test results and the presence of RFs for each herd as measured regularly within the surveillance programme. The framework incorporates knowledge at the population level on infection dynamics, test characteristics and the effect of RFs when estimating probability of infection.

In order to evaluate the capacity of the STOC free model to detect infected herds, a gold standard is required. Gold standard is the true herd status. In the context of the STOC free model, an infected herd is defined as the presence of at least one infected animal. To measure the true status of the herd, it would be necessary to test all the animals within a herd using a perfect test. However, no such data exist in the real-world. Up to now, the STOC free model has only been applied to a single French dataset, which included test results and RFs but no gold standard (Madouasse et al., 2020). An evaluation of the performance of this model under different circumstances is therefore lacking.

The use of simulated data is an effective way to evaluate the predictive accuracy of the STOC free model given the absence of gold standard information in real-world surveillance data. This approach has been used for the evaluation of latent-class models for the estimation of infection prevalence in dairy herds(McAloon et al., 2019). Data simulation allows a simplified system to be created where the true herd status is known. Simulated surveillance data, i.e. test results and RF information collected at regular intervals, can be used as input for the STOC free model as an alternative to real-world surveillance data. The performance of the model can be then evaluated by looking at errors in herd status classification, by comparing true herd status to the status predicted by the model. Furthermore, compared to real data, using simulated data enables a wide range of epidemiological

situations and surveillance modalities to be evaluated. It makes it possible to investigate the potential of the model to be used for different diseases where performance of CPs differs.

The objective of this work was to evaluate the capacity of the STOC free model, which takes account of both dynamics of testing and risk factor information, to improve the detection of infected and newly infected herds compared to test results alone (i.e. the added value of the model in sensitivity). Among infected herds, newly infected are the ones which were not infected at the previous test event. We assumed that the added value of the model in terms of the detection of newly and previously infected herds could be different depending on the epidemiological context (impact of relative risk associated and frequency of risk factor and disease dynamics) and test performances (sensitivity and specificity). Simulated data were used to generate a wide range of realistic CPs (different CP corresponding either to different diseases or to the results of different testing strategies for a disease in different contexts). We quantified the number of additional infected herds detected by the STOC free model compared to test results.

3 Material and methods

3.1 Overall design strategy

Firstly, a dynamic model was developed to simulate herd-level infection and surveillance data under a wide variety of CP scenarios corresponding either to different diseases or to the results of different testing strategies for a disease in different contexts. The simulated surveillance data were then used as input for the STOC free Bayesian Hidden Markov Model, which was run to generate outputs on model parameters estimates and predicted herd status for probability of infection at the last time-step. Finally, model and test performance were compared. The overall design strategy is presented in Figure 1.

3.2 Simulation of herd infection and surveillance data model

We simulated the dynamics of herd infection status depending on the presence of a single RF associated with an increased probability of becoming infected, generating data on herd status and test results at each time-step. Initially, RF presence/absence was simulated. Herd infection status at the first time-step was based on the chosen simulated prevalence of infection. Then, at a given time-step, non-infected herds could become infected according to a probability of new infection between time-steps, which varied depending on RF occurrence. The probability that an infected herd would remain infected between two sequential time-steps was determined by a simulation parameter that represented this probability (of infection not being resolved between two different time-steps). Infection status for a given herd at a given time-step determined the result of a test, assuming a given herd-level test sensitivity and specificity.

3.2.1 Simulation of herd status at each time-step

Infection dynamics were simulated by herd status change. Herd status was simulated as a binary event, with 0 and 1 denoting absence and presence of infection, respectively. Herd status was assumed to undergo Markovian dynamics with status at time *t* depending on status at time *t-1* and RF occurrence. In each scenario, the overall herd infection prevalence was held constant over the time-steps to evaluate the STOC free model in different situations over a short period. Keeping the prevalence constant prevents the infection of either dying out or rapidly increasing over the number of time-steps and allows a comparable number of infected herds to be detected. For consistency, the probability of new infection between time-steps was a function of both overall herd infection prevalence and the probability of a herd remaining infected between time-steps to allow overall prevalence to remain constant over time.

Status simulation can be described by the following set of equations. In herd *h* at time *t*, the infection status $S_{h,t}$, was sampled from a Bernoulli distribution:

$$S_{h,t} \sim \text{Bernoulli}(\pi_{h,t}) ,$$

with $\pi_{h,t}$ being the probability of being infected at time-step *t* for herd *h*. For a given herd at time *t=1*, the probability of infection was:

$$\pi_{h,t=1} = P ,$$

with *P* being the herd infection prevalence for that scenario. For a given herd *h* at time *t>1*, $\pi_{h,t}$ depended on previous status and infection dynamics parameters:

$$\pi_{h,t} = (1 - S_{h,t-1})\tau_1^{h,t} + S_{h,t-1}\tau_2 ,$$

with $S_{h,t-1}$ being the status of herd *h* at the previous time-step, τ_2 being the probability of remaining infected between time-steps (fixed variable in each scenario), and $\tau_1^{h,t}$ the probability of new infection between time-steps which was defined as a function of herd-level risk factor exposure and defined as:

$$\tau_1^{h,t} = (1 - X_{h,t-1})\beta + X_{h,t-1}\beta\gamma ,$$

where β was the probability of new infection when the risk factor was absent, i.e. $X_{h,t-1} = 0$ and $\beta\gamma$ was the probability of new infection when the RF was present, i.e. $X_{h,t-1} = 1$. Thus, γ was the relative risk of new infection in herds exposed to the RF. Exposure to the RF (X) was considered a random dichotomous variable simulated as:

$$X_{h,t} \sim \text{Bernoulli}(F) ,$$

with F being the RF frequency in the data set.

Assuming an endemic situation with a constant prevalence over time-steps, at each time-step in each scenario the average number of newly infected herds was constrained to be equal to the average number of herds eliminating the infection. Therefore, the following condition had to be met:

$$E(\tau_1^{h,t})(1-P) = (1-\tau_2)P \ ,$$

where $E(\tau_{h,t}^1)$ was the expectation for the probability of new infection. This amounts to applying the following constraint on the overall probability of new infection:

$$E(\tau_1^{h,t}) = \frac{(1-\tau_2)P}{1-P} \ .$$

From the definition of $\tau_1^{h,t}$ and the frequency of the RF, *F*, at a given time-step, the expected probability of new infection was:

$$E(\tau_1^{h,t}) = (1-F)\beta + F\beta\gamma \ ,$$

where γ was the relative risk of new infection in herds exposed to the RF and β the probability of new infection in herds that were not exposed to the RF. The frequency of the RF (F) and the relative risk of new infection in herds exposed to the RF (γ) are inputs in the simulation. The probability of new infection in herds that were not exposed to the RF (β) can be computed as:

$$\beta = \frac{E(\tau_1^{h,t})}{1+F(\gamma-1)} \ .$$

3.2.2 Simulation of test results

A test result was simulated for each herd at each time-step as a function of the simulated herd status, the herd-level test sensitivity and specificity. Test result in herd *h* at time *t* was sampled from a Bernoulli distribution:

$$T_{h,t} \sim \text{Bernoulli}(p(T^{+}_{h,t}))\ ,$$

with $p(T^{+}_{h,t})$ being the probability of being tested positive defined by:

$$p(T^{+}_{h,t}) = S_{h,t}Se + (1 - S_{h,t})(1 - Sp)\ ,$$

with Se and Sp being respectively herd-level test sensitivity (probability for an infected herd to be tested positive) and specificity (probability for an uninfected herd to be tested negative).

3.3 Input scenario: differing infection and epidemiological situation

We simulated various scenarios to represent different diseases in different contexts and different tests performances for which STOC free model could be used. Different range of values for the 10 different parameters of the data simulation are presented in Table 1. For all scenarios, the number of simulated herds was set at 5,000 and the number of simulated time-steps to 6. At each time-step, test results and RF information were available. The choice of parameter values was based on knowledge and discussion with a group of infectious disease experts, from different countries involved in the STOC free consortium, to represent variation in context for different endemic situations.

Various epidemiological situations were simulated to represent various endemic infections and contexts. We simulated two prevalence values, 0.3 and 0.1, representing territories in the beginning of their CP and territories already in an advanced stage of control, respectively. The probability of remaining infected depends on the effectiveness of herd-level eradication measures in the CP. We

211 consider high values, from 0.75 to 0.9, consistent with endemic infection dynamics. For consistency
212 with a constant prevalence of infection, the probability of becoming infected (τ_1) was calculated for
213 all combinations of P and $\tau_2$values (4 values).

214 Various effect of RFs on infection dynamics have been simulated to account for variability between
215 CP. We simulated low to high RF frequency setting a maximum frequency of 0.5 considering that a
216 more frequent risk factor would not be discriminatory between herds. In contrast, we have set a
217 minimum frequency at 0.1 because a very rare RF (below 0.1) will only bring information for a small
218 number of herds. The relative risk of new infection in herds exposed to the RF (γ) ranged from 1.5 to
219 5, given that RF association may be variable depending on the infection and territory (van Roon et
220 al., 2020a).

221 The test sensitivities and specificities considered in this paper measure test performance for the
222 detection of infection at the herd-level. These parameters depend on test characteristics at the animal
223 level, the number of animals tested and within-herd prevalence (Christensen and Gardner, 2000).
224 Therefore, herd-level sensitivity and specificity can differ from specific test characteristics and
225 context (Duncan and Humphry, 2016; Nielsen and Toft, 2008). We simulated herd-level sensitivity
226 from 0.4 to 0.9 and herd-level specificity from 0.8 to 0.95. Low herd-level sensitivity values
227 represent infections for which highly sensitive tests are not available, e.g. paratuberculosis (Nielsen
228 and Toft, 2008). We considered a sensitivity of 0.9 as the maximum value. In case of higher
229 sensitivity, we hypothesized that there would be limited added value from the STOC free model.
230 After taking into account the complete testing process, which often includes retesting of positive
231 herds, high values of specificity were considered appropriate. Low diagnostic specificity is less
232 common in CPs.

Combinations of parameters values represented the simulation of 216 different scenarios. Simulation of herd infection and surveillance data model were implemented in R software (R Core Team, 2017).

3.4 Description and use of the STOC free model

The model described by Madouasse et al., 2020, represents infection presence at herd level as a latent status over time-steps. The latent status is evaluated at regular time intervals through testing. Tests may be imperfect, i.e. with a sensitivity and a specificity less than 1. The variable of interest (the latent status) has a Markovian dynamic: the latent status at a given time-step depends on both the latent status at the previous time-step and actions taken or RF occurrence since the previous time-step. Risk factors are incorporated as predictors for new infection. The model predicts the probability of infection in the final time-step for each herd in the CP. Data collected before the final time-step are used as historical data for the estimation of the different model parameters, including previous latent statuses. Parameters estimation and prediction are performed in a Bayesian framework.

3.4.1 Model Structure

To describe the STOC free model and explain how predictions were performed, we use the following notation: $\hat{\beta}$ is the estimated value of β and $\tilde{y}$ is the predicted value of y.

