

A hierarchical model to evaluate pest treatments from prevalence and intensity data

ISSN 2824-7795

Armand Favrot¹ MIA Paris-Saclay, INRAE AgroParisTech Université Paris-Saclay, France
David Makowski² MIA Paris-Saclay, INRAE AgroParisTech Université Paris-Saclay, France

Date published: 2023-10-06 Last modified:

Abstract

In plant epidemiology, pest abundance is measured in field trials using metrics assessing either pest prevalence (fraction of the plant population infected) or pest intensity (average number of pest individuals present in infected plants). Some of these trials rely on prevalence, while others rely on intensity, depending on the protocols. In this paper, we present a hierarchical Bayesian model able to handle both types of data. In this model, the intensity and prevalence variables are derived from a latent variable representing the number of pest individuals on each host individual, assumed to follow a Poisson distribution. Effects of pest treatments, time trend, and between-trial variability are described using fixed and random effects. We apply the model to a real dataset in the context of aphid control in sugar beet fields. In this dataset, prevalence and intensity were derived from aphid counts observed on either factorial trials testing different types of pesticides treatments or field surveys monitoring aphid abundance. Next, we perform simulations to assess the impacts of using either prevalence or intensity data, or both types of data simultaneously, on the accuracy of the model parameter estimates and on the ranking of pesticide treatment efficacy. Our results show that, when pest prevalence and pest intensity data are collected separately in different trials, the model parameters are more accurately estimated using both types of trials than using one type of trials only. When prevalence data are collected in all trials and intensity data are collected in a subset of trials, estimations and pest treatment ranking are more accurate using both types of data than using prevalence data only. When only one type of observation can be collected in a pest survey or in an experimental trial, our analysis indicates that it is better to collect intensity data than prevalence data when all or most of the plants are expected to be infested, but that both types of data lead to similar results when the level of infestation is low to moderate. Finally, our simulations show that it is unlikely to obtain accurate results with fewer than 40 trials when assessing the efficacy of pest control treatments based on prevalence and intensity data. Because of its flexibility, our model can be used to evaluate and rank the efficacy of pest treatments using either prevalence or intensity data, or both types of data simultaneously. As it can be easily implemented using standard Bayesian packages, we hope that it will be useful to agronomists, plant pathologists, and applied statisticians to analyze pest surveys and field experiments conducted to assess the efficacy of pest treatments.

Keywords: bayesian model, epidemiology, hierarchical model, pest control, trial, survey

¹ Contents

² 1 Introduction ²

¹Corresponding author: armand.favrot@inrae.fr

²Corresponding author: david.makowski@inrae.fr

3	2 Material and Method	4
4	2.1 Description of the data	4
5	2.2 Model	4
6	2.2.1 Specification	4
7	2.2.2 Inference on real data	7
8	2.2.3 Simulations	10
9	3 Results	16
10	3.1 Results obtained with real data	16
11	3.1.1 Model convergence and posterior distributions	16
12	3.1.2 Estimated values of pesticide treatment efficacies	16
13	3.2 Results obtained by simulation	19
14	3.2.1 Interest of combining trials with prevalence and trials with intensity	19
15	3.2.2 Interest of adding intensity when prevalence is measured in all trials	21
16	3.2.3 Is it better to measure intensity or prevalence in new pest surveys?	22
17	4 Conclusion	24
18	Author contributions	26
19	Funding	26
20	Data availability	26
21	Acknowledgements	26
22	References	26
23	Session information	26

24 1 Introduction

25 In plant epidemiology, pest and disease presence can be measured in a host population using different
 26 metrics. A first metric measures the presence/absence of the pest in the individuals (plants) of the
 27 host population. This metric is often called prevalence or incidence (Madden and Hughes (1999),
 28 Shaw et al. (2018)). The prevalence describes the proportion of the host population in which the pest
 29 is present. This metric is relevant and widely used, but it does not account for the number of pest
 30 individuals per host individual. With the prevalence, a plant infected by one single pest individual
 31 (e.g., an insect) and a plant infected by many pest individuals both represents one infected plant. For
 32 this reason, pest abundance is sometimes assessed using another metric representing the average
 33 number of pest individuals per host individual. This metric is called intensity or severity (Madden and
 34 Hughes (1999), Shaw et al. (2018)), and describes the intensity of the disease in the target population.
 35 These two metrics do not generally have the same requirement in terms of working time; measuring
 36 intensity takes indeed much more time than measuring prevalence because it is very tedious to count
 37 all pest individuals, especially when individuals are small, numerous, and/or difficult to detect.

38 Pest prevalence and intensity are commonly measured in factorial field trials to test the efficacy
 39 of different treatments. In this paper, we place ourselves in an important application framework
 40 which is the evaluation of alternative pesticide treatments to neonicotinoids against aphids in sugar
 41 beet. Indeed, neonicotinoids had been a popular chemical treatment to control aphids for many
 42 years, especially in sugar beets, a major crop in Europe. Recently, neonicotinoids were recognized
 43 as presenting high risks for the environment with a negative impact on a wide range of non-target

44 organisms, including bees (Wood and Goulson (2017), Pisa et al. (2015)), and this family of pesticides
45 has been banned in several European countries. In order to find a substitute to neonicotinoids,
46 a number of factorial field trials were conducted to compare the efficacy of different alternative
47 treatments during several years in different countries. Each trial consists of a set of plots divided
48 into several blocks, themselves divided into several strips on which different pesticide treatments are
49 randomly allocated. One strip always remains untreated to serve as a control. In each strip, aphid
50 prevalence, aphid intensity or both metrics are measured in a sample of plants (usually, 10-20 plants
51 per strip). Depending on the protocol and on the working time constraint, either one type or both
52 types of metrics are measured. Consequently, for a given pest treatment, some trials may report
53 prevalence data while others report intensity data or both types of data. This heterogeneity raises
54 several issues. A first issue concerns the statistical analysis of the trials reporting prevalence and
55 intensity. Although it is easy to fit a generalized linear model to each type of data separately, it is less
56 straightforward to fit a single model to the whole set of trials in order to obtain a single ranking of the
57 pest treatments taking into account both types of data at the same time. Generally, factorial trials
58 assessing treatment efficacy are analyzed with statistical models that take into account one of the two
59 metrics but not both. Prevalence data are thus commonly analyzed using binomial generalized linear
60 models and intensity data are frequently analyzed with Poisson generalized linear models (Michel,
61 Brun, and Makowski (2017), Laurent et al. (2023), Agresti (2015)). As far as we know, no statistical
62 model has been proposed to assess treatment efficacy based on the simultaneous analysis of prevalence
63 and intensity data. In Osiewalski and Marzec (2019), the authors introduced a switching model
64 designed to handle two count variables, one of which may be degenerate. This model was employed to
65 characterize the counts of cash payments and bank card payments in Poland, utilizing data from both
66 cardholders and non-cardholders. A generalized form of the bivariate negative binomial regression
67 model was developed in Gurmu and Elder (2000), allowing for a more flexible representation of the
68 correlation between the dependent variables. This model was applied to describe the number of visits
69 to a doctor and the number of visits to non-doctor health professionals. It outperformed existing
70 bivariate models across various model comparison criteria. In order to analyze data related to crash
71 counts categorized by severity, Park and Lord (2007) employed a multivariate Poisson-lognormal
72 model, effectively addressing both overdispersion and a fully generalized correlation structure within
73 the dataset. However, it should be noted that these models did not include any binomial distribution
74 and thus could not be used to deal with proportion data, such as pest prevalence. Another issue
75 concerns the practical value of combining both prevalence and intensity data. It is unclear whether
76 the simultaneous analysis of prevalence and intensity data may increase the accuracy of the estimated
77 treatment efficacy compared to the use of a single type of data, and whether this may increase the
78 probability of identifying the most efficient treatments. Finally, it is also unclear how future trials
79 should be designed, in particular how many trials are required to obtain accurate estimations, and
80 whether intensity data should be preferred to prevalence data.

81 In this paper, we propose a new flexible statistical model that can be used to rank pest treatments
82 from trials including prevalence data, intensity data, or both. We apply it to a real dataset including
83 trials testing the efficacy of pesticides against aphids infesting sugar beets, considering contrasted
84 scenarios of data availability, and we show how the proposed model can be used to evaluate the
85 efficacy of different treatments. Based on simulations, we then quantify the reduction of mean
86 absolute errors in the estimated treatment efficacies resulting from the use of both prevalence and
87 intensity data during the statistical inference, compared to the use of exclusively either prevalence
88 or intensity data. The rest of the paper is organized as follows. First, we present the structure of the
89 dataset including real prevalence and intensity data. Next, we describe in detail the proposed model,
90 the inference method, and the simulation strategy. After checking the convergence of the fitting
91 algorithm, we show how the model can be used to assess treatment efficacy. We finally present the
92 results based on simulated data and we make several recommendations.

93 **2 Material and Method**

94 **2.1 Description of the data**

95 Data were collected in 32 field trials conducted in France, Belgium and the Netherlands to compare
96 several treatments against aphids in sugar beets. Each trial consists in a plot located in a given site at
97 a given year (site-year) divided into one to four blocks. Each of these blocks is itself divided into
98 strips where different treatments are tested, one of these treatments being an untreated control and
99 the others corresponding to different types of insecticide. In each strip of each block, the number
100 of aphids was counted on a sample of 10 beet plants (intensity). The number of infested plants
101 (prevalence) was measured as well, but only in 15 trials out of 32. The total numbers of intensity
102 and prevalence data are equal to 1128 and 561, respectively. Note that the number of aphids was not
103 counted on each beet plant but in the whole plant sample. Intensity and prevalence were monitored
104 at different times after treatments. As shown in Figure 1 A, the dataset is unbalanced as less data are
105 available for the treatment Mavrik-jet compared to the others. Figure 1B shows that the intensity
106 and prevalence tend to increase with time.

107 **2.2 Model**

108 **2.2.1 Specification**

109 We introduce an unobserved variable representing the number of pest individuals (here, aphids) on
110 each plant in a sample of N plants (here, sugar beets). This variable is noted W and is assumed to
111 follow a Poisson probability distribution whose mean value is a function of time.

112 We use the following indexes: i for the trial, j for the treatment, k for the block, t for the time and s
113 for the plant number. The distribution of W_{ijklts} is defined as:

$$W_{ijklts} \sim \mathcal{P}(\lambda_{ijklt}) \quad (1)$$

$$\log \lambda_{ijklt} = \alpha_0 + \beta_{0i} + \gamma_{0j} + (\alpha_1 + \gamma_{1j}) X_t + u_{ij} + \epsilon_{ijklt} \quad (2)$$

114 with

- 115 • $\beta_{0i} \sim \mathcal{N}(0, \sigma_0^2)$
116 • $u_{ij} \sim \mathcal{N}(0, \chi^2)$
117 • $\epsilon_{ijklt} \sim \mathcal{N}(0, \eta^2)$

118 The random variables are all assumed independent. The parameters α_0 , α_1 , γ_{0j} , and γ_{1j} are considered
119 as fixed. This model serves as a tool for conducting inference on a population of trials, from which the
120 subset of trials comprising our dataset is assumed to constitute a random sample. In essence, the trials
121 contained within our database are leveraged to estimate the parameter values that characterize a target
122 population, where the tested pest control treatments will be actually implemented. Consequently, all
123 parameters contingent on individual trials have been defined as random effects.