Latent state. We consider two latent states: 0 for uninfected herds and 1 for infected herds. For a given herd h at a given time t, status $\hat{S}_{h,t}$ follows a Bernoulli distribution:

$$\hat{S}_{h,t} \sim \text{Bernoulli}(\hat{\pi}_{h,t}) \,,$$

with $\hat{\pi}_{h,t}$ being the probability of being infected. At *t=1*, a beta prior is used for $\hat{\pi}_{h,t=1}$, representing initial prevalence:

$$\hat{\pi}_{h,t=1} \sim \text{Beta}(\alpha_{\pi}, \beta_{\pi}) \,).$$

Infection dynamics. From the second time-step on, the probability of being infected at t depends on the latent state at $t-1$. Herds that were uninfected at $t-1$ (i.e. $\hat{S}_{h,(t-1)}=0$) can become infected with probability of new infection $\hat{\tau}_1^{h,t}$. Infected herds remain infected with a probability of remaining infected $\hat{\tau}_2$:

$$\hat{\pi}_{h,t} = (1-\hat{S}_{h,(t-1)})\hat{\tau}_1^{h,t} + \hat{S}_{h,(t-1)}\hat{\tau}_2 \, .$$

A beta prior is used for the probability of remaining infected, which is constant over time and herds:

$$\hat{\tau_2} \sim \text{Beta}(\alpha_{\tau_2}, \beta_{\tau_2}).$$

Probability of new infection. The probability of new infection $\tau_{i,t}^1$ is modelled as a function of the presence or absence of the RF $X_{h,t-1}$ using a logistic regression:

$$\text{logit}(\hat{\tau}_1^{h,t}) = \hat{\theta}_1 + \hat{\theta}_2 X_{h,t-1} \, .$$

Normal priors are used for logistic regression parameters $(\hat{\theta}_1, \hat{\theta}_2)$:

$$\hat{\theta}_1 \sim \text{Normal}(\mu_1, \sigma_1) \, ,$$

$$\hat{\theta}_2 \sim \text{Normal}(\mu_2, \sigma_2).$$

Test results. Test results are considered as an imperfect measure of the latent status. We consider two herd-level test results: positive or negative (discrete). Each result follows a Bernoulli distribution with a probability $p(T^+)_{h,t}$ of being positive:

$$T_{h,t} \sim Bernoulli\left(p\left(T_{h,t}^+\right)\right) ,$$

with $p\left(T_{h,t}^{+}\right)$ depending on estimate latent status at *t* and test characteristics: herd-level sensitivity ($\widehat{Se}$) and specificity ($\widehat{Sp}$):

$$p(T_{h,t}^{+}) = \widehat{Se}\hat{S}_{h,t} + (1 - \widehat{Sp})(1 - \hat{S}_{h,t}).$$

Beta priors are used for test characteristics parameters:

$$\widehat{Se} \sim \text{Beta}(\alpha_{Se}, \beta_{Se}) ,$$

$$\widehat{Sp} \sim \text{Beta}(\alpha_{Sp}, \beta_{Sp}).$$

3.4.2 Predicting the probability of infection

The model predicts the herd-level probability of being infected at the last time-step using status prediction from the previous month, estimated infection dynamic parameters, and estimated test specificity and sensitivity.

First, the model predicts the probability of being herd status positive (noted $p(\tilde{S}_{h,t}^{+*})$) depending on previous predicted status ($\hat{S}_{h,t-1}^{+}$) and estimated infection dynamics parameter($\tilde{\tau}_{h,t}^{1}, \hat{\tau}_2$):

$$p(\tilde{S}_{h,t}^{+*}) = p(\tilde{S}_{h,t}^{+*}|p\left(\hat{S}_{h,t-1}^{+}, \tilde{\tau}_1^{h,t}, \hat{\tau}_2\right)) ,$$

with

$$\tilde{\tau}_1^{h,t} = logit^{-1}(\hat{\theta}_1 + \hat{\theta}_2 X_{h,t-1}) .$$

Then, it combines this prediction to test results to compute the final predicted probability of being infected (noted $p(\tilde{S}_{h,t}^{+})$):

$$p\left(\tilde{S}_{h,t}^{+}\middle|T_{h,t}^{+}, \tilde{S}_{h,t}^{+*}\right) = T_{h,t}^{+}.\frac{\widehat{Se}.p(\tilde{s}_{h,t-1}^{+})}{\widehat{Se}.p\left(\tilde{s}_{h,t-1}^{+}\right)+(1-\widehat{Sp}).\left(1-p\left(\tilde{s}_{h,t-1}^{+}\right)\right)} + (1 - T_{h,t}^{+}).\frac{(1-\widehat{Se}).p(\tilde{s}_{h,t-1}^{+})}{(1-\widehat{Se}).p\left(\tilde{s}_{h,t-1}^{+}\right)+\widehat{Sp}.(1-p(\tilde{s}_{h,t-1}^{+}))} ,$$

289 with $T_{h,t}^{+}$being test results at final step time, and $\widehat{Se}$ and $\widehat{Sp}$ being test characteristics parameters
290 estimated by the model. The way to estimate these predicted probability and test results is presented
291 in supplementary materials.

292 **3.4.3 Choice of prior distribution**

293 The STOC free model requires prior distributions for six different parameters: $\widehat{Se}$, $\widehat{Sp}$, $\widehat{\tau_2}$, $\hat{\pi}_{h,t=1}$,
294 $\hat{\theta}_1$and $\hat{\theta}_2$. The distributions and distribution parameters used are summarized in Table 2. We used
295 Beta distributions for parameters bounded between 0 and 1. The Beta distribution requires two
296 parameters. The α and β parameters were estimated using the mean and variance. In our model, a
297 Beta prior was used for the probability of being infected at time-step 1 $\hat{\pi}_{h,t=1}$,, test characteristics $\widehat{Se}$
298 and $\widehat{Sp}$ or the probability of remaining negative $\widehat{\tau_2}$ .We used true input simulated parameter values as
299 the means. The mean value was associated with low variance to build informative priors. We
300 consider that in the case of using real data accurate information would be available to construct such
301 prior. We used a Normal prior for the logistic regression parameter ($\hat{\theta}_1$and $\hat{\theta}_2$) centred on the true
302 value. Types of priors and distribution parameters used are summarized in Table 2. Example of the
303 95% credibility intervals are displayed in the supplementary material.

304 **3.5 Evaluation of STOC free model output**

305 For each scenario, the STOC free model produced different outputs. The model returns Markov
306 Chain Monte Carlo (MCMC) samples from the posterior distributions model parameters and
307 probabilities of being infected at the last time-step. Model parameters include parameters related to
308 infection dynamics, association between RF and probability of new infection and test characteristics.
309 Estimations of these model parameters are performed from historical data on test results and RFs (in
310 our case, data from the first five time-steps) as well as from the prior distributions for the different
311 model parameters. First, we evaluated the convergence of the MCMC chains as well as the

consistency between estimated model parameters and the parameters used for simulating the data. Then, from the posterior distributions of the herd-level probabilities of infections, rules were defined to categorize herds as infected or uninfected. Error rates of the STOC free model were computed and compared to simulated test results to enable computation of model performance.

3.5.1 Evaluation of model parameter estimation

3.5.1.1 Assessing MCMC convergence

The STOC free model were implemented in the JAGS computer programme (Plummer, 2003). The model was applied to each scenario, running 4 chains in parallel. We removed the first 1,000 iterations as burn-in. Then 5,000 more iterations were run, of which one in five iterations was stored for analysis, to reduce the size of the output file. For each parameter, the posterior distribution was built with 4000 iterations (1000 for each chain). We used the Gelman-Rubin statistics ($\hat{r}$) to assess convergence of the chains (Gelman and Rubin, 1992). This statistic was computed for the five parameters estimated by the model ($\widehat{Se}$, $\widehat{Sp}$, $\widehat{\tau_2}$, $\hat{\theta}_1$and $\hat{\theta}_2$). We considered that scenarios with $\hat{r}$ values less than 1.05 had converged. Scenarios that did not reach convergence using 1,000 iterations of burn-in were run again using 5,000 iterations of burn-in. Scenarios that did not reach convergence after this second step were excluded for the rest of the analysis. To again run these scenarios with more iterations would have been too time consuming.

3.5.1.2 Verification of parameter estimation

Parameter estimation was verified by comparing the posterior distributions to the empirical parameter values within the simulated populations. In the data simulation process the value of a parameter can differ between the chosen value for simulating a scenario and the resulting simulated population value. For example, because of the stochasticity in the simulations, for a chosen sensitivity of 0.7 as a simulation parameter there could have been 695 simulated test positives out of 1000 simulated

infections. In this case, the empirical sensitivity of 0.695 was used as the reference to evaluate the posterior distribution for sensitivity.

3.5.2 Evaluation of model prediction performances

The STOC free model returns distributions of the predicted posterior probability of being infected for the 5000 herds at the last time-step (Figure 1). In order to evaluate the performance of the model for the prediction of true infection status, these probability distributions were discretised into *predicted infected* or *predicted uninfected* status. First, each herd posterior probability of being infected was summarised. The median probability per herd was used as the summary value as it was the variable that best discriminated between uninfected and infected herds (results not shown). Then, a cut-off value was applied to the summary values to classify herds as predicted infected or uninfected. The general framework of the prediction performances analysis is presented in Figure 2.

Two different indices were used to select the cut-off value, corresponding to two different objectives and are described below. Those two methods are based on knowledge of true herd status. In our study, we used the simulated herd status as the gold standard, which allowed the number of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) to be calculated. TP herds are infected herds classified infected, FN are infected herds classified uninfected, FP are uninfected herds classified infected and TN are uninfected herds classified uninfected. We computed them using the STOC free model or the test, represented by corresponding subscript (e.g. $TP_{STOCfree}$ and TP_{test}).

3.5.2.1 Identification of cut-off value using Youden's index

Firstly, we used all herd predictions at the last time-step to estimate the cut-off value which minimized classification error (i.e. false positive and false negative). The cut-off choice was

357 determined using the criterion below (Youden, 1950), noting that it is a trade-off between sensitivity
358 and specificity:

359 $$Youden's \quad index = max(Se + Sp) ,$$

360 with

361 $$Se = \frac{TP_{STOCfree}}{TP_{STOCfree}+FP_{STOCfree}} ,$$

362 and

363 $$Sp = \frac{TN_{STOCfree}}{TN_{STOCfree}+FN_{STOCfree}} .$$

364 We ran this analysis using pROC packages in R software.

365 We compared STOC free model performances to test performances. We firstly compared the number
366 of accurately classified herds (TN+TP) by the STOC free model and by the test. Then, we explored
367 the impact of the simulation parameter values on the additional number of infected herds (TP)
368 detected by the STOC free model compared to test results.

369 We applied this cut-off value to a sub-group of the population, specifically only herds that were not
370 infected at the step before prediction (i.e. candidate herds for new infection), using true simulated
371 herd status, to allow us to distinguish between herds remaining uninfected and newly infected herds.
372 We compared STOC free model performances to test performances by doing the same analysis as
373 described above.

374 **3.5.2.2 Alternative cut-off optimizing detection of newly infected herds**

375 We explored an alternative method to choose a cut-off value designed to evaluate the performances
376 of the model for detection of newly infected herds compared to testing. We selected herds that were

candidates to be newly infected. With this approach, we firstly constrained the cut-off value to detect at least one more newly infected herd compare to test results:

$$Number\ of\ additional\ TP = TP_{STOCfree} - TP_{test} > 0\ .$$

For cut-off values that verified this condition, we computed the associated additional number of false positive (FP):

$$Number\ of\ additional\ FP = FP_{STOCfree} - FP_{test}\ .$$

Finally, we computed the NewI cost index. This index was based on a trade-off between the additional numbers of true positive herds and of false positive herds in the STOC free model compared to test results:

$$NewI\ cost\ index = \frac{Number\ \ of\ \ additional\ \ FP}{Number\ \ of\ \ additional\ \ TP}\ .$$

We chose the cut-off value with the lowest value of NewI cost index. This NewI cost index represents the additional number of false positive for each additional true positive detected by the model compared to the test results. When the NewI cost index is negative, the STOC free model classifies less herds as FP and more herds as TP compared to the test results. A NewI cost index of 1 implies that using the STOC free model we had one additional FP for each additional TP. When the NewI cost index is positive (and more than one), there is more than one additional FP for each additional TP using the STOC free model.

In addition, cut-off values selected with both methods (Youden index and NewI cost index) were compared.

4 Results

4.1 Evaluation of model parameter estimation

4.1.1.1 Assessing MCMC convergence

Of the 216 scenarios, 131 had a $\hat{r}$ < 1.05 for all parameters ($\widehat{Se}$, $\widehat{Sp}$, $\widehat{\tau_2}$, $\hat{\theta}_1$and $\hat{\theta}_2$) which confirmed convergence. For the 85 other scenarios, at least one of the five estimated parameters had a $\hat{r}$ > 1.05. For most of these scenarios (71/85), $\theta_1$chains did not converge. There were fewer scenarios where Se, Sp, τ_2 and θ_2 chains did not converge (21, 28, 43 and 37 of 85 scenarios, respectively). These scenarios were re-run using a greater number of iterations during burn-in. From those 85 scenarios, 41 subsequently converged.

The proportion of scenarios that finally converged (with either 1,000 or 5,000 iterations) varied between values of the simulation parameters (Figure 3). About half of the scenarios (38/72) with a test sensitivity of 0.4 did not converge, and about a third of the scenarios (34/108) with a test specificity of 0.8 did not converge. The values of these two simulation parameters (Se and Sp) had the biggest impact on convergence (Figure 3). Considering both parameters, it appears that higher specificity values helped the model to converge for lower and medium, but to a lesser extent with sensitivity values of 0.4 and 0.7. However, it did not make any difference for scenarios with higher sensitivity values (Figure 4).

4.1.1.2 Checking parameters estimation

Of the 172 scenarios for which model convergence was validated, the simulated parameter value was not within the 95% credibility interval of the posterior distribution, for at least one of the simulation parameters, in 13 scenarios. For each of these 13 scenarios, the parameter for which this was the case varied. This corresponds to 14 parameters i.e. 1.6% of the cases (14/860) for which the true value is outside the 95% credibility interval. The gap between the 95% confidence interval of the posterior

distribution and the simulated population value was low for each of the 14 parameters (supplementary material).

4.2 Evaluation of model prediction performances

Performances of the model were analysed for the 172 scenarios that did converge. Table 3 summarizes the number of scenarios for each simulation parameter value remaining at this step.

4.2.1 Ability to detect infected herds in the whole population

With the cut-off based on the Youden index to select the "best" cut-off to classify the whole population, the model accurately classified more infected herds in 152 of the 172 scenarios compared to test results alone (Figure 5). The difference between the model and test results varied from 125 fewer to 509 additional infected herds detected. On average the model detected an additional 105 truly infected herds. This represented a proportion of infected herds additionally detected by the STOC free model from -0.085 to 0.358, with a mean value of 0.110, corresponding to the added value in sensitivity of the surveillance scheme provided by the model (Figure 6).

For all scenarios with herd test sensitivity (Se) of 0.4 and 0.7, the STOC free model detected more infected herds than the test results (Figure 6). For 12 out of 34 scenarios with low sensitivity, the STOC free model detected an additional 0.3 proportion of infected herds than the test, with a mean value of 0.258. Conversely, when sensitivity was high (0.9) the mean value of additional proportion of infected herds was 0.022. Additionally, for all but two scenario with a herd test specificity (Sp) of 0.95, the STOC free model detected more infected herds than the test (Figure 6). The proportion of herds additionally detected was similar whatever the values of the infection dynamics parameters (prevalence (P), incidence rate (τ_1), and probability to remain infected (τ_2)) and RF link parameters (frequency (F) and relative associated risk (γ)) (Figure 6).

4.2.2 Classification of uninfected herds

With the cut-off based on the Youden index, the number of herds classified as false positives increased in 126 scenarios with the model (Figure 5). Only 27 of the 172 scenarios had a higher number of both infected and uninfected herds that were accurately classified. They were mainly associated with medium and high values of sensitivity (0.7, 0.9), the lowest value of specificity (0.8) and the highest value of probability of remaining infected (τ^2) (0.9).

4.2.3 Ability to detect newly infected herds among candidates to new infection

4.2.3.1 Using Youden index

With the cut-off based on the Youden index, the STOC free model accurately classified more newly infected herds in 65 scenarios compared to the test results (Figure 7). The difference between the model and test results varied from 82 fewer to 88 more newly infected herds detected. On average, the model detected 5 fewer herds than the test. This corresponded to a proportion of newly infected herds additionally detected by the STOC free model from -0.603 to 0.370, with a mean value of -0.046 (Figure 8).

Interestingly, for all scenarios with herd test sensitivity of 0.4, the STOC free model detected more newly infected herds than the test results, with the additional proportion of newly infected herds detected ranging from 0.008 to 0.370 (Figure 8). For 48 of the 98 simulated scenarios with a herd test specificity of 0.95, the model detected more truly newly infected herds than the test alone.

4.2.3.2 Using NewI cost index

We developed a new index to select cut-off values, with the constraint to detect at least one more newly infected herd compared to the test. For 13 of the 172 scenarios, no cut-off value allowed the detection of at least one additional newly infected herd. For all the 159 remaining scenarios, using this index allowed the detection of an additional proportion of newly infected herds, ranging from

0.003 to 0.429, with a mean value of 0.071 (Figure 9). This corresponded to the detection of 1 to 156 additional newly infected herds with a mean value of 14 herds. In 24 scenarios, the proportion of additional newly infected herds that were detected was higher than 0.15 (Figure 9). By construction, the test sensitivity value limits the potential number of additional newly infected herds that can be detected by the model (e.g. with a sensitivity of 0.9, the maximum potential proportion of newly infected herds additionally detected is 0.1). On average, the model captured proportions increased by 0.125, 0.076, and 0.034 for sensitivity values of 0.4, 0.7 and 0.9, respectively (Figure 9).

Using the NewI cost index, the cut-off value allows systematically for a better detection of newly infected herds compared to test results but is associated with a cost in false positives. Only 3 scenarios had a negative cost index, whereby it was able to detect more newly infected herds while having less false positives (Figure 10). For all the other scenarios, the additional detection of newly infected herds was always associated with a positive NewI cost index, i.e. a number of additional false positives for each additional true positive detected (Figure 10). This NewI cost index ranged from - 266 to 1055. On average, the cost index value was 98 meaning that for each additional newly infected herd detected, there were an additional 98 false positive herds compared to test results. NewI cost index was <100 for 73% of the scenarios (116/159) (Figure 10). Extremely high values of the cost index (above 500) were associated with a sensitivity of 0.9 for 5 scenarios (Figure 10). These extreme values were also associated with lower proportions of additionally detected newly infected herds (Figure 11.A)). When the proportion of herds additionally detected was above 0.1, the cost index was <100 except in three (Figure 11.A). All scenarios (43) with a high number of newly infected herds (corresponding to τ_1=0.107) had a NewI cost index below 100 (Figure 10 and Figure 11.B).

4.2.4 Comparison of cut-off values

The cut-off values varied substantially between scenarios for both indexes (Figure 12). Use of the Youden index resulted in higher cut-off values (mean cut-off equal0.14 against 0.05 for cost index) (Figure 12). No association between input parameter values (test characteristics, disease dynamics and risk factors parameters) and selection of a cut-off value was found (supplementary material).

5 Discussion

Our simulation study illustrates the added value of a Bayesian Hidden Markov model, the STOC free model, compared to test results alone to detect infected herds in many different contexts. This model was able to predict herd-level probabilities of infection in about 80% of the investigated scenarios. Situations in which the model did not converge and therefore could not provide estimates of the probabilities of infection were mainly related to low sensitivity values. When it converged, the model detected more infected herds compared to test alone in 152 of the 172 scenarios and detected more newly infected herds in only 65 of 172 scenarios. In these scenarios, the STOC free model sensitivity was higher than the herd-level test sensitivity.

Test sensitivity had a great impact on the added value of the STOC free model. Indeed, following a test, the total number of infected and newly infected herds still to be detected (false negatives) increases as test sensitivity decreases. The STOC free model was able to detect an important proportion of these undetected infected herds. On average, the model detected around 25% more infected herds when sensitivity was low (0.4) and around 2% more infected herds when the sensitivity was high (0.9), i.e. around 40% and 20% of the herds still to be detected in our simulations (as assumed for the given levels of sensitivity). The range of herd-level test sensitivities evaluated in this study covers the known range of sensitivities for endemic diseases for which control programmes are in place.

509 An increase in the number of newly infected herds detected by the STOC free model was associated
510 with an increase in the number of false positive herds detected in all but one scenario. We quantified
511 the proportion of additional false positives for each additional newly infected herd detected using a
512 cost index. This cost index increased with high test sensitivity and low prevalence corresponding to
513 small numbers of false test negatives. In five scenarios with a high herd-level test sensitivity, the cost
514 index was substantial (above 500, i.e. 500 false positives for each additional true positive herd
515 detected by the model). This tends to advise against using the STOC free model when test sensitivity
516 is high. On the other hand, the cost index was lower (below 100) with low test sensitivity and high
517 incidence, i.e. when the number of newly infected herds still to be detected was high. For decision
518 support, the level of acceptability in terms of extra false positives would differ according to the
519 consequences in a given control programme, and to the possibilities and resources necessary to
520 confirm a herd status with complementary testing.

521 Different reasons could explain the fact that the STOC free model did not reach convergence in a
522 number of scenarios. In this study, we limited the number of burn-in and sampling iterations to
523 reduce computing time (around 3.5 hours per scenario). For scenarios that did not meet our
524 convergence criterion, re-running the model with more burn-in iterations allowed convergence in
525 around 50% of cases. Adding more iterations could address the remaining convergence issues. A
526 larger population (number of herds) would increase available data (especially in terms of numbers of
527 infected herds) to estimate parameters values. We did not further investigate these hypotheses due to
528 computing time constraint for both simulation and analysis. Low test performance also led to
529 convergence issues. Indeed, as test sensitivity and specificity decrease, the contribution of test results
530 to defining the latent status decreases whereas the contribution of model parameters accounting for
531 new infection and elimination of infection increases. In our study, given the relatively wide prior
532 distributions put on the association between the risk factor and the probability of new infection, this

association was estimated from the data. This means that in the scenarios in which test performance was poor, the contribution of surveillance data to estimation and prediction could be expected to be small, which could have made it more difficult for the model to converge. Such estimation issues have already been described in state-space models, when measurement error is high (Auger-Méthé et al., 2016). In such cases, increase the sample size (e.g. the number of herds) or adding prior information could reduce this issue. In our study, informative priors were used for measurement error parameters (sensitivity and specificity) assuming that relevant epidemiological quantities would be known beforehand. To decrease convergence issue, it could be hypothesised that a good knowledge of the strength of association between risk factors and the probability of new infection facilitates convergence by reducing uncertainty around latent statuses. This knowledge would need to be translated into narrow prior distributions.

The frequency and strength of the risk factor did not influence the STOC free model performances, contrary to our assumptions. The inclusion of RFs was expected to improve the detection of newly infected herds when they strongly contribute to the risk of new infections (high strength of association). This added value was especially expected to be important when test sensitivity is poor, because knowing that a RF is present could compensate for the lack of sensitivity. In our study, only one RF was included to establish its influence on model performance. More RFs can easily be added to the logistic regression if necessary. The choice of RFs to be included must be based on specific knowledge of infection dynamics within the CP.

A cut-off value is needed to classify herds as infected or uninfected from the distributions of probabilities of infection predicted by the STOC free model. The cut-off value varied depending on the method of selection and the simulated context. In the field, the “best” cut-off value would also depend on the objective of the CP. The Youden index equally values sensitivity and specificity without other constraint (i.e. separates at best infected versus non infected herds), while our NewI

cost index ensures the detection of a higher number of newly infected herds than the test alone. For most scenarios, the cut-off value identified with the cost index was lower than the cut-off value identified with the Youden index. Indeed, given an endemic situation, the probability of becoming infected is lower than the probability of remaining infected. The specific detection of newly infected herds, that have not been detected by the test, requires a lower cut-off value compared to the cut-off value selected without this constraint. To compute our cost index associated with the detection of a higher number of newly infected herds we gave the same weight to false positives and false negatives. These two types of misclassification have different consequences: in the context of cattle trade, introducing a false negative into a disease free herd is more damaging than not allowing a false positive to be introduced. Whatever the method of selection used, cut-off values were highly variable between the simulated contexts. According to our study, it does not seem possible to determine a cut-off value directly from CP characteristics. However, we can argue that low cut-off values should be favoured where the objective is safe trade, i.e. limiting false negative herds. In real data where no gold standard is available, the choice of the cut-off value has to rely on another method. This point is an important question when this framework is applied to real data and it needs more exploration.

Applying the STOC free model to real CPs also requires previous knowledge about the distributions of the model parameters. The choice of prior distributions will be crucial because when the prior distributions deviate too much from the true parameter values, this may lead to convergence issues or bias in the posterior distributions. In this simulation study, true parameter values were known, allowing prior distributions to be centred on the true parameter values. In the context of real CPs, test characteristics are almost always assessed before designing the CP. However, even if information is often available, its interpretation must be made in relation to the targeted latent status which may differ from the definition used in the literature and can be challenging (Duncan et al., 2016). Test characteristics may change depending on the latent status of interest. Information on risk factor of

581 introduction is often available as controlling them is a key measure in CPs to reduce the spread of
582 infection between herds (Lindberg and Houe, 2005). Quantitative data can be derived from the
583 literature (e.g. risk factor study, meta-analysis) but are highly variable between territories and not
584 always available for a specific territory (van Roon et al., 2020a). The model makes it possible to use
585 more or less precise priors according to the available information in the population of interest.

586 Within a CP, the dynamics of the infection (incidence and clearance of infection) as well as the
587 contribution of risk factors are expected to change over time given that the majority of CPs generally
588 act on both preventing new infections and eliminating the pathogen from infected herds. Depending
589 on the CP, these changes may be observed over different periods of time. Example of CPs against
590 BVDV have shown that the decrease in prevalence and incidence in European countries occurred
591 over different time lapses (Houe et al., 2014; Joly et al., 2001; Presi et al., 2011). Risk factor
592 contribution (frequency and strength) may also change during a CP. For example, neighbourhood
593 risk of introduction is linked to infection prevalence in the area. When the prevalence decreases in the
594 territory, the strength of association between having contact with neighbouring herds and becoming
595 infected will decrease, while the frequency of contacts between herds remains the same. In our study,
596 infection dynamics and the contribution of the risk factors remained stable over time to simplify
597 parameter estimations. The changes in infection dynamics and contribution of RFs to new infections
598 could be accommodated by running the model over short time periods (e.g. 1 to 3 years), using the
599 parameter posterior distributions for one period as the prior distributions for the next one.