124 The observed variables (intensity and prevalence) can be expressed as a function of W . We note:

- 125 • Y_{ijklt} the number of pest individuals (aphids) in the sample of N_i plants collected in trial i ,
126 treatment j , block k , at time t
127 • Z_{ijklt} the number of infested plants among the N_i plants collected in trial i , treatment j , block k ,
128 at time t

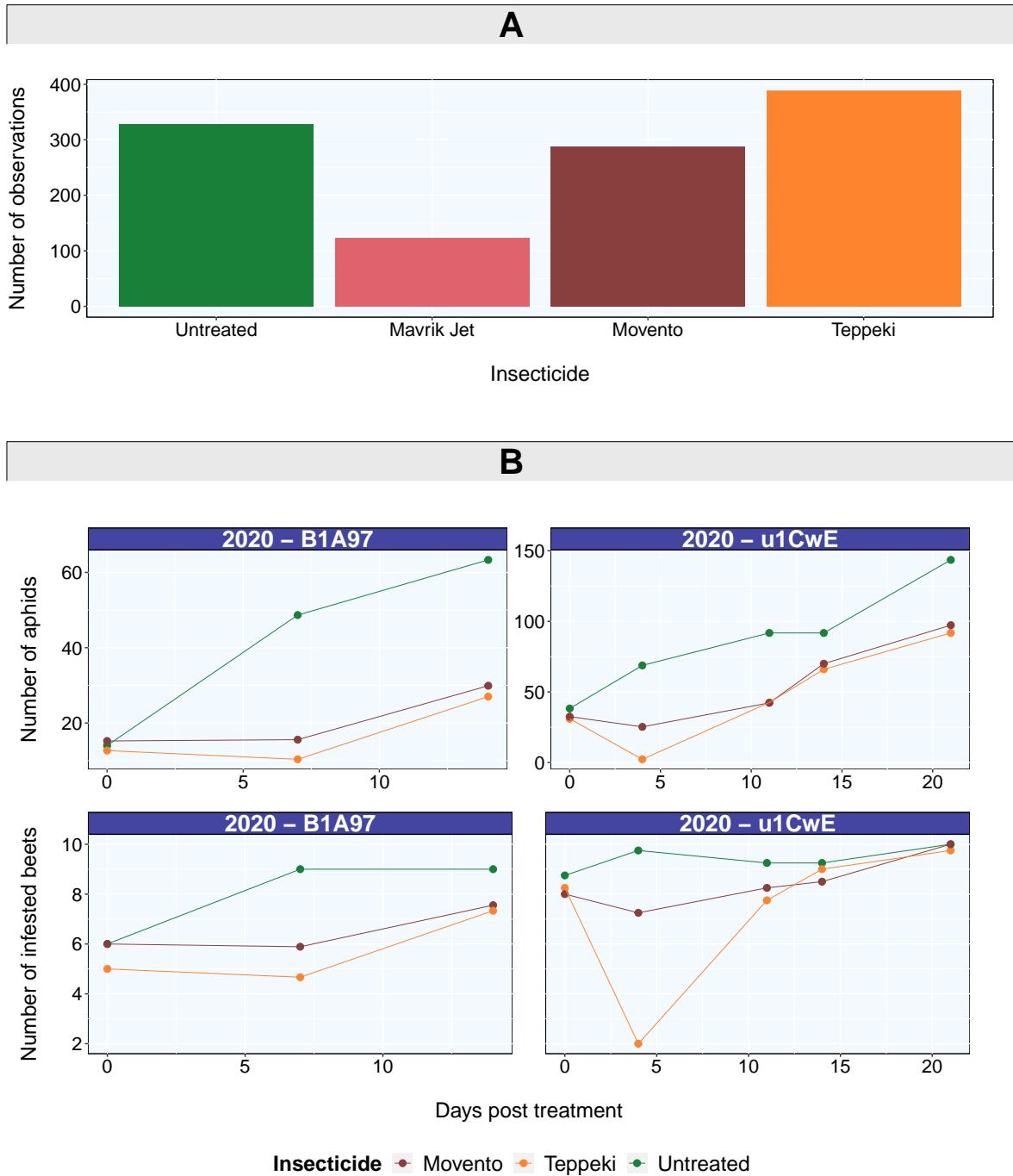


Figure 1: Description of the dataset. **A** Number of observations according to the type of insecticide. **B** Examples of observed number of aphids averaged over the blocks (intensity) and number of infested beets out of ten (prevalence) averaged over the blocks, at different dates for two trials.

¹²⁹ Then, assuming the Ws are independent, we have:

$$Y_{ijkt} = \sum_{s=1}^{N_i} W_{ijkts} \quad Y_{ijkt} \sim \mathcal{P}(N_i \lambda_{ijkt}) \quad (3)$$

$$Z_{ijkt} = \sum_{s=1}^{N_i} \mathbf{1}_{W_{ijkts}>0} \quad Z_{ijkt} \sim \mathcal{B}(N_i, \pi_{ijkt}) \quad (4)$$

¹³⁰ where $\pi_{ijkt} = 1 - \exp(-\lambda_{ijkt})$. The different quantities used in the model are defined in Table 1.

Table 1: Description of the indices, inputs and parameters used in the model

i	trial index
j	treatment index
k	block index
t	time index
s	plant index
N_i	sample size (number of plants)
λ_{ijkt}	mean number of pest individuals (aphids) on one plant
π_{ijkt}	probability for a plant to be infested
α_0	mean number of pest individuals (aphids) in the untreated group
β_{0i}	trial effect
γ_{0j}	effect of treatment j at time 0 (time of treatment)
α_1	growth parameter of the number of pest individuals for the untreated group
γ_{1j}	effect of the treatment on the time effect (interaction between treatment j and time)
X_t	number of days post treatment
u_{ij}	random interaction between trial and treatment
ϵ_{ijkt}	residuals

¹³¹ From this model we defined the efficacy of the j th treatment at time t (t days after pesticide application)
¹³² by the quantity (Laurent et al. (2023)):

$$\text{Ef}_{jt} = \left(1 - \exp(\gamma_{0j} + \gamma_{1j} \times X_t) \right) \times 100 \quad (5)$$

¹³³ The quantity Ef_{jt} corresponds to the expected percentage reduction of pest individuals (aphid numbers)
¹³⁴ for the j th treatment compared to the untreated group, over trials and blocks.

¹³⁵ Our Poisson log linear model includes an additive random dispersion term associated to each
¹³⁶ individual observation (ϵ_{ijkt} in Equation 2). This is a standard and well-recognized approach to deal
¹³⁷ with over-dispersion (Harrison (2014)). In order to check the model assumptions, we performed
¹³⁸ a posterior predictive check of our model to check that the data were compatible with the model
¹³⁹ assumptions. Posterior predictive check is frequently used to look for systematic discrepancies
¹⁴⁰ between real and simulated data (Gelman et al. (1995)). To do so, we computed the probability of
¹⁴¹ exceeding each individual data with the fitted model (Equation 2). The computed probabilities were
¹⁴² all falling in the range 0.22-0.93 (except for the observations equal to 0, for which the probability
¹⁴³ of being greater was equal to 1), and were thus not extreme. This result indicates that the model
¹⁴⁴ specified is not incompatible with the observed data and that the over-dispersion was correctly taken
¹⁴⁵ into account. In addition, we fitted another model including a negative binomial distribution instead
¹⁴⁶ of a Poisson distribution. The results were almost identical between the two models.

147 2.2.2 Inference on real data

148 The model parameters were estimated using Bayesian inference with a Markov chain Monte-Carlo
149 method. We performed the inference using R, with the package rjags (Plummer (2022)). For each of
150 the six dataset listed in Table 2, we fitted the model (Equation 2 - Equation 4) with the following weakly
151 informative priors: $\mathcal{N}(0, 10^3)$ for the parameters $\alpha_0, \gamma_0, \alpha_1, \gamma_1$ and $\mathcal{U}([0, 10])$ for the parameters
152 σ_0, χ, η . We used two Markov chains with 2×10^5 iterations (after an adapting phase of 10^5 iterations),
153 and we centered the time variable t to facilitate convergence.

154 The convergence of the MCMC algorithm was checked by inspecting the mixing of the two Markov
155 chains and monitoring the Gelman-Rubin diagnosis statistics (Gelman and Rubin (1992)). We then
156 computed the posterior mean of the pesticide treatment efficacy (defined by Equation 5) as well as its
157 95% credibility interval. The code used to fit the model is provided below.

i Note

The following code presents the inference on an extract of the real dataset, which includes both trials of figure 1B (2020 - B1A97 ; 2020 - u1CwE). It is a demo for the “50% Y - 50% Z” scenario and we set here the number of adaptation and iteration to 2000 in order to reduce computation time.

158

```
# Jags code for the model #####
modelstringYZ = "
model {

    # Likelihood #####
    for (i in 1:Q){
        Y[i] ~ dpois(N[i] * lb[i])
        Z[i] ~ dbinom(pi[i], N[i])

        log(lb[i]) = beta0[ID[i]] + gamma0[INSEC[i]] + (alpha1 +
            gamma1[INSEC[i]]) * TIME[i] + u[ST[i]] + epsi[i]

        pi[i] = 1 - exp(- lb[i])
        epsi[i] ~ dnorm(0, pi_eps)
    }

    for (j in 1:K){
        beta0[j] ~ dnorm(alpha0, tau0)
    }

    for (c in 1:M){
        u[c] ~ dnorm(0, invchi)
    }

    gamma0[1] = 0
    gamma1[1] = 0

    # Priors #####
    for (s in 2:L){
        gamma0[s] ~ dnorm(0, 0.001)
    }
}
```

```

        gamma1[s] ~ dnorm(0, 0.001)
    }

alpha0 ~ dnorm(0, 0.001)
alpha1 ~ dnorm(0, 0.001)
sigma0 ~ dunif(0, 10)
chi ~ dunif(0, 10)
eta ~ dunif(0, 10)

# Derived Quantities #####
tau0 = pow(sigma0, -2)
invchi = pow(chi, -2)
pi_eps = pow(eta, -2)

for (h in 2:L){
    for(t in 1 : T){
        Eff[h, t] = (1 - exp(gamma0[h] + gamma1[h] *
            TIME_unique[t])) * 100
    }
}
}

""

writeLines(modelstringYZ, con = "Files_for_code/modelYZ.txt")
#####
##### Inference example on the extract of the real dataset #####
real_data_extract = readRDS(file = "data/real_data_extract.rds")

data = real_data_extract %>%
    mutate(tscaled = scale(DPT), st = paste(ID, Insecticide))

# Building scenarios -----
scenarioY = data %>% mutate(Z = NA)
scenarioYhalfZhalf = data %>% mutate(Y = ifelse(ID == "2020 - B1A97", Y, NA),
                                         Z = ifelse(ID == "2020 - B1A97", NA, Z))
scenarioYhalf = data %>% filter(ID == "2020 - B1A97") %>% mutate(Z = NA)
scenarioZhalf = data %>% filter(ID == "2020 - u1CwE") %>% mutate(Y = NA)
# -----

data = scenarioYhalfZhalf

Y = data$Y; Q = length(Y); N = data$N; Z = data$Z

ID = as.numeric(as.factor(as.character(data$ID)));
INSEC = as.numeric(as.factor(as.character(data$Insecticide)));
TIME = as.numeric(data$tscaled);
ST = as.numeric(as.factor(as.character(data$st)));