600 The decrease of infection prevalence and incidence with time during a CP can influence
601 performances of the STOC free model. Here, the cost index was higher when incidence and
602 prevalence were low, reflecting a lower positive predictive value, when the number of true positive
603 herds decreases in a population (similarly to surveillance based on tests only). Therefore, we
604 speculate that the use of the STOC free model will be more interesting with disease present at an

605 endemic level in a population rather than when CP results in decreased prevalence close to
606 eradication.

607 6 Conclusion

608 This simulation study demonstrated the capacity of a Hidden Markov Model using disease dynamics
609 and risk factor information from surveillance programmes to detect more infected herds and newly
610 infected herd than test results alone. The added value of the model depends on the context in which a
611 control programme is conducted. It was greatest in situations with low sensitivity tests. However,
612 these situations were also the ones in which the convergence of the model was the most difficult. The
613 added value of the model did not depend on the strength and frequency of the risk factor. The use of
614 the model is likely to be beneficial especially in the early stages of a control programme (when
615 prevalence and incidence are at moderate level) rather than close to eradication.

616 7 References

617 Auger-Méthé, M., Field, C., Albertsen, C.M., Derocher, A.E., Lewis, M.A., Jonsen, I.D., Flemming,
618 J.M., 2016. State-space models' dirty little secrets: Even simple linear Gaussian models can
619 have estimation problems. Sci. Rep. 6, 1–10. https://doi.org/10.1038/srep26677

620 Cameron, A.R., 2012. The consequences of risk-based surveillance: Developing output-based
621 standards for surveillance to demonstrate freedom from disease. Prev. Vet. Med. 105, 280–286.
622 https://doi.org/10.1016/j.prevetmed.2012.01.009

623 Christensen, J., Gardner, I.A., 2000. Herd-level interpretation of test results for epidemiologic studies
624 of animal diseases. Prev. Vet. Med. 45, 83–106. https://doi.org/10.1016/S0167-5877(00)00118-
625 5

626 Duncan, A.J., Gunn, G.J., Humphry, R.W., 2016. Difficulties arising from the variety of testing

627 schemes used for bovine viral diarrhoea virus (BVDV). Vet. Rec. 178, 292 LP – 292.
628 https://doi.org/10.1136/vr.103329

629 Duncan, A.J., Humphry, R.W., 2016. Difficulties arising from the variety of testing schemes used for
630 bovine viral diarrhoea virus. Vet. Rec. 178, 292–292. https://doi.org/10.1136/vr.103329

631 Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat.
632 Sci. 7, 457–511. https://doi.org/10.2307/2246134

633 Houe, H., Nielsen, S.S., Nielsen, L.R., 2014. Control and Eradication of Endemic Infectious diseases
634 in Cattle. London: College Publications.

635 Joly, A., Beaudeau, F., Seegers, H., 2001. Evaluation de la prévalence et de la dynamique de
636 l'infection par le virus de la maladie des muqueuses en Bretagne à l'aide d'un test ELISA sur
637 lait de grand mélange. Epidémiologie Santé Anim. 7–14.

638 Lindberg, A., Houe, H., 2005. Characteristics in the epidemiology of bovine viral diarrhea virus
639 (BVDV) of relevance to control. Prev. Vet. Med. 72, 55–73.
640 https://doi.org/10.1016/j.prevetmed.2005.07.018

641 Madouasse, A., Mercat, M., van Roon, A., Graham, D., Guelbenzu, M., Berends, I.S., van Schaik,
642 G., Nielen, M., Frössling, J., Ågren, E., Humphry, R., Eze, J., Gunn, G., Henry, M., Gethmann,
643 J., More, S., Fourichon, C., 2020. A modelling framework for the prediction of the herd-level
644 probability of infection from longitudinal data. bioRxiv 1–36.
645 https://doi.org/10.1101/2020.07.10.197426

646 Madouasse, A., Mercat, M., Van Roon, A., Graham, D., Guelbenzu, M., Berends, I.S., Van Schaik,
647 G., Nielen, M., Frössling, J., Ågren, E., Humphry, R.W., Eze, J., Gunn, G.J., Henry, M.K.,
648 Gethmann, J., More, S.J., Toft, N., Fourichon, C., 2021. A modelling framework for the

649 prediction of the herd-level probability of infection from longitudinal data. bioRxiv 197426, ve,
650 1–31. https://doi.org/10.1101/2020.07.10.197426

651 McAloon, C.G., Doherty, M.L., Whyte, P., Verdugo, C., Toft, N., More, S.J., O'Grady, L., Green,
652 M.J., 2019. Low accuracy of Bayesian latent class analysis for estimation of herd-level true
653 prevalence under certain disease characteristics—An analysis using simulated data. Prev. Vet.
654 Med. 162, 117–125. https://doi.org/10.1016/j.prevetmed.2018.11.014

655 Nielsen, S.S., Toft, N., 2008. Ante mortem diagnosis of paratuberculosis: A review of accuracies of
656 ELISA, interferon-γ assay and faecal culture techniques. Vet. Microbiol. 129, 217–235.
657 https://doi.org/10.1016/j.vetmic.2007.12.011

658 Plummer, M., 2003. JAGS : A program for analysis of Bayesian graphical models using Gibbs
659 sampling JAGS : Just Another Gibbs Sampler. Proc 3rd Int Work Distrib Stat Comput (DSC
660 2003).

661 Presi, P., Struchen, R., Knight-Jones, T., Scholl, S., Heim, D., 2011. Bovine viral diarrhea (BVD)
662 eradication in Switzerland-Experiences of the first two years. Prev. Vet. Med. 99, 112–121.
663 https://doi.org/10.1016/j.prevetmed.2011.01.012

664 R Core Team, 2017. R: A language and Computing.

665 Raaperi, K., Orro, T., Viltrop, A., 2014. Epidemiology and control of bovine herpesvirus 1 infection
666 in Europe. Vet. J. 201, 249–256. https://doi.org/10.1016/j.tvjl.2014.05.040

667 Rangel, S.J., Paré, J., Doré, E., Arango, J.C., Côté, G., Buczinski, S., Labrecque, O., Fairbrother,
668 J.H., Roy, J.P., Wellemans, V., Fecteau, G., 2015. A systematic review of risk factors associated
669 with the introduction of Mycobacterium avium spp. paratuberculosis (MAP) into dairy herds.
670 Can. Vet. J. 56, 169–177.

van Roon, A.M., Mercat, M., van Schaik, G., Nielen, M., Graham, D.A., More, S.J., Guelbenzu-Gonzalo, M., Fourichon, C., Madouasse, A., Santman-Berends, I.M.G.A., 2020a. Quantification of risk factors for bovine viral diarrhea virus in cattle herds: A systematic search and meta-analysis of observational studies. J. Dairy Sci. 103, 9446–9463. https://doi.org/10.3168/jds.2020-18193

van Roon, A.M., Santman-Berends, I.M.G.A., Graham, D., More, S.J., Nielen, M., van Duijn, L., Mercat, M., Fourichon, C., Madouasse, A., Gethmann, J., Sauter-Louis, C., Frössling, J., Lindberg, A., Correia-Gomes, C., Gunn, G.J., Henry, M.K., van Schaik, G., 2020b. A description and qualitative comparison of the elements of heterogeneous bovine viral diarrhea control programs that influence confidence of freedom. J. Dairy Sci. 103, 4654–4671. https://doi.org/10.3168/jds.2019-16915

Whittington, R., Donat, K., Weber, M.F., Kelton, D., Nielsen, S.S., Eisenberg, S., Arrigoni, N., Juste, R., Sáez, J.L., Dhand, N., Santi, A., Michel, A., Barkema, H., Kralik, P., Kostoulas, P., Citer, L., Griffin, F., Barwell, R., Moreira, M.A.S., Slana, I., Koehler, H., Singh, S.V., Yoo, H.S., Chávez-Gris, G., Goodridge, A., Ocepek, M., Garrido, J., Stevenson, K., Collins, M., Alonso, B., Cirone, K., Paolicchi, F., Gavey, L., Rahman, M.T., De Marchin, E., Van Praet, W., Bauman, C., Fecteau, G., McKenna, S., Salgado, M., Fernández-Silva, J., Dziedzinska, R., Echeverría, G., Seppänen, J., Thibault, V., Fridriksdottir, V., Derakhshandeh, A., Haghkhah, M., Ruocco, L., Kawaji, S., Momotani, E., Heuer, C., Norton, S., Cadmus, S., Agdestein, A., Kampen, A., Szteyn, J., Frössling, J., Schwan, E., Caldow, G., Strain, S., Carter, M., Wells, S., Munyeme, M., Wolf, R., Gurung, R., Verdugo, C., Fourichon, C., Yamamoto, T., Thapaliya, S., Di Labio, E., Ekgatat, M., Gil, A., Alesandre, A.N., Piaggio, J., Suanes, A., De Waard, J.H., 2019. Control of paratuberculosis: Who, why and how. A review of 48 countries. BMC Vet. Res. 15, 1–29. https://doi.org/10.1186/s12917-019-1943-4

695 Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3, 32–35. https://doi.org/10.1002/1097-
696 0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

697

698

699 Table 1: Parameter values for scenario simulation.

Parameter	Description	Value	Condition[1]
$nherds$	Number of herds	5000	-
$nTests$	Number of test times per herd	6	-
Se	Herd-level sensitivity	0.4, 0.7, 0.9	-
Sp	Herd-level specificity	0.8, 0.95	-
P	Prevalence of infection	0.1, 0.3	-
τ_2	Probability of remaining infected	0.75, 0.9	-
τ_1	Probability of becoming infected	0.011, 0.028, 0.043, 0.107	Depends on τ_2 and P values
γ	Relative risk associated with X	1.5, 2, 5	-
F	Frequency of X	0.1, 0.25, 0.5	-
β	Probability of new infection for an uninfected herd without X	0.004, 0.005, 0.007, 0.008, 0.009, 0.010, 0.014, 0.019, 0.020, 0.021, 0.022, 0.025, 0.027, 0.029, 0.031 0.034, 0.036, 0.038, 0.039, 0.041, 0.054, 0.071, 0.076, 0.086, 0.086, 0.095, 0.097, 0.102	Depends on τ_1, γ and F value

700 [1] The condition column details the dependencies between parameters, for parameters whose values
701 are derived from the combination of values of other parameters.

702 Table 2: Prior distribution for the model parameters.

Parameter	Description	Distribution	Mean	Variance
$\widehat{Se}$	Herd-level test sensitivity	$Beta$	True value	0.0025
$\widehat{Sp}$	Herd-level test specificity	$Beta$	True value	0.0025
$\widehat{\tau_2}$	Probability for an infected herd not to eliminate the infection	$Beta$	True value	0.0025
$\widehat{\theta_1}$	Intercept (risk factor)	$Normal$	True value	1
$\widehat{\theta_2}$	Coefficient (risk factor)	$Normal$	True value	1
$\widehat{\pi_{i,1}}$	Probability of being infected at time 1	$Beta$	Prevalence true value	0.0225

703

Table 3: Number of scenarios that converged depending on each value of the simulated parameters.

Parameter	Value	Initial number of scenarios	Number of scenarios that converged
Se	0.4	72	34
	0.7	72	66
	0.9	72	72
Sp	0.8	108	74
	0.95	108	98
P	0.1	108	91
	0.3	108	81
τ_1	0.0111	54	45
	0.0278	54	46
	0.0429	54	38
	0.1071	54	43
τ_2	0.75	108	89
	0.9	108	83
F	0.1	72	62
	0.25	72	57
	0.5	72	53
γ	1.5	72	61
	2	72	59
	5	72	52

Figure captions

Figure 1: Representation of the design strategy. Variables in rectangles represent observational data (risk factor and test results). Variables in circles represent herd infection statuses: true simulated status in solid line and latent estimated/predicted status in dashed line. Observational data simulated using the simulation model are used as input for the STOC free model. Herd statuses predicted by the STOC free model on the 6th time-step are compared to the corresponding simulated statuses, considered as the gold standard.