```

```

K = length(unique(ID)); L = length(unique(INSEC));
M = length(unique(ST));

df_TIME = suppressMessages(data %>%
                           group_by(DPT, tscaled) %>%
                           summarise(n = n()) %>% as.data.frame)

TIME_unique = approx(df_TIME$DPT, df_TIME$tscaled, xout = c(6, 12))$y;
T = length(unique(TIME_unique))

data_jags = list(
  "Y" = Y, "Z" = Z, "Q" = Q, "ID" = ID, "INSEC" = INSEC,
  "TIME" = TIME, "ST" = ST, "K" = K, "L" = L, "M" = M,
  "N" = N, "T" = T, "TIME_unique" = TIME_unique
)

nadapt = 2000; niter = 2000

model <- jags.model("Files_for_code/modelYZ.txt", data = data_jags,
                     n.chains = 2, n.adapt = nadapt)

samples <- coda.samples(model,
                        variable.names = c("gamma0", "gamma1", "Eff"),
                        n.iter = niter, thin = 10)
#####

```

159
160 In practice, it is common that only Y or Z data are available in some of the trials. In this case, the
161 resulting dataset includes observations of Y in some trials and observations of Z in others. The dataset
162 may even include one type of observations only, either Y or Z , in all trials. Here, we define four
163 scenarios with contrasted levels of Y and Z availability in order to evaluate the consequences of using
164 different types of datasets. We consider four data subsets defined from the real dataset including
165 trials with observations of Y , with observations of Z , or with both types of observation in different
166 proportions (Table 2). The data subset “100% Y - 0% Z ” includes Y data collected in the 32 trials. The
167 data subset “50% Y - 0% Z ” includes Y data collected in the 17 trials for which no Z observation is
168 available. The data subset “0% Y - 50% Z ” includes the Z data collected in the 15 trials for which Z
169 observations are available. The data subset “50% Y - 50% Z ” includes Y data collected in 17 trials and
170 Z data collected in the other 15 trials. The latter data subset does not include any trial reporting both
171 Y and Z data. Throughout our analysis, missing data were assumed to be missing at random.
172 The hierarchical model defined above is fitted to each dataset in turn. Each fitted model is then used
173 to compute the posterior mean and 95% credibility interval of Ef_{jt} for each treatment at $t=6$ and 12
174 days after pesticide application.

Table 2: Four data subsets defined from the original dataset (real data).

Type of dataset	Description
100% Y - 0% Z	Y observations available in the 32 trials and no Z
50% Y - 0% Z	Y observations available in 17 trials and no Z
0% Y - 50% Z	Z observations available in 15 trials and no Y

Type of dataset	Description
50% Y - 50% Z	Y observations available in 17 trials and Z observations available in the other 15 trials

175 2.2.3 Simulations

176 Simulations are carried out to further investigate the impact of the type and amount of data available
 177 on the accuracy of the parameters and the ability of the model to identify the most and least effective
 178 treatments.

179 We define three numbers of trials, equal to 20, 40 and 80, successively. For data simulations, the
 180 model parameters are set equal to those estimated from the real dataset “100% Y - 0% Z” defined
 181 in Table 2 (posterior means). For each number of trials, we generate virtual data from the model
 182 (Equation 1 - Equation 4) and estimate the model parameters, according to the following procedure:

- 183 • Draw values of β_{0i} , u_{ij} and ϵ_{ijkt} in their distributions for each trial, 3 treatments (+ the untreated
 184 control), 3 dates ($t=0, 6, 12$), and 4 blocks
- 185 • Calculate λ_{ijkt} from Equation 2,
- 186 • Draw values of W_{ijkt} in its Poisson distribution for 10 plants ($s=1, \dots, 10$)
- 187 • Calculate Y_{ijkt} and Z_{ijkt} from the Ws for each trial, treatment, date, block.
- 188 • Generate the eight data subsets corresponding to the scenarios defined in Table 3 (including
 189 all values of W, the generated values of Y only, the generated values of Z only, both Y and Z
 190 values but not W, the values of Y in 50% of the trials, the values of Z in 50% of the trials, the
 191 values of Y in 50% of the trials and the values of Z in the other 50%),
- 192 • Fit the model (Equation 1 - Equation 4) to each of the data subsets according to the procedure
 193 described above based on MCMC.

194 At the end of this procedure, we get eight sets of estimated parameters, corresponding to the eight
 195 scenarios defined in Table 3. This procedure is repeated 1000 times with each time a different seed
 196 between 0 and 999. However, the computations performed with jags failed for 26 replicates and thus
 197 974 replicates were available for the analysis.

198 For each number of trials and each scenario defined in Table 3, the accuracy of the estimated
 199 parameters γ (from which depend the treatment efficacies) is evaluated by computing an absolute
 200 error, averaged over the three treatments ($j=1$ corresponding to the control) as:

$$E_\gamma = \frac{1}{2 \times 3} \sum_{j=2}^4 \left(\frac{|\gamma_{0j} - \hat{\gamma}_{0j}|}{|\gamma_{0j}|} + \frac{|\gamma_{1j} - \hat{\gamma}_{1j}|}{|\gamma_{1j}|} \right) \quad (6)$$

201 where the true parameter values are set equal to the posterior means computed with the real dataset,
 202 and the parameter estimates ($\hat{\gamma}_{0j}$ and $\hat{\gamma}_{1j}$) are the posterior means computed for the j th treatment. For
 203 each trial number and each scenario defined in Table 3, the 974 values of E_γ obtained for the 974
 204 generated data subsets are then averaged. The average values obtained for the eight scenarios are
 205 finally compared to determine the type of data leading to the most accurate estimated parameter
 206 values.

207 In addition, we compare the eight scenarios according to another criterion measuring the difference
 208 between the estimated efficacy values and the true efficacy value (averaged over the three pesticide
 209 treatments considered), as follows:

$$E_{\text{Ef}_t} = \frac{1}{3} \sum_{j=2}^4 \frac{|\text{Ef}_{jt} - \hat{\text{Ef}}_{jt}|}{|\text{Ef}_{jt}|} \quad (7)$$

210 where the true treatment efficacy is defined by Equation 5 (setting the parameters γ_s to the posterior
 211 means obtained with the real dataset) and the estimated efficacy (Ef_{jt}) is the posterior mean computed
 212 with the simulated dataset. The 974 values of E_{Ef_t} obtained from the 974 simulated datasets are then
 213 averaged for each trial number and each scenario. Finally, we evaluate the proportions of cases
 214 where the true best treatment (i.e., the treatment with the highest efficacy) is correctly identified.

215 In order to determine whether the difference of performance obtained with the types of data Y and Z
 216 depends on the pest abundance, we perform an additional series of simulations with three values
 217 for the model parameter α_0 , equal to -1 , 1 and 2 , successively. These three values of α_0 define three
 218 contrasted levels of pest abundance (the higher α_0 , the higher the abundance). We used the procedure
 219 outlined above considering only two scenarios of data availability, namely “ $100\% Y - 0\% Z$ ” and “ 0%
 220 $Y - 100\% Z$ ”. This procedure is implemented with each value of α_0 in turn. The results are used to
 221 compare the model performances using either Y or Z for parameter estimation, depending of the pest
 222 abundance specified by α_0 .

Table 3: Eight scenarios compared using simulated data.

Type of dataset	Description
100% W	W observations available in all the trials
100% Y - 0% Z	Y observations available in all the trials
0% Y - 100% Z	Z observations available in all the trials
100% Y - 100% Z	Y and Z observations available in all the trials
50% Y - 0% Z	Y observations available in half of the trials
0% Y - 50% Z	Z observations available in half of the trials
50% Y - 50% Z	Y observations available in half of the trials and Z observations available in the other half of the trials
50% Y - 100% Z	Y observations available in half of the trials and Z observations available in all the trials

223 These simulations required a significant amount of computation time and were conducted on the
 224 INRAE MIGALE server. For a single seed and 20 trials, the computations took 80 minutes on a
 225 computer with an Intel Core i7 processor running at 1.90 GHz and 32 GB of RAM. The code used to
 226 perform them is presented below.

```
simu_data <- function(seed, I){

  set.seed(seed)

  ID = c(1 : I);

  data = expand_grid(ID = c(1 : I),   Block = c(1 : K),   Band = c(1 : J),
                     Beet = c(1 : N),   DPT = seq(0, 12, length.out = T))

  data = data %>%
    mutate(Insecticide = ifelse(DPT <= 0, "T 1", paste("T", Band)),
          tscaled = scale(data$DPT)) %>%
```

```

    mutate(st = paste(ID, Insecticide),
           sbit = paste(ID, Block, Insecticide, DPT))

Insecticide = data$Insecticide %>% unique %>% sort
names(gamma0) = Insecticide
names(gamma1) = Insecticide

beta0 = rnorm(I, sd = sig0);
u = rnorm(I * J, sd = chi);
epsi = rnorm(I * J * K * (T - 1) + I * K, sd = eta);

names(beta0) = data$ID %>% unique;
names(u) = data$st %>% unique
names(epsi) = data$sbit %>% unique

data = data %>% mutate(alpha0 = alpha0,
                        alpha1 = alpha1,
                        N = N,
                        beta0 = recode(ID, !!!beta0),
                        gamma0 = recode(Insecticide, !!!gamma0),
                        gamma1 = recode(Insecticide, !!!gamma1),
                        u = recode(st, !!!u),
                        epsi = recode(sbit, !!!epsi))

data$lb = exp(data$alpha0 + data$beta0 + data$gamma0 + (data$alpha1 +
                                                       data$gamma1) * data$tscaled + data$u + d)

data$W = sapply(c(1 : (I * J * K * T * N)),
                 function(x) rpois(1, data$lb[x]))

dataYZ = data %>% group_by(ID, Block, Band, DPT) %>%
  summarise(Insecticide = unique(Insecticide),
            tscaled = unique(tscaled), st = unique(st), N = unique(N),
            Y = sum(W), Z = sum(W > 0)) %>% as.data.frame

dataW = data %>% select(- alpha0, - alpha1, - beta0, - gamma0, - gamma1,
                         - u, - epsi, - lb)

return(list("dataYZ" = dataYZ, "dataW" = dataW))
}

```

227

228

Note

The following code presents the inference for one seed. As the computing time is growing fast with the number of trials and the number of sampled values, we set here the number of trials to 10 ($I = 10$) and the number of adaptations and iterations to 2000.

Our simulation results are obtained by running this code for $I = 10, 20, 40, 80$, for all seeds

229

230

between 0 and 999, with a number of adaptations and iterations equal to 36000. Scenarios “50% Y - 0% Z” and “0% Y - 50% Z” are obtained from scenarios “100% Y - 0% Z” and “0% Y - 100% Z”, respectively. For example “50% Y - 0% Z” scenario with 40 trials corresponds to “100% Y - 0% Z” with 20 trials.

```
J = 4; T = 3; N = 10; K = 4;

Block = c(1 : K); Band = c(1 : J);
Beet = c(1 : N); DPT = seq(0, 12, length.out = T);

alpha0 = 0.5; gamma0 = c(0, - 0.13, - 1.13, - 1.24);
alpha1 = 0.16; gamma1 = c(0, 0.24, - 0.14, - 0.15);
sig0 = 1.87; chi = 0.27; eta = 0.98

#####
# Building scenarios #####
seed = 1; I = 10
data = suppressMessages(simu_data(seed = seed, I = I))

scenarios = list(
  Y = data$dataYZ %>% mutate(Z = NA),
  Z = data$dataYZ %>% mutate(Y = NA),
  YhalfZhalf = data$dataYZ %>% mutate(Y = ifelse(ID <= (I / 2), NA, Y),
                                         Z = ifelse(ID > (I / 2), NA, Z)),
  YhalfZ = data$dataYZ %>% mutate(Y = ifelse(ID <= (I / 2), NA, Y)),
  YZ = data$dataYZ,
  W = data$dataW
)
#####

# Inference #####
nadapt = 2000; niter = 2000

res_inf = NULL

# Inference YZ -----
ID = as.numeric(as.factor(as.character(scenarios$Y$ID)));
INSEC = as.numeric(as.factor(as.character(scenarios$Y$Insecticide)));
TIME = as.numeric(scenarios$Y$tscaled);
ST = as.numeric(as.factor(as.character(scenarios$Y$st)));

K = length(unique(ID));
L = length(unique(INSEC));
M = length(unique(ST))

TIME_unique = unique(scenarios$Y$tscaled)[2 : 3]
T = length(unique(TIME_unique))
```

```

for(i in (1 : 5)){
  Y = scenarios[[i]]$Y; Z = scenarios[[i]]$Z;
  N = scenarios[[i]]$N; Q = length(Y)

  data_jags = list(
    "Y" = Y, "Z" = Z, "Q" = Q, "ID" = ID, "INSEC" = INSEC,
    "TIME" = TIME, "ST" = ST, "K" = K, "L" = L, "M" = M,
    "N" = N, "T" = T, "TIME_unique" = TIME_unique
  )

  model <- jags.model("Files_for_code/modelYZ.txt", data = data_jags,
                       n.chains = 2, n.adapt = nadapt)

  samples <- coda.samples(model,
                          variable.names = c("gamma0", "gamma1", "Eff"),
                          n.iter = niter, thin = 10)

  bind = list(samples);
  names(bind) = paste("samples", names(scenarios)[i], sep = "_")

  res_inf = res_inf %>% append(bind)
}

# Inference W -----
ID = as.numeric(as.factor(as.character(scenarios$W$ID)));
INSEC = as.numeric(as.factor(as.character(scenarios$W$Insecticide)));
TIME = as.numeric(scenarios$W$tscaled);
ST = as.numeric(as.factor(as.character(scenarios$W$st)));
SBIT = as.numeric(as.factor(as.character(scenarios$W$sbit)))

K = length(unique(ID)); L = length(unique(INSEC));
M = length(unique(ST)); X = length(unique(SBIT))

TIME_unique = unique(scenarios$W$tscaled)[2 : 3]
T = length(unique(TIME_unique))

W = scenarios$W$W; Q = length(W)

data_jags = list(
  "W" = W, "Q" = Q, "ID" = ID, "INSEC" = INSEC,
  "TIME" = TIME, "ST" = ST, "SBIT" = SBIT, "K" = K, "L" = L,
  "M" = M, "X" = X, "T" = T, "TIME_unique" = TIME_unique
)

model <- jags.model("Files_for_code/modelW.txt", data = data_jags,
                     n.chains = 2, n.adapt = nadapt)

samples <- coda.samples(model,
                        variable.names = c("gamma0", "gamma1", "Eff"),
                        n.iter = niter, thin = 10)

```

```

res_inf = res_inf %>% append(list("samples_W" = samples))
#####
# Formatting results #####
t = scenarios$Y$tscaled %>% unique
Eff_6_true = (1 - exp(gamma0[2 : J] + gamma1[2 : J] * t[2])) * 100
Eff_12_true = (1 - exp(gamma0[2 : J] + gamma1[2 : J] * t[3])) * 100
truth = c(Eff_6_true, Eff_12_true, gamma0[2 : J], gamma1[2 : J])

n_scenarios = length(scenarios)

esti = lapply(
  res_inf, function(x) summary(x)$statistics %>% as.data.frame %>%
    rownames_to_column %>%
    filter(!(grepl("gamma", rowname) & Mean == 0)) %>%
    select(Mean) %>% as.matrix %>% as.vector
)

b_inf = lapply(
  res_inf, function(x) summary(x)$quantiles %>% as.data.frame %>%
    rownames_to_column %>%
    filter(!(grepl("gamma", rowname) & `2.5%` == 0)) %>%
    select(`2.5%`) %>% as.matrix %>% as.vector
)

b_sup = lapply(
  res_inf, function(x) summary(x)$quantiles %>% as.data.frame %>%
    rownames_to_column %>%
    filter(!(grepl("gamma", rowname) & `2.5%` == 0)) %>%
    select(`97.5%`) %>% as.matrix %>% as.vector
)

parameters = summary(res_inf[[1]])$statistics %>% as.data.frame %>%
  rownames_to_column %>%
  filter(!(grepl("gamma", rowname) & Mean == 0)) %>%
  select(rowname) %>% as.matrix %>% as.vector

res = do.call(rbind,
  lapply(c(1 : n_scenarios),
    function(x)
      data.frame(Truth = truth, Scenario = names(scenarios)[x],
        I = I, seed = seed,
        value = esti[[x]], parameters = parameters,
        b_inf = b_inf[[x]], b_sup = b_sup[[x]])
    )
  )
#####

```

231 **3 Results**

232 **3.1 Results obtained with real data**

233 We present here the results obtained with real data. First, we check the convergence of the model for
234 the scenarios defined in Table 2. We then compare the estimated values and the credibility intervals
235 of the treatment efficacy obtained in the different scenarios.

236 **3.1.1 Model convergence and posterior distributions**

237 Figure 2 presents the Markov chains associated with the model parameters and treatment efficacies for
238 the different scenarios. The x-axis presents the iteration number and the y-axis presents the sampled
239 value. Results show that the chains are well mixed. Figure 3 presents the Gelman-Rubin statistics
240 associated with the model parameters and treatment efficacies as a function of the iterations, for the
241 different scenarios. We observe that this statistic converges to 1, which indicates the convergence
242 of the algorithm.

243 Table 4 gives a summary of the posterior distributions of the model parameters and treatment
244 efficacies obtained for the “100% Y - 0% Z” scenario. The significantly positive value of α_1 indicates
245 that the aphid numbers tend to increase with time in untreated plots. The relatively high value
246 of σ_0 (posterior mean equal to 1.87) reveals a strong variability in aphid numbers between trials.
247 The posterior mean value of χ (0.27) suggests that the treatment efficacy varies across trials. The γ_0
248 parameter is negative for all three treatments, indicating a negative effect of the treatments on the
249 aphid numbers at the time of pesticide spray. The Movento and Teppeki treatments have a similar
250 effect with a posterior mean for γ_0 equal to -1.13 and -1.24, and a standard deviation equal to 0.12
251 and 0.11, respectively. The effect of treatment Mavrik Jet is weaker as its posterior mean for γ_0 is
252 equal to -0.13 and its standard deviation is equal to 0.16. The γ_1 posterior means are negative for
253 Movento and Teppeki (-0.14 and -0.15), suggesting that the effect of these treatments tend to increase
254 with time, but the posterior mean value is positive for Mavrik Jet (= 0.24), suggesting that the effect
255 of this treatment may decrease with time. However, the 95% credibility intervals of γ_1 include zero
256 and these parameters are not very accurately estimated.

257

258

259

260

261 **3.1.2 Estimated values of pesticide treatment efficacies**

262 Figure 4 presents the posterior means and the 95% credibility intervals of treatment efficacies at
263 6 days (A) and 12 days (B) after pesticide spray, for the “100% Y - 0% Z”, “50% Y - 50% Z”, “50% Y
264 - 0% Z” and “0% Y - 50% Z” scenarios. Different scenarios are indicated by different colors. The
265 x-axis presents the efficacy and the y-axis presents the treatments. Overall, the results obtained are
266 consistent across scenarios; Teppeki and Movento show higher mean efficacies than Mavrik Jet, and
267 the credibility intervals are narrower for Teppeki and Movento than for Mavrik Jet in all scenarios.
268 The credibility interval of the “100% Y - 0% Z” scenario is narrower than that of the “50% Y - 50%
269 Z” scenario, which is itself narrower than that of the “50% Y - 0% Z” and “0% Y - 50% Z” scenarios.
270 Table 5 quantifies the differences in the sizes of the credibility intervals and present the reduction
271 percentage in the credibility interval sizes of the “100% Y - 0% Z”, “50% Y - 50% Z” and “0% Y - 50% Z”
272 scenarios, compared to the “50% Y - 0% Z” scenario, for all the treatment efficacies at 6 days (third
273 column) and at 12 days (fourth column) after pesticide spray. A positive (negative) value corresponds
274 to an increase (decrease) of the credibility interval size compared to “50% Y - 0% Z”. For example,