Figure 2: Representation of the STOC free model prediction performance analysis. At first stage posterior herd probability of being infected are summarized using the median value. Then, categorization of herds is done by applying a cut-off to the distribution of posterior median. Cut-off determination is based on two different indexes. Finally, the performance of the STOC free model is obtained by comparing the number of true positives using the STOC free model and the number of true positives obtained using test information alone.

Figure 3: Proportion of scenarios that converged for each simulation parameter value. Six of the seven simulated parameters are represented : Se (test sensitivity), Sp (test specificity), P (prevalence), τ_2 (probability of remaining infected), F (frequency of the risk factor) and γ (relative risk associated with the risk factor).

Figure 4 : Proportion of scenarios that converged for all combinations of Se (test sensitivity) and Sp (test specificity) values.

Figure 5: Difference between the number of herds accurately classified by the STOC free model and the number of herds accurately classified using test results for infected herds only, for uninfected herds and for all herds. Dark blue diamond represents the mean of each

distribution. At the dashed grey line, the STOC free model and test results accurately classified the same numbers of herds.

Figure 6: Additional proportion of infected herds accurately classified by the STOC free model relative to test results, among the total number of infected herds, depending on simulated parameter values, using cut-off found applying Youden index. The seven simulated parameters are represented : Se (test sensitivity), Sp (test specificity), F (frequency of the risk factor), γ (relative risk associated with the risk factor), P (prevalence), τ_1 (probability of being newly infected) and τ_2 (probability of remaining infected). Dark blue diamond represents the mean of each distribution. At the dashed grey line, the STOC free model and test results accurately classified the same numbers of herds.

Figure 7: Difference between the number of herds accurately classified by the STOC free model and the number of herds accurately classified using test results only for herds which were candidates for new infection at the final time-step (i.e. herds that were uninfected at the previous step time) for newly infected herds, uninfected herds and all herds, using cut-off found applying Youden index. Dark blue diamond represents the mean of each distribution. At the dashed grey line, the STOC free model and test results accurately classified the same numbers of herds.

Figure 8: Additional proportion of newly infected herds detected by the STOC free model relative to test results, among the total number of newly infected herds, depending on simulated parameter values, using cut-off found applying Youden index. The seven simulated parameters are represented : Se (test sensitivity), Sp (test specificity), F (frequency of the risk factor), γ (relative risk associated with the risk factor), P (prevalence), τ_1 (probability of being newly infected) and τ_2 (probability of remaining infected). Dark blue diamond

represents the mean of each distribution. At the dashed grey line, the STOC free model and test results accurately classified the same numbers of herds.

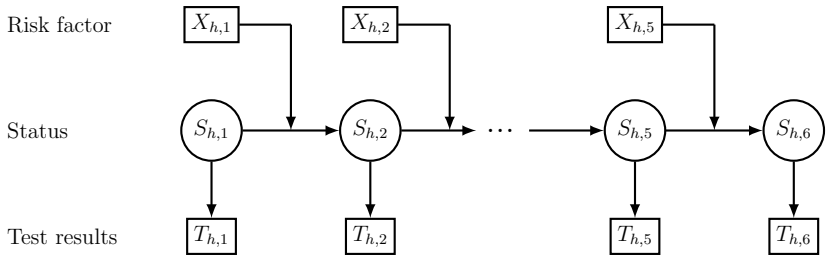
Figure 9: Additional proportion of newly infected herds detected by STOC free model relative to test results, among the total number of newly infected herds, depending on simulated parameter values, using cut-off found applying NewI cost index. The seven simulated parameters are represented : Se (test sensitivity), Sp (test specificity), F (frequency of the risk factor), γ (relative risk associated with the risk factor), P (prevalence), τ_1 (probability of being newly infected) and τ_2 (probability of remaining infected). Dark blue diamond represents the mean of distribution. At the dashed grey line, the STOC free model and test results accurately classified the same numbers of herds.

Figure 10: Cost index value, i.e. the number of additional false positive herds for each additional true positive herds by the STOC free model relative to test results, depending on simulated parameter values, using cut-off found applying NewI cost index. The seven simulated parameters are represented : Se (test sensitivity), Sp (test specificity), F (frequency of the risk factor), γ (relative risk associated with the risk factor), P (prevalence), τ_1 (probability of being newly infected) and τ_2 (probability of remaining infected). Dark blue diamond represents the mean of distribution. Under the dashed grey line cost is negative meaning that STOC free model do detect more newly infected for less false positive herds compared to test results.

Figure 11: Cost index value, i.e. the ratio of additional false positive herds on the additional true positive herds, using cut-off found applying NewI cost index, depending on the proportion of additional newly infected herds detected (A) and the number of newly infected herds which depend on the four possible values of the probability of become infected (τ_1).

Figure 12: Distribution of cut-off values for herd status classification depending on the index used. “NewI” criterion is based on a trade-off between the additional number of true positive herds and the additional number of false positive herds. “Youden” criterion is based on maximizing sensitivity and specificity based on classification of all herds. Dark blue diamond represents the mean of the distribution.

Model structure for simulation of herd infection and surveillance data



$p(X_{h,t} = 1) = F$

$p(S_{h,1} = 1) = P$

$p(S_{h,t>1} = 1|S_{h,t-1} = 0) = \tau_1^{h,t}$

$\tau_1^{h,t} = (1 - X_{h,t-1})\beta + X_{h,t-1}\beta\gamma$

$p(S_{h,t>1} = 1|S_{h,t-1} = 1) = \tau_2$

$p(T_{h,t} = 1|S_{h,t} = 1) = Se$

$p(T_{h,t} = 0|S_{h,t} = 0) = Sp$

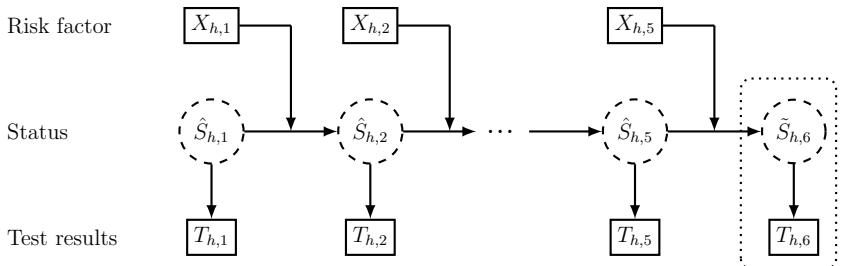
Parameters that vary among simulated scenarios

Infectious dynamics parameters: P, τ_2 and τ_1

Risk factor effect parameters: F, γ and β

Tests performances parameters: Se and Sp

STOCfree model structure for the prediction of the herd-level probability of infection



$\hat{S}_{h,t}$ ~Bernoulli($\hat{\pi}_{h,t}$)

$\hat{\pi}_{h,t>1} = (1 - \hat{S}_{h,(t-1)})\hat{\tau}_1^{h,t} + \hat{S}_{h,(t-1)}\hat{\tau}_2$

$logit(\hat{\tau}_1^{h,t}) = \hat{\theta}_1 + \hat{\theta}_2 X_{h,t-1}$

$T_{h,t}$ ~Bernoulli($p(T_{h,t}^+)$)

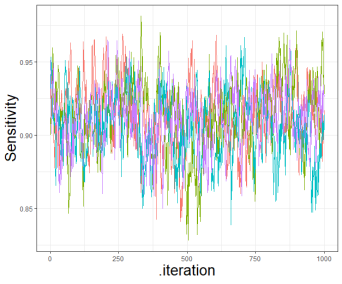
$p(T_{h,t}^+) = \hat{Se}\hat{S}_{h,t} + (1 - \hat{Sp})(1 - \hat{S}_{h,t})$

With prior on $\hat{\pi}_{h,t=1}$, $\hat{Se}$, $\hat{Sp}$, $\hat{\tau}_2$, $\hat{\theta}_1$, $\hat{\theta}_2$

Evaluation of model parameter estimation

$\hat{Se}, \hat{Sp}, \hat{\tau}_2, \hat{\theta}_1, \hat{\theta}_2$

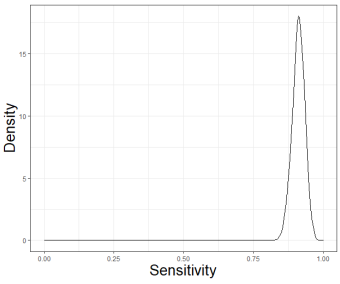
Markov Chain



Assessing MCMC convergence

Gelman-rubin statistics

Posterior distribution

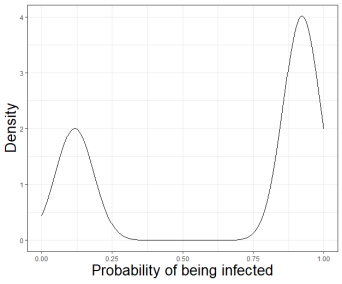


Assessing parameter estimation accuracy

Posterior distribution vs simulated population value

Evaluation of model prediction

Posterior herd level probability of being infected $p(\tilde{S}_{h,6})$

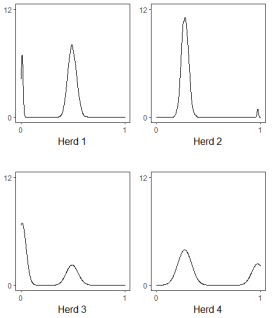


Assessing prediction performances

Comparison of (newly) infected herds detected by : - the test

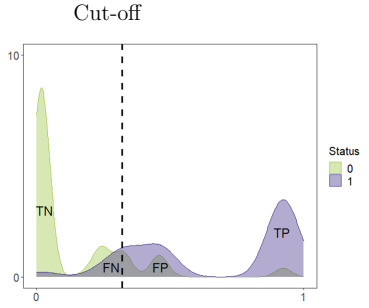
- the STOCfree model

Herd level posterior probability of being infected



Median probability

Distribution of median probability of being infected

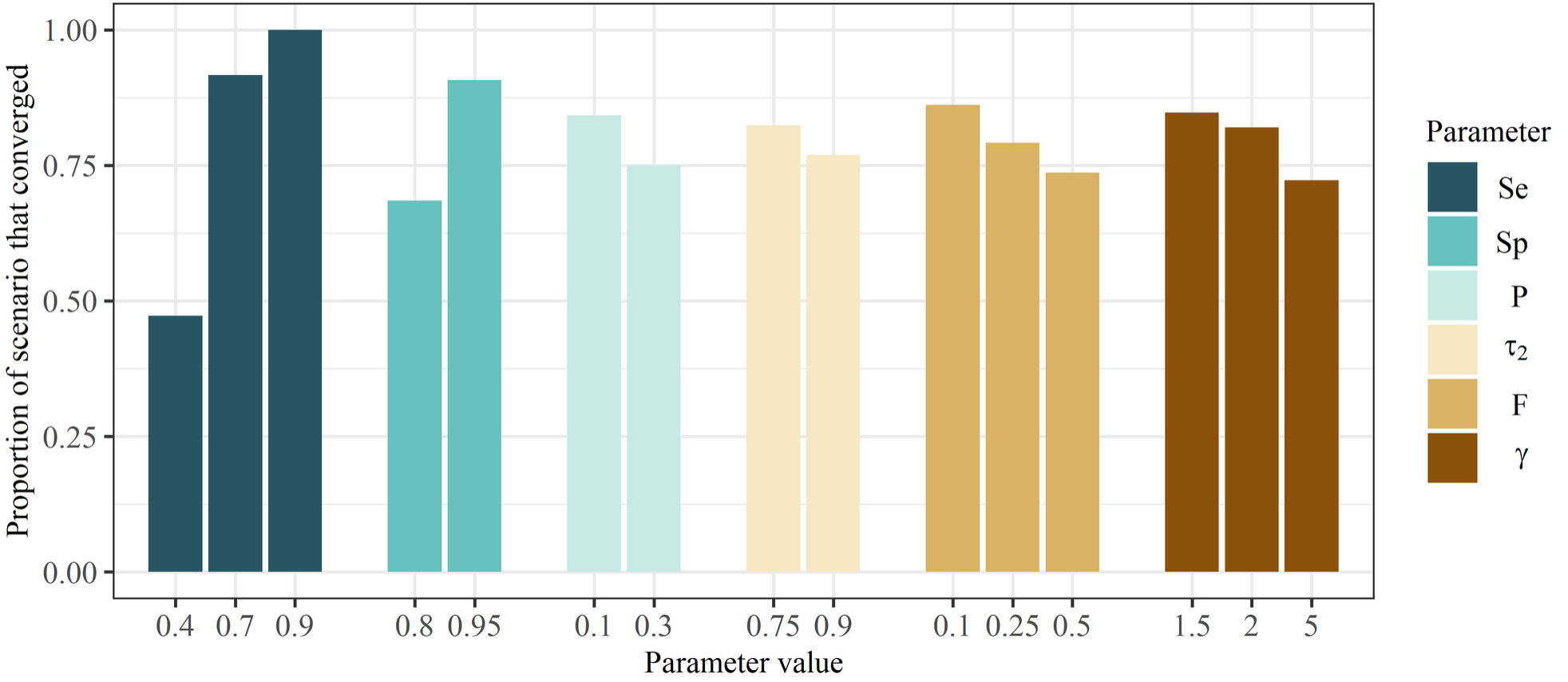


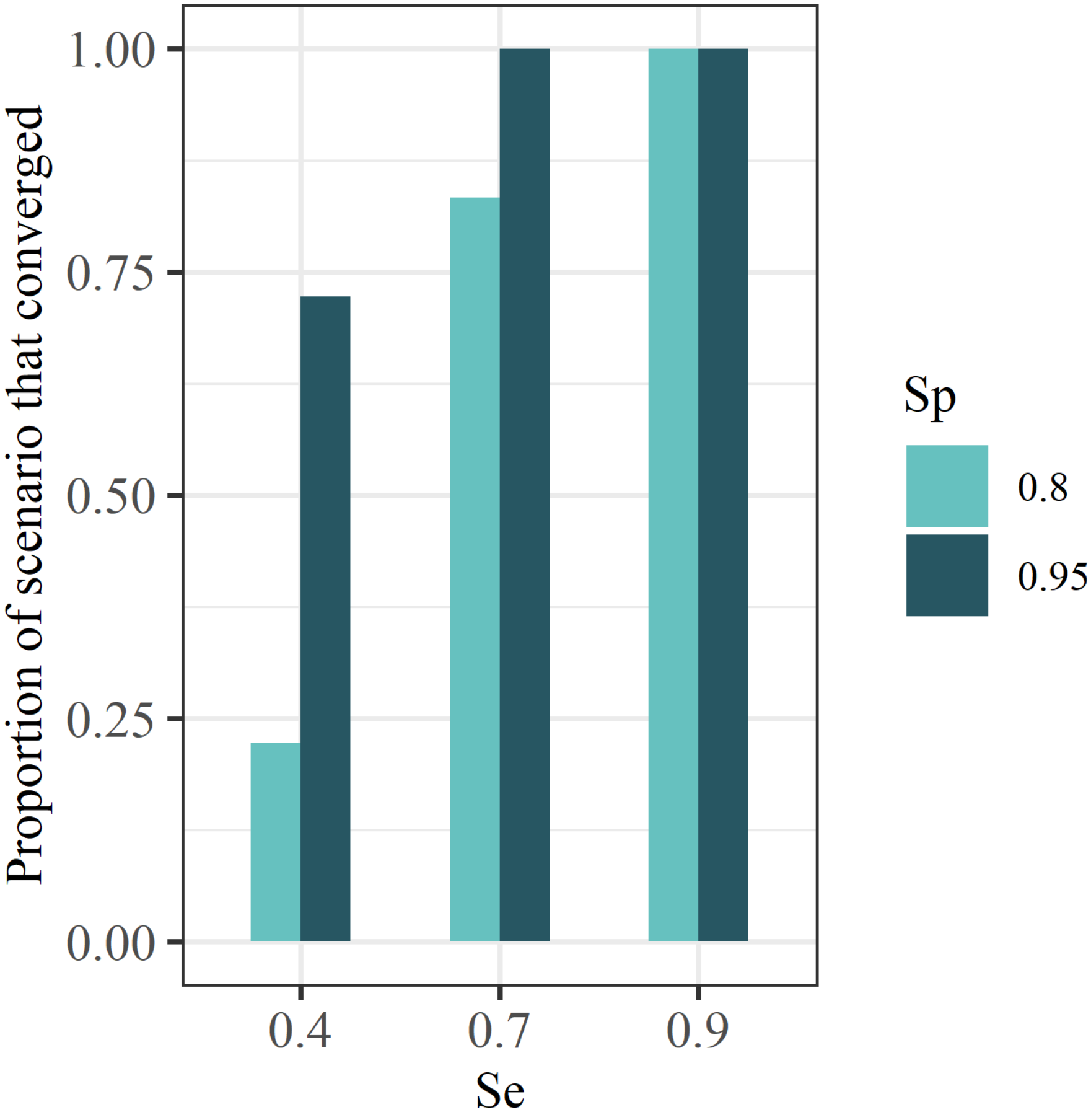
Herd classification

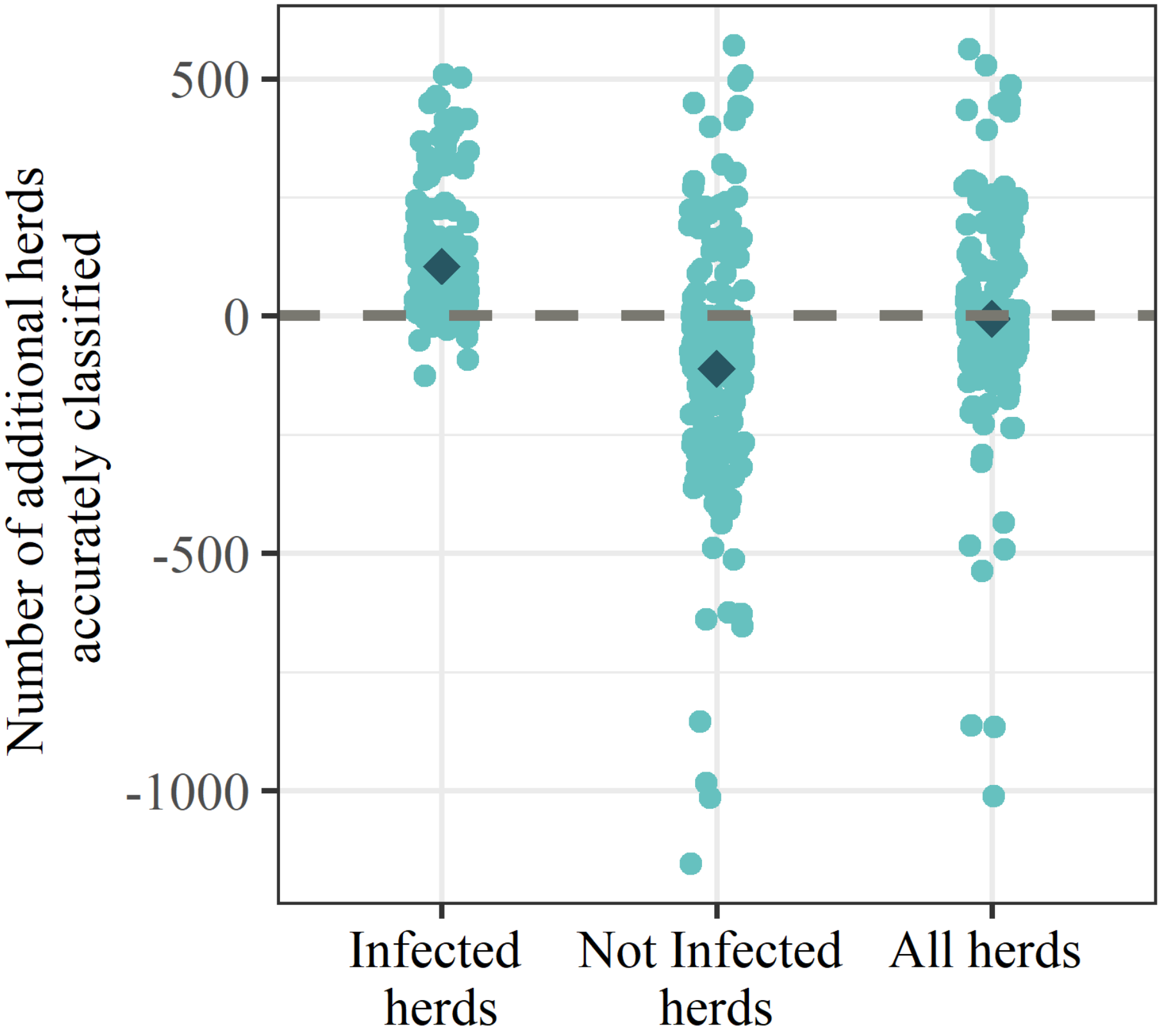
STOC free model performances

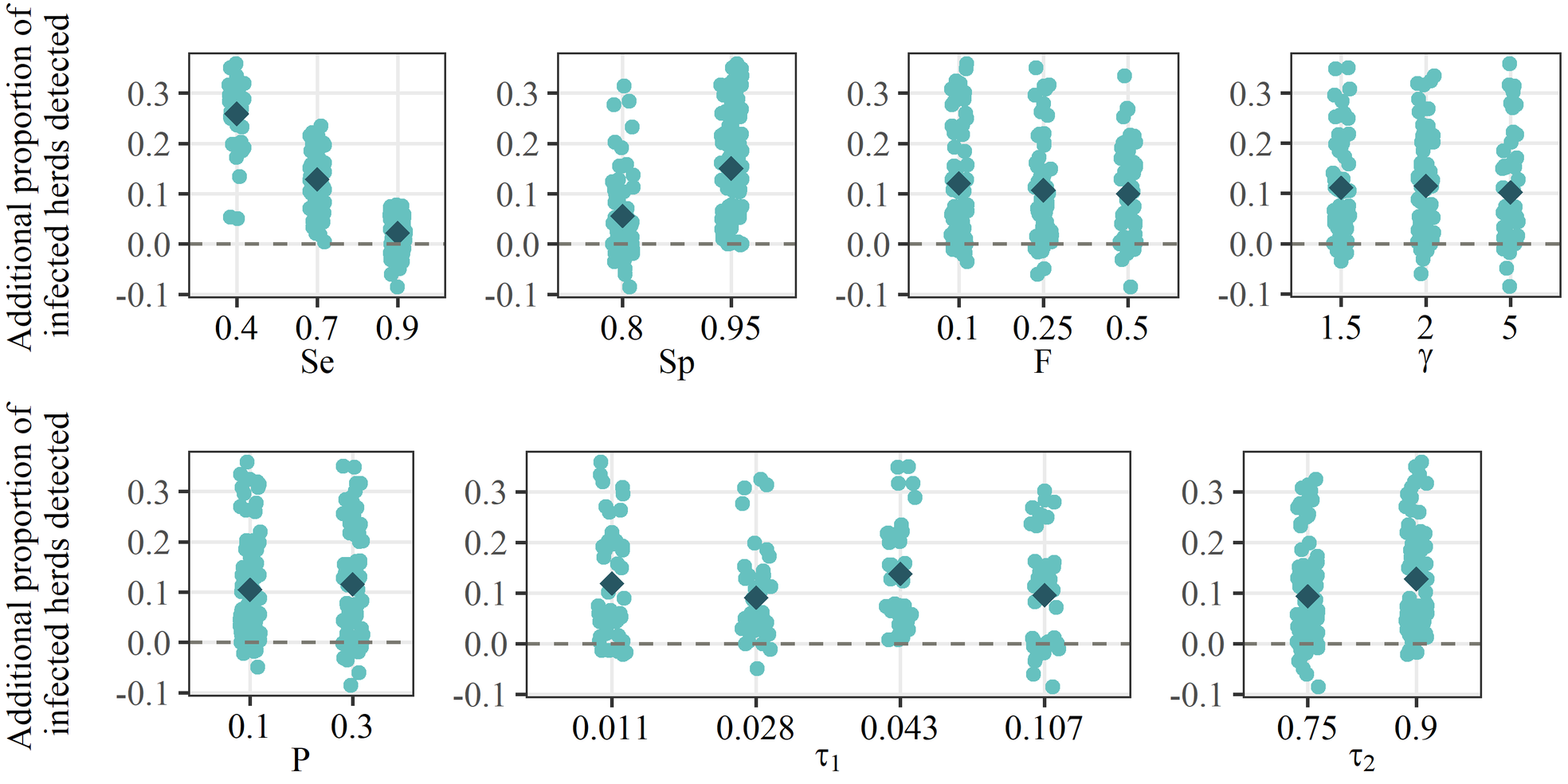
Proportion of infected herds additionally detected compared to test:

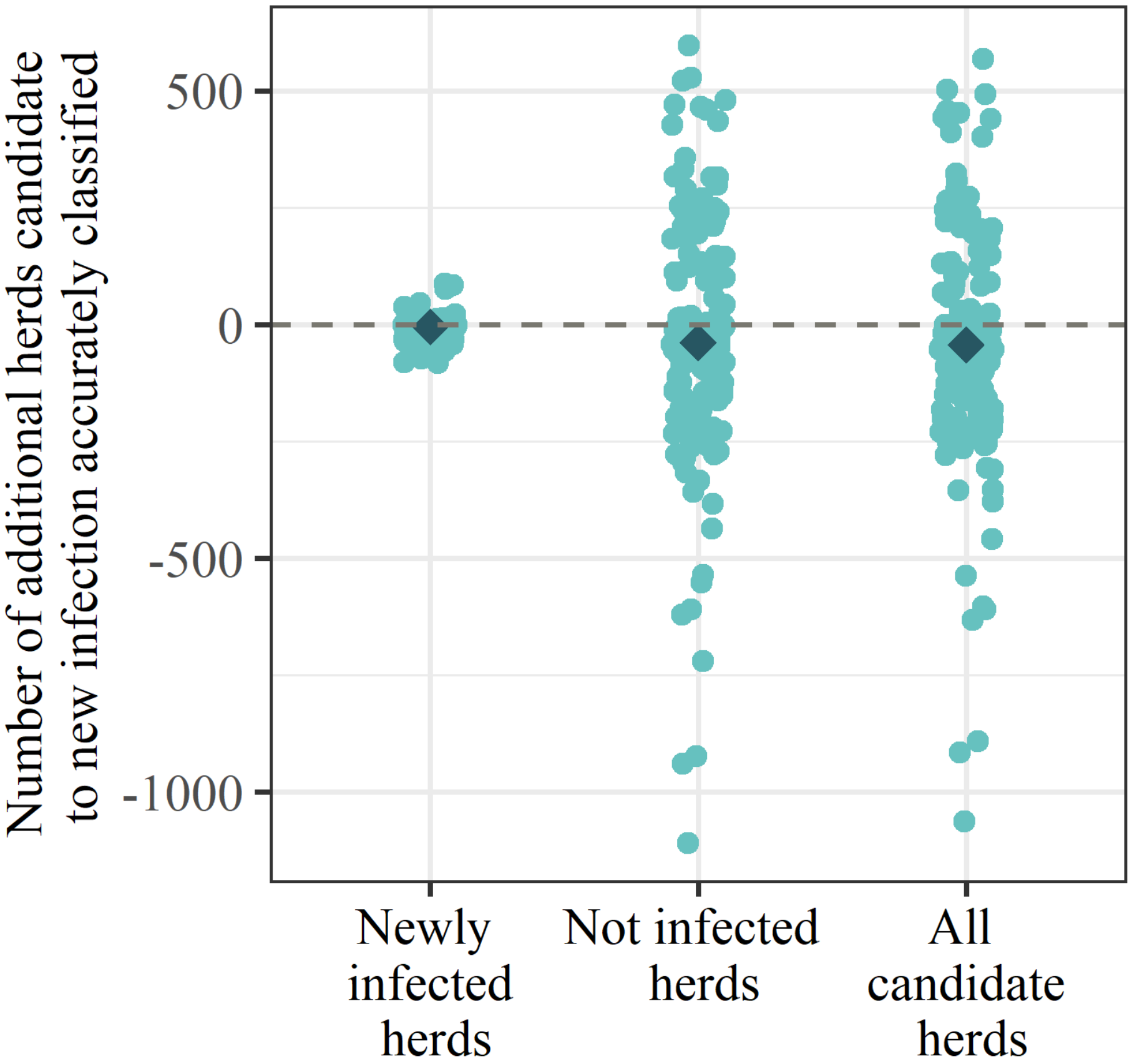
$$\frac{TP_{STOCfree} - TP_{test}}{Number\ of\ infected\ herds}$$

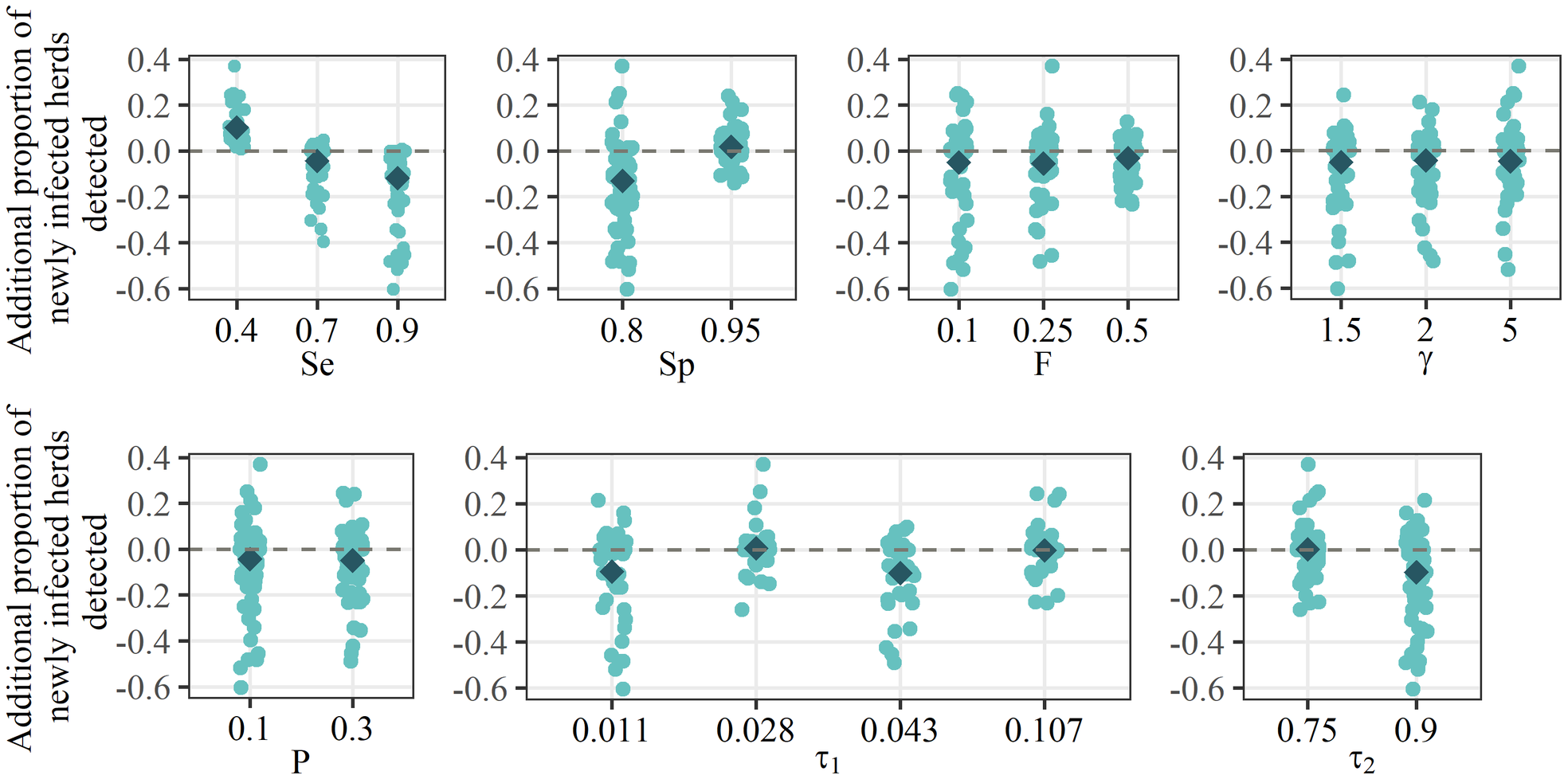

Proportion of scenario that converged
1.00
0.75
0.50
0.25
0.00
0.4
0.7
0.9
0.8
0.95
0.1
0.3
0.75
0.9
0.1
0.25
0.5
1.5
2
5
Parameter value
Parameter
Se
Sp
P
τ2
F
γ

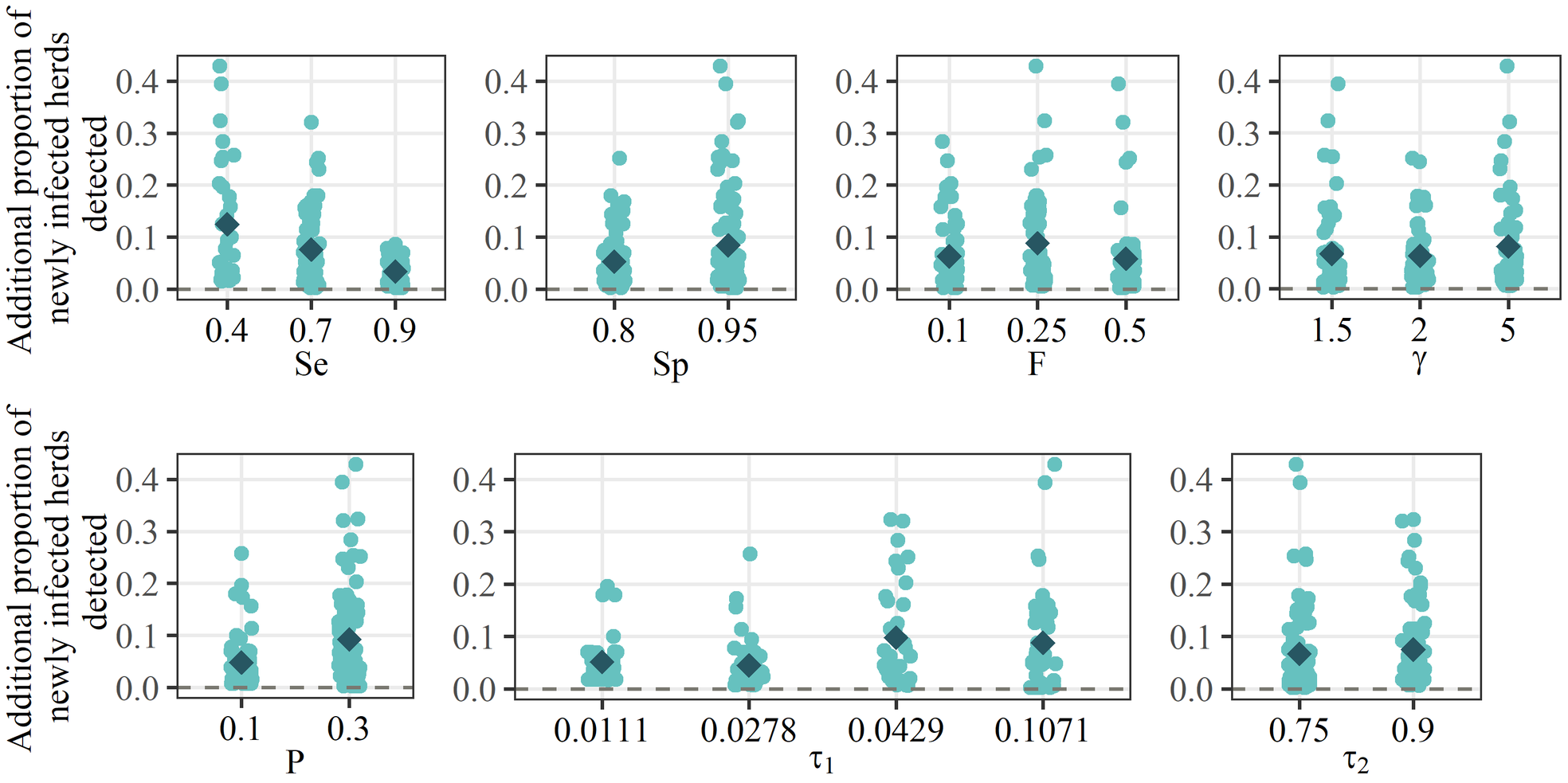

1.00
0.75
0.50
0.25
0.00
Proportion of scenario that converged
0.4
0.7
0.9
Se
Sp
0.8
0.95