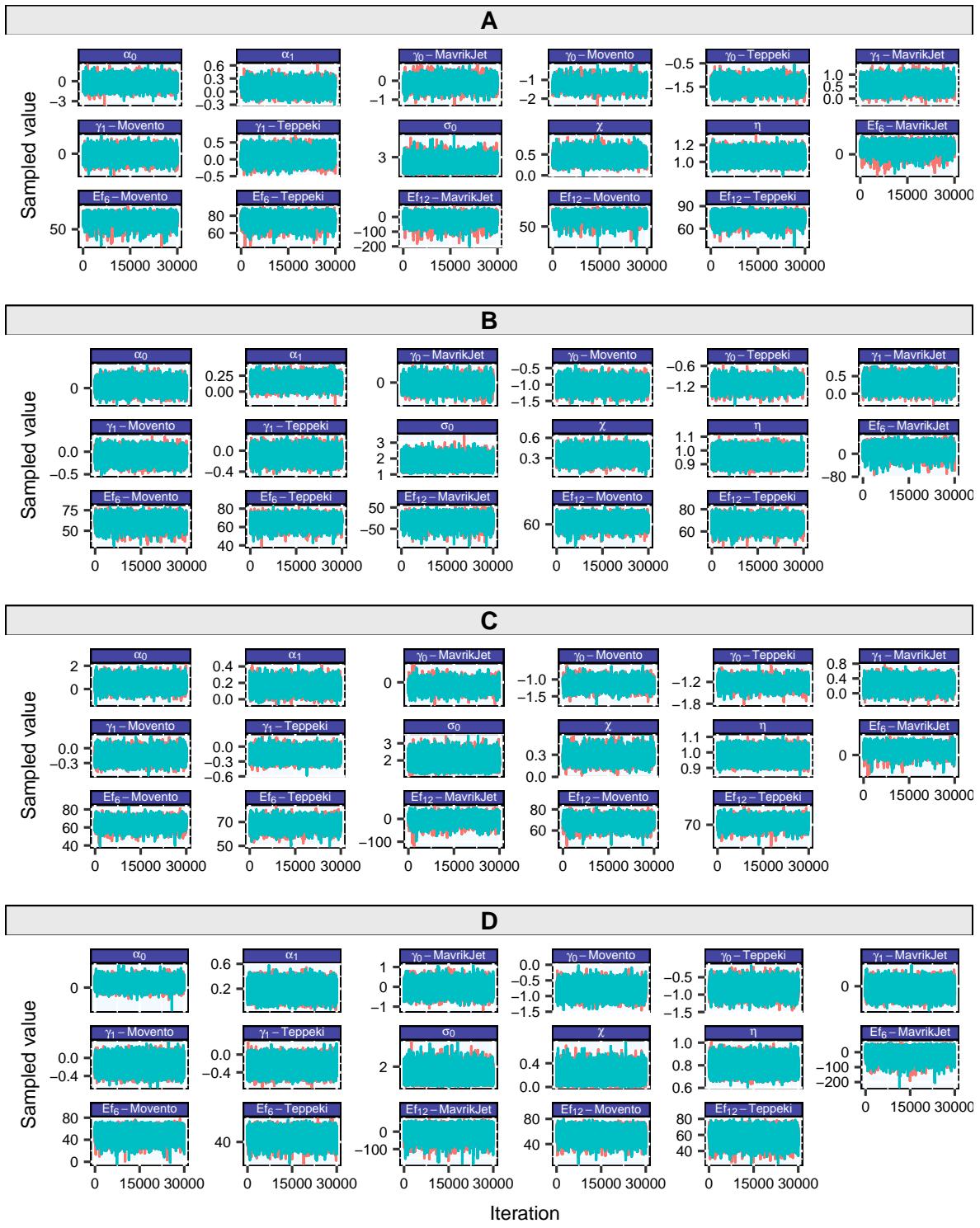


Figure 2: Model convergence - Markov chain for the model parameters and treatment efficacies, in the scenarios ‘50% Y - 0% Z’ (**A**), ‘50% Y - 50% Z’ (**B**), ‘100% Y - 0% Z’ (**C**) and ‘0% Y - 50% Z’ (**D**). The x-axis presents the iteration number and the y-axis presents the sampled value.

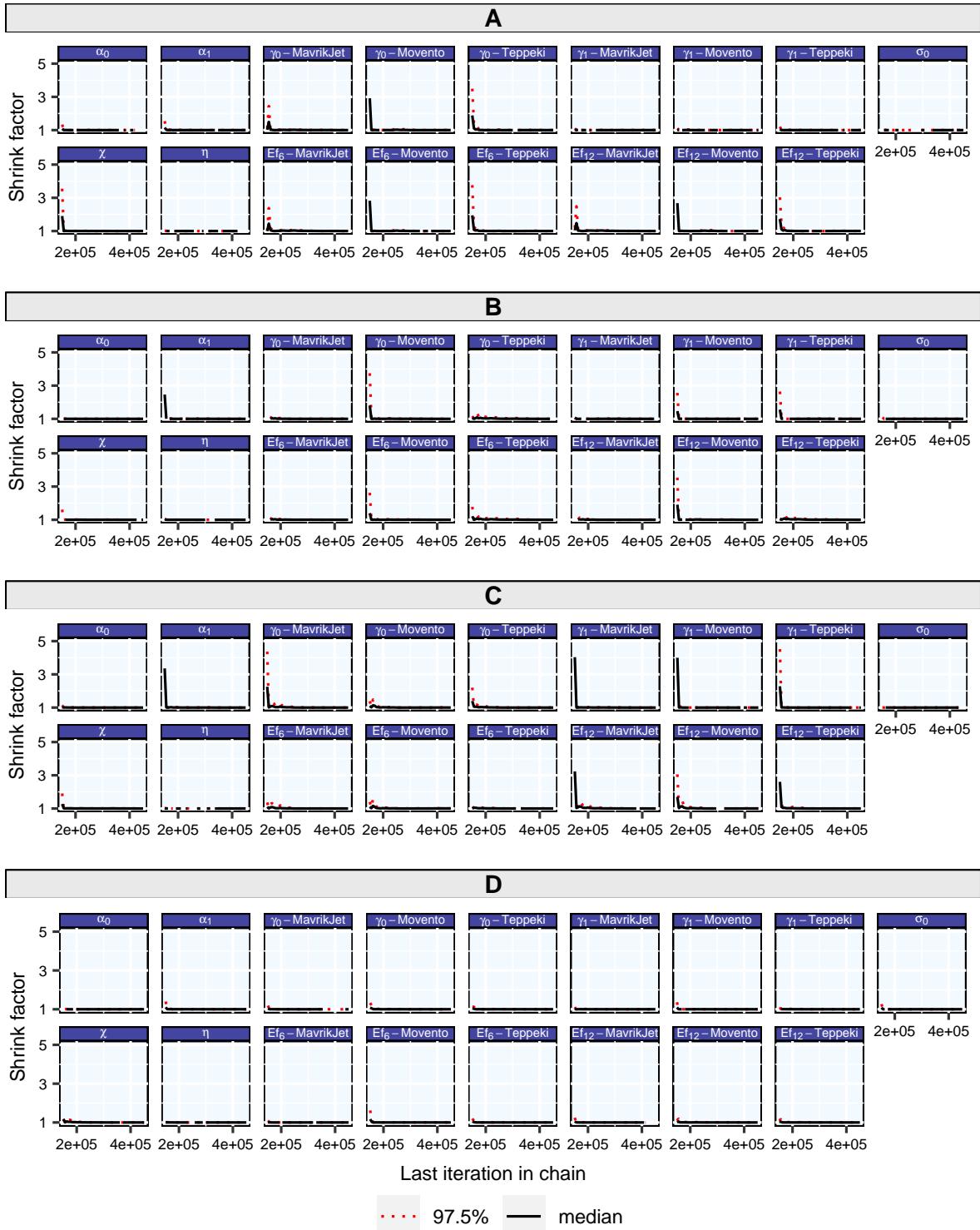


Figure 3: Model convergence - Gelman-Rubin statistics for the parameters of the model (Equation 1 - Equation 4) and for treatment efficacies (Equation 5), according to the scenarios ‘50% Y - 0% Z’ (**A**), ‘50% Y - 50% Z’ (**B**), ‘100% Y - 0% Z’ (**C**) and ‘0% Y - 50% Z’ (**D**). The x-axis presents the iteration number and the y-axis presents the Gelman-Rubin statistic.

Table 4: Summary of the posterior distributions obtained with the 100% Y - 0% Z scenario for the model parameters and treatment efficacies: posterior mean, standard deviation, 2.5 and 97.5 quantiles.

Parameter	Mean	SD	2.5%	97.5%
α_0	0.49	0.34	-0.19	1.17
α_1	0.16	0.06	0.04	0.29
σ_0	1.87	0.26	1.44	2.46
χ	0.28	0.06	0.15	0.40
η	0.98	0.03	0.92	1.03
γ_0 - untreated	0.00	0.00	0.00	0.00
γ_0 - Mavrik Jet	-0.12	0.17	-0.46	0.20
γ_0 - Movento	-1.13	0.12	-1.37	-0.88
γ_0 - Teppeki	-1.24	0.11	-1.46	-1.02
γ_1 - untreated	0.00	0.00	0.00	0.00
γ_1 - Mavrik Jet	0.24	0.13	-0.01	0.49
γ_1 - Movento	-0.15	0.10	-0.33	0.04
γ_1 - Teppeki	-0.15	0.09	-0.32	0.02

275 the credibility interval size obtained for the efficacy of Movento at 6 days is 40.5% smaller with the
 276 “100% Y - 0% Z” scenario than with the “50% Y - 0% Z” scenario. Overall, the credibility interval sizes
 277 obtained with the “100% Y - 0% Z” scenario are 25 to 45% smaller than those obtained with the “50% Y
 278 - 0% Z” scenario. This indicates that the accuracy of the estimated treatment efficacies is improved by
 279 doubling the number of Y observations. Results show that the credibility intervals are often larger
 280 with “0% Y - 50% Z” than with “50% Y - 0% Z”, suggesting that better esimtations are obtained using
 281 Y than Z when only a limited number of data is available, at least in this case study. Interestingly,
 282 the sizes of the credibility intervals are smaller by about 25% with “50% Y - 50% Z” compared to
 283 “50% Y - 0% Z”, showing that the combination of Y and Z observations collected in different trials in
 284 beneficial and lead to a reduction of the uncertainty in the estimated treatment efficacy. The latter
 285 result illustrates how the estimation of treatment efficacy could be improved by combining trials
 286 with prevalence data and trials with intensity data.

```
#|echo: false
cbind(df_len6, df_len12 %>% select(- Data, - Insecticide)) %>%
  kbl(booktabs = TRUE, caption = "Differences in the sizes of the 95 credibility intervals (CI) of"
  kable_styling()
```

287 3.2 Results obtained by simulation

288 3.2.1 Interest of combining trials with prevalence and trials with intensity

289 In this section, we consider the situation where only one type of observation is available per trial
 290 - pest prevalence or pest intensity. We compare the accuracy of the estimated parameters and
 291 estimated levels of treatment efficacy obtained by combining both types of trials compared to the
 292 results obtained using each set of trials separately.

293 The parameters used to generate the data are given in table 6.

294 Figure 5 represents the E_γ (Equation 6) (A), $E_{E_{f_6}}$ (Equation 7) (B) and $E_{E_{f_{12}}}$ (Equation 7) (C) evaluation
 295 criteria for the “0% Y - 50% Z”, “50% Y - 0% Z” and “50% Y - 50% Z” scenarios (different scenarios are
 296 indicated by different colors. The x-axis presents the number of trials and the y-axis the value of the

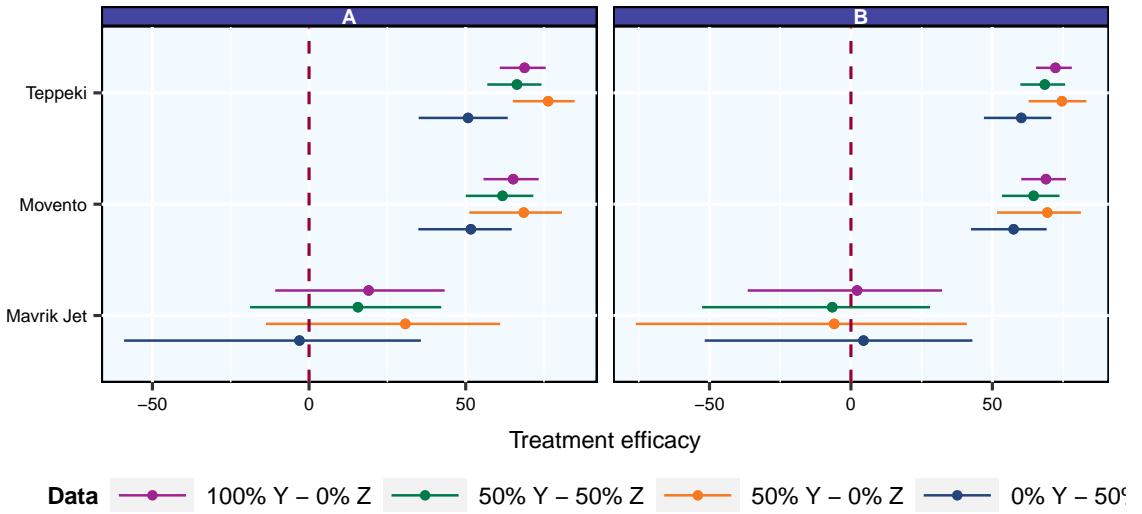


Figure 4: Estimated treatment efficacies after 6 days (**A**) and after 12 days (**B**), with their credibility intervals. Colors correspond to the different scenarios.

Table 5: Differences in the sizes of the 95 credibility intervals (CI) of the estimated treatment efficacies for the scenarios 50% Y - 50% Z, 0% Y - 50% Z and 100% Y - 0% Z, compared to 50% Y - 0% Z. The difference is given in percentage. A positive (negative) value indicates an increase (decrease) of the credibility interval size. The third column indicates differences for the efficacy at 6 days after pesticide spray, and the fourth column indicates the difference for the efficacy at 12 days.

Insecticide	Data	For efficacy at 6 days	For efficacy at 12 days
Mavrik Jet	0% Y - 50% Z	26.7	-19.1
Movento	0% Y - 50% Z	0.8	-9.8
Teppeki	0% Y - 50% Z	44.4	16.9
Mavrik Jet	100% Y - 0% Z	-27.7	-41.3
Movento	100% Y - 0% Z	-40.5	-46.7
Teppeki	100% Y - 0% Z	-25.6	-37.8
Mavrik Jet	50% Y - 50% Z	-18.4	-31.1
Movento	50% Y - 50% Z	-27.1	-31.4
Teppeki	50% Y - 50% Z	-12.1	-22.4

Table 6: Parameters used to generate virtual data.

Parameters	α_0	α_1	γ_{00}	γ_{01}	γ_{02}	γ_{03}	γ_{10}	γ_{11}	γ_{12}	γ_{13}	σ_0	η	χ
Values	0.50	0.16	0.00	-0.13	-1.13	-1.24	0.00	0.24	-0.14	-0.15	1.87	0.98	0.27

297 criterion, averaged over the simulated data sets. For each number of trials and for each criterion, we
 298 observe that scenario “50% Y - 50% Z” gives a more accurate estimate than scenario “50% Y - 0% Z”
 299 which itself gives a more accurate estimate than scenario “0% Y - 50% Z”. For example, for the efficacy
 300 at 6 days with 40 trials, the mean absolute error of scenario “50% Y - 0% Z” is 10% less than the mean
 301 absolute error of scenario “0% Y - 50% Z” (0.38 vs 0.42). The mean absolute error of scenario “50% Y -
 302 50% Z” is 35% less than that of scenario “0% Y - 50% Z” (0.26 vs 0.42). The values of the three criteria
 303 decrease with the number of trials. The E_γ criterion decreases from 0.62 with 20 trials to 0.32 with 80
 304 trials for the “50% Y - 50% Z” scenario. A value of 20 trials is therefore not sufficient to obtain an
 305 accurate estimate of the parameters.

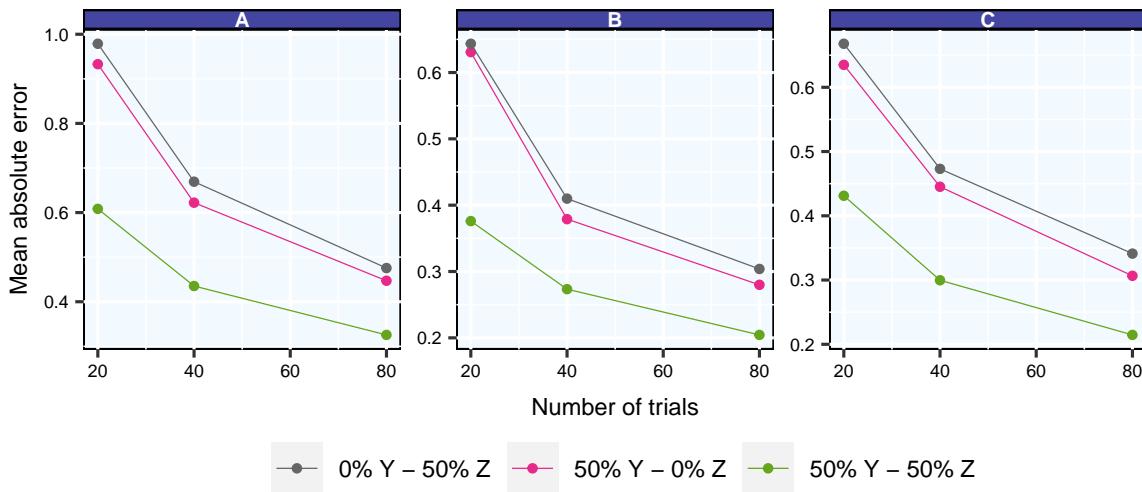


Figure 5: Values of the E_γ (Equation 6) (A), E_{Ef_0} (Equation 7) (B) and $E_{Ef_{12}}$ (Equation 7) (C) mean absolute error criteria for the ‘0% Y - 50% Z’, ‘50% Y - 0% Z’ and ‘50% Y - 50% Z’ scenarios. The x-axis presents the number of trials and the y-axis presents the mean absolute error, averaged over the 974 simulated datasets. Different colors correspond to different scenarios.

306 Figure 6 presents the percentages of cases where the best treatment at 6 days (A) and 12 days (B)
 307 has been correctly identified for the “50% Y - 0% Z”, “0% Y - 50% Z” and “50% Y - 50% Z” scenarios.
 308 The x-axis presents the number of trials and the y-axis the percentage of cases where the treatment
 309 identification is correct. In general, the best treatment is better identified when the number of trials
 310 increases. With the “50% Y - 50% Z” scenario, the best treatment at 6 days is well identified in 69% of
 311 cases with 20 trials and in 85% of cases with 80 trials. For each number of trials, the percentage of
 312 correctly identification is higher for the “50% Y - 50% Z” scenario than for the other two, and the
 313 scenario “50% Y - 0% Z” generally gives better results than the scenario “0% Y - 50% Z”, except at
 314 6 days with 20 trials. For example, at 12 days after treatment and with 40 trials, the percentage of
 315 correct identification is 5% higher with scenario “50% Y - 50% Z” than with scenario “50% Y - 0% Z”
 316 (78 vs 73), and 4% higher with the scenario “50% Y - 0% Z” than with scenario “0% Y - 50% Z” (73 vs
 317 69). These results show the interest of combining prevalence and intensity data for assessing the
 318 efficacy of treatments and identifying the best treatments.

319 3.2.2 Interest of adding intensity when prevalence is measured in all trials

320 We now consider a situation where prevalence is measured in each trial and intensity is measured in
 321 only some of these trials. We compare the results obtained when the data are combined and when
 322 they are used separately. As the prevalence data are usually more accessible in practice and the
 323 intensity data more costly, it is important to evaluate the interest of adding intensity data in the

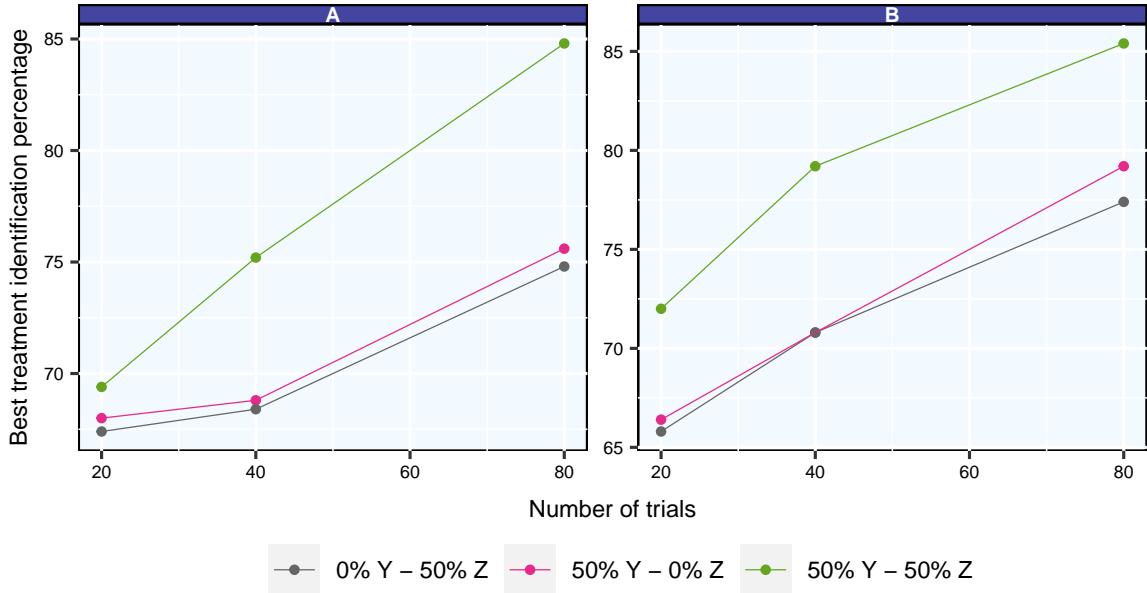


Figure 6: Comparison of proportion cases where the best treatment is correctly identified in the ‘0% Y - 50% Z’, ‘50% Y - 0% Z’ and ‘50% Y - 50% Z’ scenarios. The x-axis represents the number of trials and the y-axis represents the percentage of cases where the best treatment has been correctly identified at 6 days (A) and 12 days (B), over the 974 simulated datasets. Different colors correspond to different scenarios.