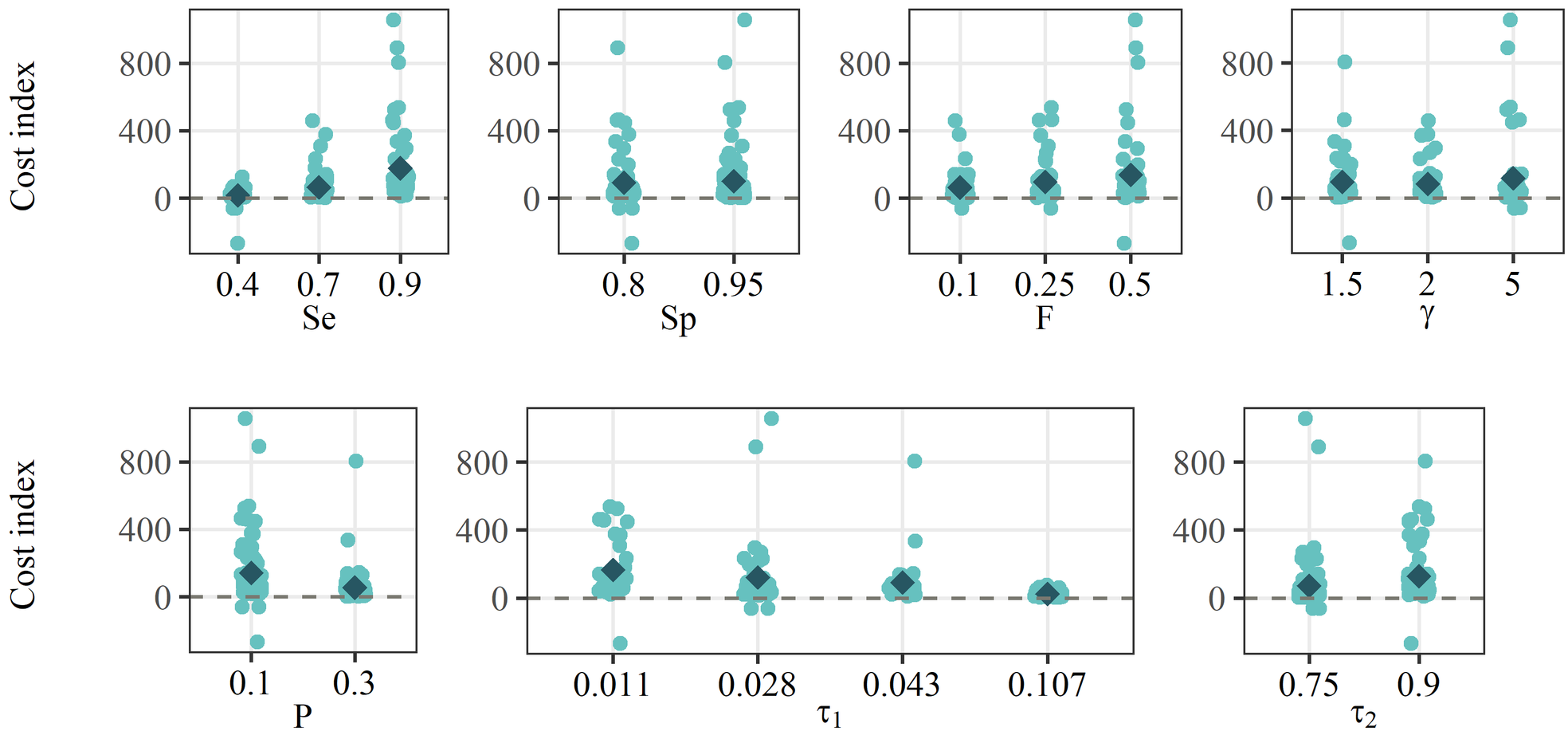

Number of additional herds accurately classified
500
0
-500
-1000
Infected herds
Not Infected herds
All herds

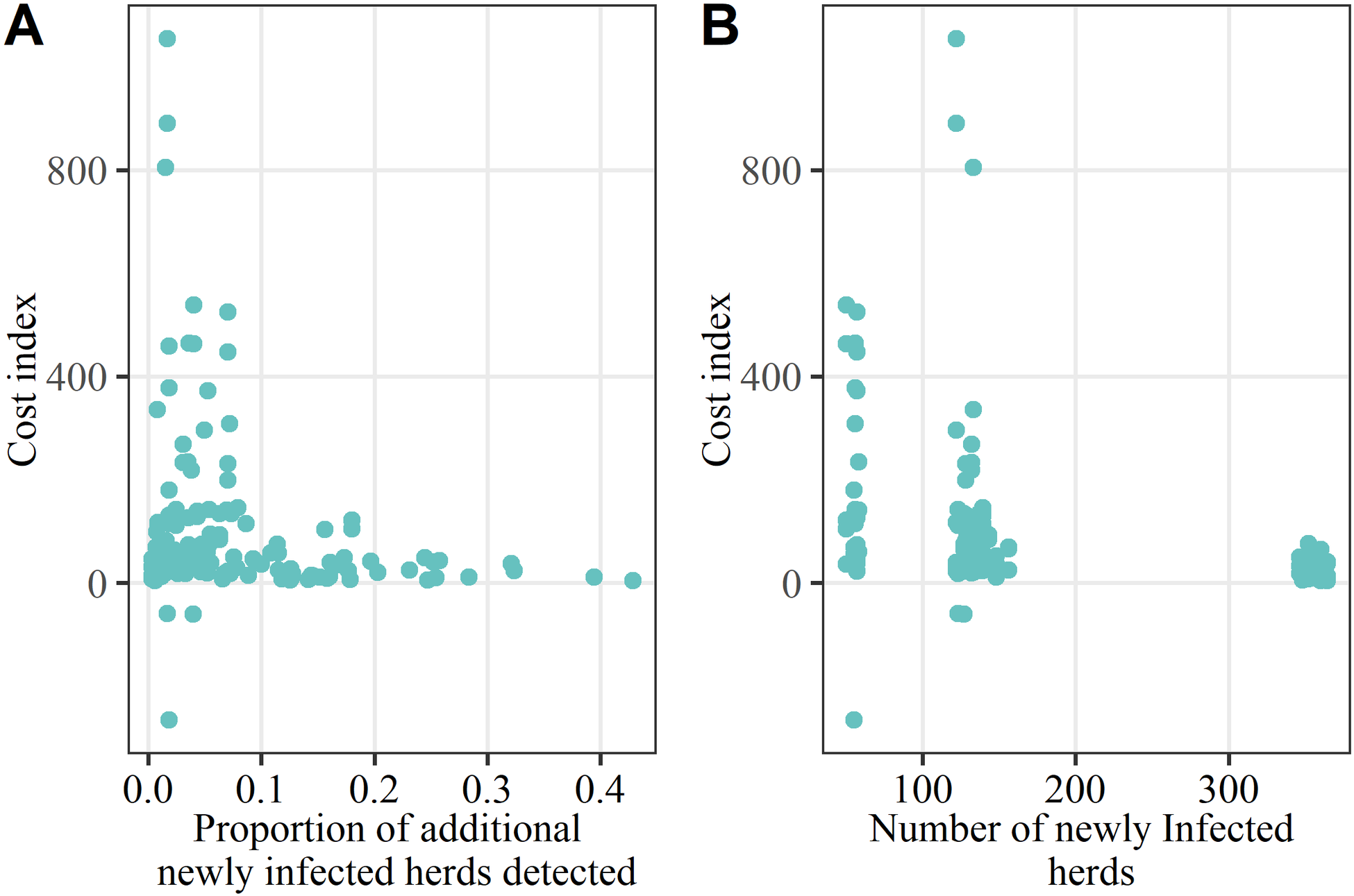

Additional proportion of infected herds detected
0.3
0.2
0.1
0.0
-0.1
0.4
0.7
0.9
Se
0.8
0.95
Sp
0.1
0.25
0.5
F
1.5
2
5
γ
Additional proportion of infected herds detected
0.1
0.3
P
0.011
0.028
0.043
0.107
τ1
0.75
0.9
τ2

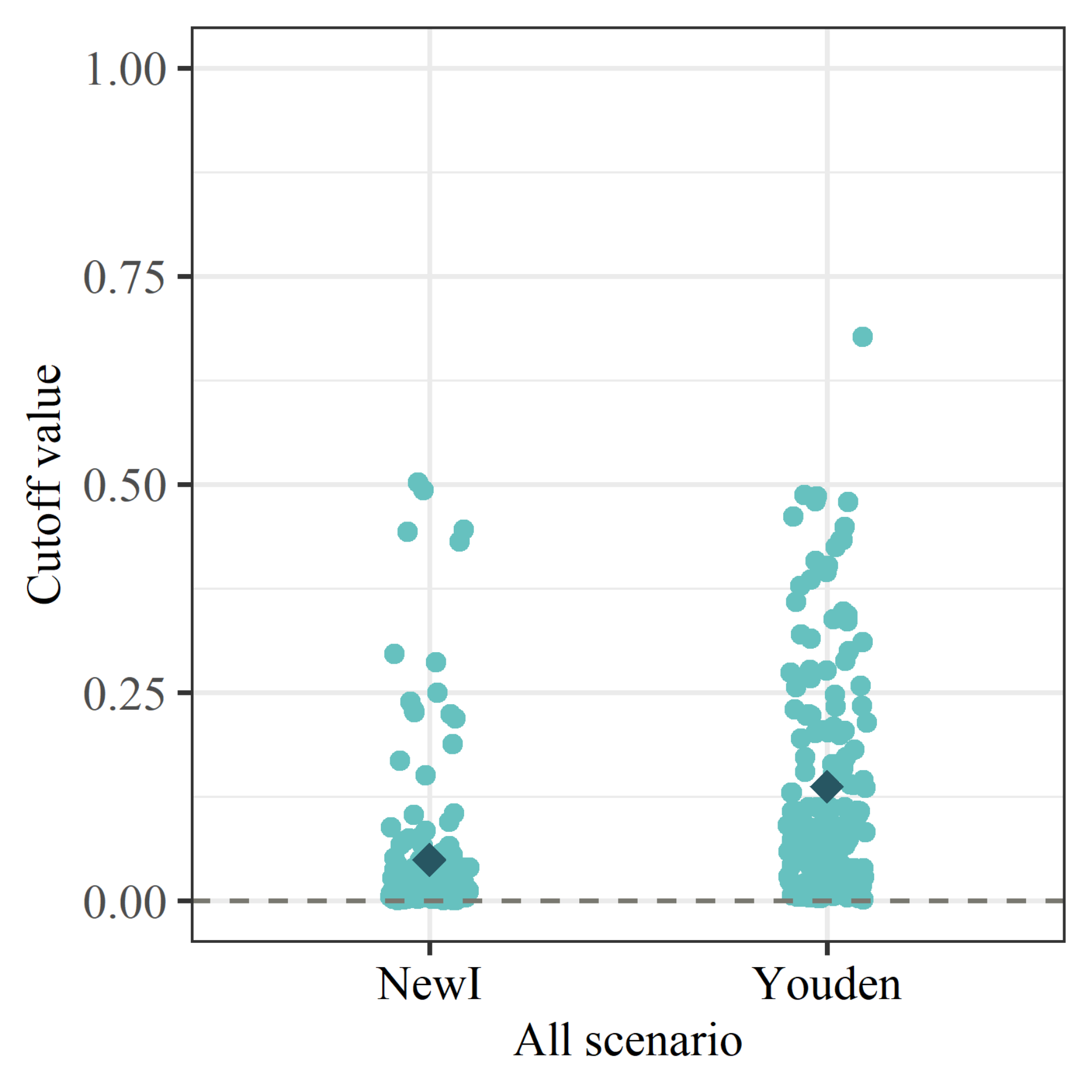

Number of additional herds candidate to new infection accurately classified
500
0
-500
-1000
Newly infected herds
Not infected herds
All candidate herds


Additional proportion of newly infected herds detected
0.4
0.2
0.0
-0.2
-0.4
-0.6
0.4
0.7
0.9
Se
0.8
0.95
Sp
0.1
0.25
0.5
F
1.5
2
5
γ
Additional proportion of newly infected herds detected
0.1
0.3
P
0.011
0.028
0.043
0.107
τ1
0.75
0.9
τ2


Additional proportion of newly infected herds detected
0.0
0.1
0.2
0.3
0.4
0.4
0.7
0.9
Se
0.8
0.95
Sp
0.1
0.25
0.5
F
1.5
2
5
γ
0.1
0.3
P
0.0111
0.0278
0.0429
0.1071
τ1
0.75
0.9
τ2


Cost index
800
400
0
0.4
0.7
0.9
Se
0.8
0.95
Sp
0.1
0.25
0.5
F
1.5
2
5
γ
Cost index
0.1
0.3
P
0.011
0.028
0.043
0.107
τ1
0.75
0.9
τ2


A
B
800
400
0
Cost index
0.0
0.1
0.2
0.3
0.4
Proportion of additional newly infected herds detected
100
200
300
Number of newly Infected herds


1.00
0.75
0.50
0.25
0.00
Cutoff value
NewI
Youden
All scenario