324 statistical analysis.

325 The parameters used to generate the data are the same as in 3.2.1.

326 Figure 7 presents the evaluation criteria E_γ (Fig 7A), E_{Ef_6} (Fig 7B) and $E_{Ef_{12}}$ (Fig. 7C) for the scenarios
 327 “0% Y - 100% Z”, “50% Y - 0% Z”, “50% Y - 100% Z” and “50% Y - 50% Z”. The x-axis presents the number
 328 of trials and the y-axis presents the value of the criterion, averaged over the number of simulated
 329 data sets. Results show that the mean absolute errors are lower in scenarios “50% Y - 100% Z” and
 330 “50% Y - 50% Z” than in “0% Y - 100% Z”, and that the mean absolute errors are lower in the “0%
 331 Y - 100% Z” scenario than in the “50% Y - 0% Z” scenario. For example, considering the treatment
 332 efficacy at 12 days with 40 trials (Fig. 7C), the mean absolute errors are 13% lower in scenarios “50%
 333 Y - 100% Z” and “50% Y - 50% Z” than in “0% Y - 100% Z” (0.30 vs 0.34), and the mean absolute error is
 334 11% lower in “0% Y - 100% Z” than in “50% Y - 0% Z” (0.34 vs 0.38). Clearly, adding intensity data to
 335 prevalence data improves the accuracy of the estimations. The mean absolute errors decrease with
 336 the number of trials. For example, the E_γ criterion decreases from 0.62 with 20 trials to 0.32 with 80
 337 trials for the “50% Y - 50% Z” scenario. As noted above, 20 trials is clearly not sufficient to obtain
 338 accurate results.

339 3.2.3 Is it better to measure intensity or prevalence in new pest surveys?

340 In order to optimize the design of new pest surveys that might be conducted in the future, we
 341 determine which type of observations should be favored. For that purpose, we compare the results
 342 obtained with the “100% Y - 0% Z”, “0% Y - 100% Z”, “100% Y / 100% Z” and “100% W” scenarios, for
 343 different values of α_0 that defines the average number of infested plants. With $\alpha_0 = -1$, the proportion
 344 of infested plants is generally much lower than one, while with $\alpha_0 = 2$, 100% of plants are generally
 345 infested. The case $\alpha_0 = 1$ leads to intermediate levels of infestation.

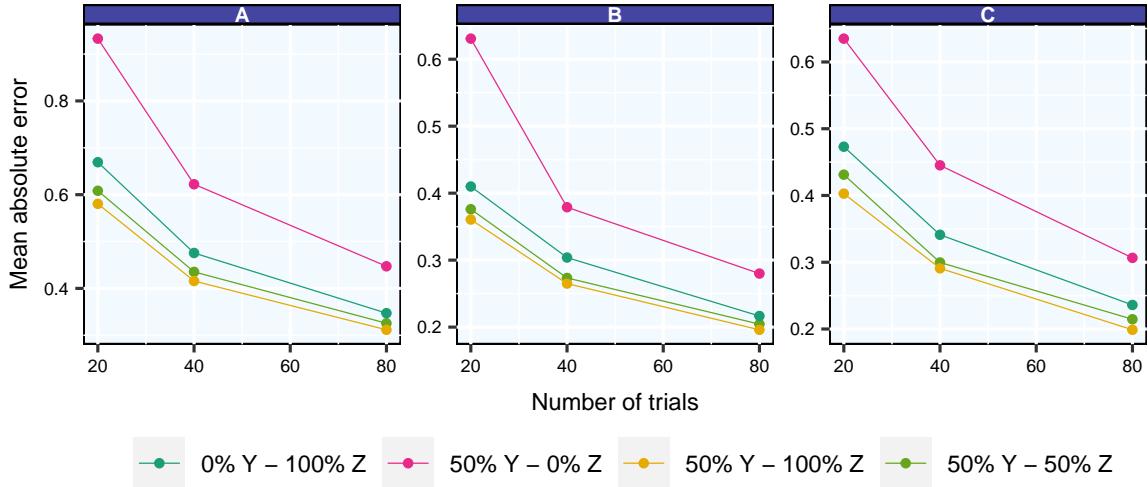


Figure 7: Values of E_γ (Equation 6) (A), $E_{E_{f_6}}$ (Equation 7) (B) and $E_{E_{f_{12}}}$ (Equation 7) (C) for the ‘0% Y - 100% Z’, ‘50% Y - 0% Z’, ‘50% Y - 100% Z’ and ‘50% Y - 50% Z’ scenarios. The x-axis presents the number of trials and the y-axis presents the absolute error averaged over the 974 simulated datasets. Different colors correspond to different scenarios.

³⁴⁶ The three parameter sets used to generate the data are given in Table 7 and are labeled A, B and C.

```
#|echo: false

tab7 = data.frame(Parameters = c(
  # alpha_0 et alpha_1
  c("\u03b1\u2080", "\u03b1\u2081"),
  # gamma_0
  c("\u03b3\u2080\u2080", "\u03b3\u2080\u2080\u2081", "\u03b3\u2080\u2080\u2080\u2082", "\u03b3\u2080\u2080\u2080\u2081"),
  # gamma_1
  c("\u03b3\u2081\u2080", "\u03b3\u2081\u2080\u2081", "\u03b3\u2081\u2080\u2081\u2082", "\u03b3\u2081\u2080\u2081\u2080"),
  # sigma_0, eta et chi
  c("\u03c3\u2080", "\u03b7", "\u03c7")
),
"Values set A" = c(-1, alpha1, gamma0, gamma1, sig0, eta, chi),
"Values set B" = c(1, alpha1, gamma0, gamma1, sig0, eta, chi),
"Values set C" = c(2, alpha1, gamma0, gamma1, sig0, eta, chi)
)

tab7 %>%
  t %>%
  `rownames<-`(`Set", "A", "B", "C")) %>%
  kbl(booktabs = TRUE, caption = "Parameters considered for the design of future pest surveys.")
  column_spec(2, bold = TRUE) %>%
  kable_styling()
```

Table 7: Parameters considered for the design of future pest surveys.

Set	α_0	α_1	γ_{00}	γ_{01}	γ_{02}	γ_{03}	γ_{10}	γ_{11}	γ_{12}	γ_{13}	σ_0	η	χ
A	-1.00	0.16	0.00	-0.13	-1.13	-1.24	0.00	0.24	-0.14	-0.15	1.87	0.98	0.27
B	1.00	0.16	0.00	-0.13	-1.13	-1.24	0.00	0.24	-0.14	-0.15	1.87	0.98	0.27
C	2.00	0.16	0.00	-0.13	-1.13	-1.24	0.00	0.24	-0.14	-0.15	1.87	0.98	0.27

Figure 8 (A.1, B.1 and C.1) shows the mean absolute error E_y (Equation 6) as a function of the number of trials for the four scenarios “100% Y - 0% Z”, “0% Y - 100% Z”, “100% Y / 100% Z” and “100% W”. Figure 8 (A.2, B.2 and C.2) shows the distributions of infested plants with 40 trials corresponding to the three values of α_0 reported in Table 7. In case A (Table 7), the distribution of Z is such that Z is rarely close to 1 and often lower than 0.5 (Figure 8 A.2). In case C, the distribution of Z is such that Z is often very close to 1 (100% of plants infested). Case B is intermediate. The accuracy of the estimated values of the model parameters γ is better with scenario “100% Y - 0% Z” than with scenario “0% Y - 100% Z”, for all number of trials. The advantage of “100% Y - 0% Z” is stronger in case of high pest prevalence (i.e., cases B and C) but very small in case of low pest prevalence (case A). For example, with 20 trials, the mean absolute error is 27% lower in the scenario “100% Y - 0% Z” than in “0% Y - 100% Z” for parameter set C (0.55 vs. 0.75), 10% lower for parameter set B (0.55 vs. 0.62), and not different for parameter set A (0.64). The “100% W” scenario leads to similar results as “100% Y - 0% Z”, regardless of α_0 and the number of trials. Results obtained with “100% Y / 100% Z” are generally similar to those obtained with “100% Y - 0% Z” and “100% W” but better than those obtained with the scenario “0% Y - 100% Z” in cases B and C. Here again, results show that 20 trials are not sufficient to obtain accurate parameter estimates.

4 Conclusion

In order to evaluate pest treatment efficacy, numerous trials are conducted to monitor pest prevalence and intensity. Quite often, only one type of data is available and, when both prevalence and intensity are available, they are usually analysed separately. In this paper, we proposed an alternative approach based on a hierarchical statistical model able to analyze intensity and prevalence data, simultaneously.

We successfully applied the model to a real dataset including prevalence and incidence data collected to evaluate three pesticide treatments against aphids in sugar beets. The model was fitted to this dataset using a Markov chain Monte Carlo algorithm, and convergence was quickly achieved after a few thousands iterations. Results showed that the use of both prevalence and intensity data led to a substantial reduction of the uncertainty in the parameter estimates, compared to the use of a single type of data.

Results obtained from simulated data confirmed that, when pest prevalence and pest intensity are collected separately in different trials, the model parameters are more accurately estimated combining both prevalence and intensity trials than using one type of trials only. We also found that, when prevalence data are collected in all trials and intensity data are collected in a subset of trials, estimations and pest treatment ranking are more accurate using both types of data than using prevalence data only. Moreover, when only one type of observation can be collected in a pest survey or in an experimental trial, our analysis indicates that it is usually better to collect intensity data than prevalence data, especially in situations where all or most of the plants are expected to be infested. Finally, our simulations show that it is unlikely to obtain accurate results with fewer than 40 trials when assessing the efficacy of pest control treatments based on prevalence and intensity data.

Although our framework was illustrated to compare the efficacy of plant pest treatments, it could be

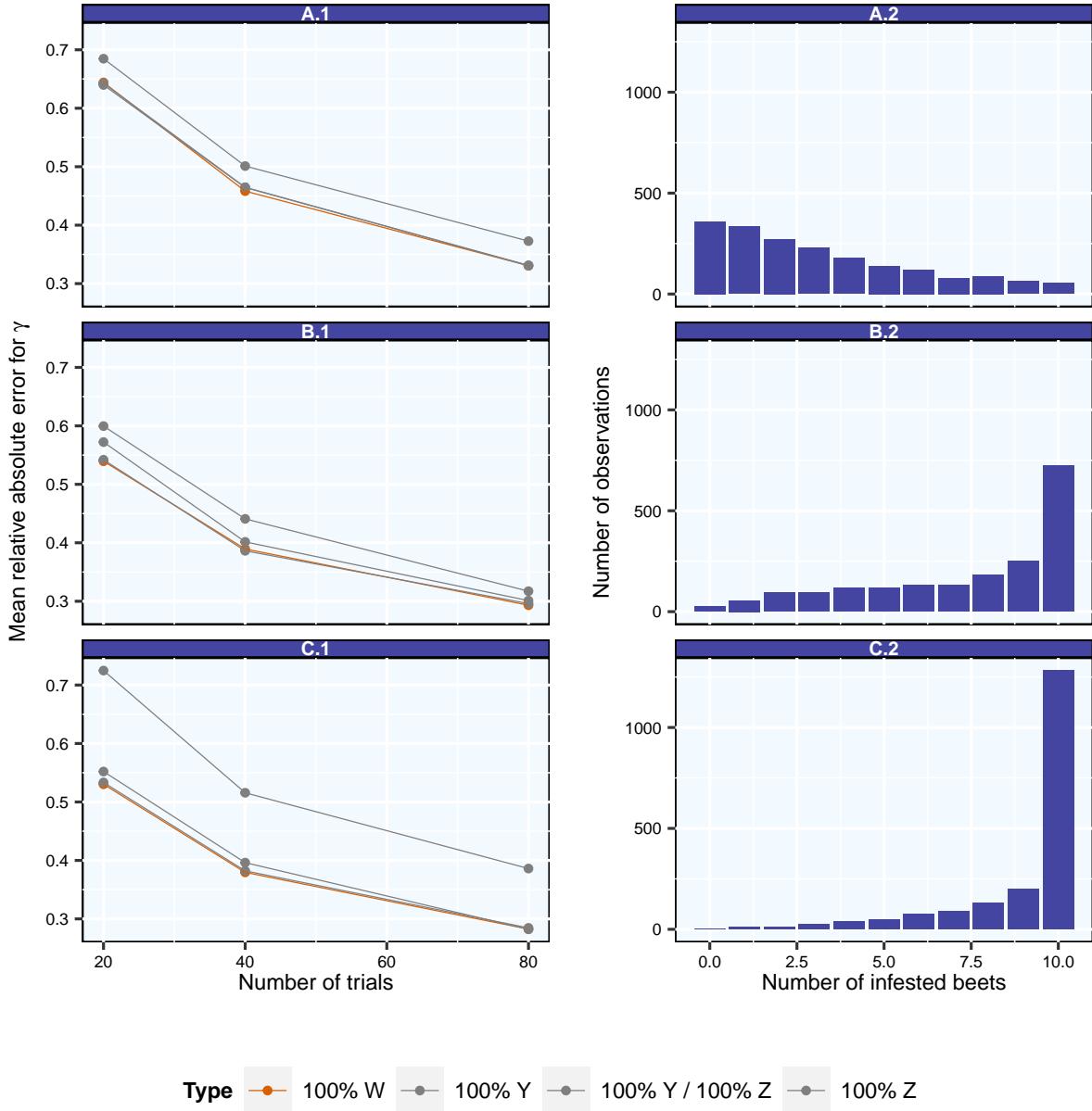


Figure 8: Comparison of the ‘100% Y - 0% Z’, ‘0% Y - 100% Z’, ‘100% Y / 100% Z’ and ‘100% W’ scenarios according to the distribution of Z and the number of trials, using the E_γ criterion (Equation 6). A, B and C correspond to different Z distributions which are given by A.2, B.2 and C.2 (distribution for a number of trials equal to 40). A, B and C respectively correspond to $\alpha_0 = -1, 1$ and 2 . The details of the simulation parameters are given in table 7. A1, B1 and C1 represent the absolute error E_γ averaged over the 974 simulated datasets as a function of the number of trials. Colors correspond to the different scenarios.

385 applied to other areas of research in the future, in particular for optimizing designs used in animal
386 and human epidemiology. It is imperative to note that the ultimate selection of a design should
387 be contingent upon the consideration of localized constraints. As the model codes are made fully
388 available, we believe that these codes could be used by different institutes to compare many different
389 designs in the future, not only the types of designs considered in our paper. Of particular significance
390 is the capability of our model to optimize sample sizes, with its impact contingent on the relative
391 importance of within-trial variability compared to between-trial variability.

392 **Author contributions**

393 AF and DM designed the study. AF performed the computations. AF and DM wrote the paper.

394 **Funding**

395 This work was partly funded by the project SEPIM (PNRI) and by the RMT SDMAA.

396 **Data availability**

397 Simulated data and model parameters are available without restriction. The original experimental
398 data may be available under request.

399 **Acknowledgements**

400 We are grateful to Anabelle Laurent, Elma Raaijmakers, Kathleen Antoons and to the institute ITB
401 (<https://www.itbfr.org/>) for their comments on this project.

402 We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale
403 bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing help and/or computing
404 and/or storage resources.

405 **References**

406 **Session information**

```
sessionInfo()

407 R version 4.3.1 (2023-06-16)
408 Platform: x86_64-pc-linux-gnu (64-bit)
409 Running under: Ubuntu 22.04.3 LTS

410
411 Matrix products: default
412 BLAS:    /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
413 LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p0.3.20.so;  LAPACK version 3.10.0

414
415 locale:
416 [1] LC_CTYPE=C.UTF-8          LC_NUMERIC=C           LC_TIME=C.UTF-8
417 [4] LC_COLLATE=C.UTF-8        LC_MONETARY=C.UTF-8   LC_MESSAGES=C.UTF-8
418 [7] LC_PAPER=C.UTF-8         LC_NAME=C             LC_ADDRESS=C
419 [10] LC_TELEPHONE=C         LC_MEASUREMENT=C.UTF-8 LC_IDENTIFICATION=C
```

```

420
421 time zone: UTC
422 tzcode source: system (glibc)
423
424 attached base packages:
425 [1] grid      stats     graphics grDevices datasets  utils      methods
426 [8] base
427
428 other attached packages:
429 [1] kableExtra_1.3.4 latex2exp_0.9.4 rjags_4-13      coda_0.19-4
430 [5] gridExtra_2.3  forcats_0.5.1   stringr_1.4.0    dplyr_1.1.2
431 [9] purrr_1.0.1   readr_2.1.2    tidyverse_1.3.1 ggplot2_3.3.6
432 [13] tidyverse_1.3.1
433
434 loaded via a namespace (and not attached):
435 [1] gtable_0.3.0    xfun_0.30       lattice_0.20-45   tzdb_0.3.0
436 [5] vctrs_0.6.3     tools_4.3.1      generics_0.1.2     fansi_1.0.3
437 [9] pkgconfig_2.0.3 dbplyr_2.1.1    webshot_0.5.3     readxl_1.4.0
438 [13] assertthat_0.2.1 lifecycle_1.0.3   farver_2.1.0      compiler_4.3.1
439 [17] munsell_0.5.0   htmltools_0.5.2   yaml_2.3.5       pillar_1.9.0
440 [21] crayon_1.5.1    ellipsis_0.3.2   tidyselect_1.2.0   rvest_1.0.2
441 [25] digest_0.6.29   stringi_1.7.6    labeling_0.4.2    cowplot_1.1.1
442 [29] fastmap_1.1.0   colorspace_2.1-0   cli_3.6.0       magrittr_2.0.3
443 [33] utf8_1.2.2     broom_0.8.0      withr_2.5.0      scales_1.2.0
444 [37] backports_1.4.1 lubridate_1.8.0    rmarkdown_2.14    modelr_0.1.8
445 [41] httr_1.4.3      cellranger_1.1.0  hms_1.1.1       evaluate_0.15
446 [45] knitr_1.39     haven_2.5.0      viridisLite_0.4.0 rlang_1.1.1
447 [49] glue_1.6.2      DBI_1.1.2       xml2_1.3.3      renv_1.0.2
448 [53] reprex_2.0.1    svglite_2.1.0    rstudioapi_0.13 jsonlite_1.8.0
449 [57] R6_2.5.1       systemfonts_1.0.4 fs_1.5.2

```

- 450 Agresti, Alan. 2015. *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons.
- 451 Gelman, Andrew, John B Carlin, Hal S Stern, and Donald B Rubin. 1995. *Bayesian Data Analysis*.
Chapman; Hall/CRC.
- 453 Gelman, Andrew, and Donald B Rubin. 1992. “Inference from Iterative Simulation Using Multiple
454 Sequences.” *Statistical Science*, 457–72.
- 455 Gurmu, Shiferaw, and John Elder. 2000. “Generalized Bivariate Count Data Regression Models.”
456 *Economics Letters* 68 (1): 31–36.
- 457 Harrison, Xavier A. 2014. “Using Observation-Level Random Effects to Model Overdispersion in
458 Count Data in Ecology and Evolution.” *PeerJ* 2: e616.
- 459 Laurent, Anabelle, Armand Favrot, Fabienne Maupas, Cédric Royer, and David Makowski. 2023.
460 “Assessment of Non-Neonicotinoid Treatments Against Aphids on Sugar Beets.” *Crop Protection*
461 164: 106140. <https://doi.org/10.1016/j.cropro.2022.106140>.
- 462 Madden, LV, and G Hughes. 1999. “Sampling for Plant Disease Incidence.” *Phytopathology* 89 (11):
463 1088–1103.
- 464 Michel, Lucie, Francois Brun, and David Makowski. 2017. “A Framework Based on Generalised
465 Linear Mixed Models for Analysing Pest and Disease Surveys.” *Crop Protection* 94: 1–12.
- 466 Osiewalski, Jacek, and Jerzy Marzec. 2019. “Joint Modelling of Two Count Variables When One of
467 Them Can Be Degenerate.” *Computational Statistics* 34: 153–71.
- 468 Park, Eun Sug, and Dominique Lord. 2007. “Multivariate Poisson-Lognormal Models for Jointly

- 469 Modeling Crash Frequency by Severity." *Transportation Research Record* 2019 (1): 1–6.
- 470 Pisa, Lennard W, Vanessa Amaral-Rogers, Luc P Belzunces, Jean-Marc Bonmatin, Craig A Downs,
471 Dave Goulson, David P Kreutzweiser, et al. 2015. "Effects of Neonicotinoids and Fipronil on
472 Non-Target Invertebrates." *Environmental Science and Pollution Research* 22 (1): 68–102.
- 473 Plummer, Martyn. 2022. *Rjags: Bayesian Graphical Models Using MCMC*. <https://CRAN.R-project.org/package=rjags>.
- 475 Shaw, Allison K, Julie Sherman, F Keith Barker, and Marlene Zuk. 2018. "Metrics Matter: The Effect
476 of Parasite Richness, Intensity and Prevalence on the Evolution of Host Migration." *Proceedings
477 of the Royal Society B* 285 (1891): 20182147.
- 478 Wood, Thomas James, and Dave Goulson. 2017. "The Environmental Risks of Neonicotinoid Pest-
479 cides: A Review of the Evidence Post 2013." *Environmental Science and Pollution Research* 24 (21):
480 17285–325.