Bayesian correlated models for assessing the prevalence of viruses in organic and non-organic agroecosystems

Elena Lázaro[1] , Carmen Armero[1], Luis Rubio[2]

Abstract

Virus diseases constitute one of the most important limiting factors in horticultural production. Cultivation of horticultural species under organic management has increased in importance in recent years. However, the sustainability of this new production method needs to be supported by scientific research, especially in the field of virology. We studied the prevalence of three important virus diseases in agroecosystems with regard to its management system: organic *versus* non-organic, with and without greenhouse. Prevalence was assessed by means of a Bayesian correlated binary model which connects the risk of infection of each virus within the same plot and was defined in terms of a logit generalized linear mixed model (GLMM). Model robustness was checked through a sensitivity analysis based on different hyperprior scenarios. Inferential results were examined in terms of changes in the marginal posterior distributions, both for fixed and for random effects, through the Hellinger distance and a derived measure of sensitivity. Statistical results suggested that organic systems show lower or similar prevalence than non-organic ones in both single and multiple infections as well as the relevance of the prior specification of the random effects in the inferential process.

[1]Department of Statistics and Operations Research, Faculty of Mathematics, Universitat de València, Spain. elena.lazaro@uv.es, carmen.armero@uv.es

[2]Valencian Institute for Agricultural Research, Spain. lrubio@ivia.es

MSC:

Keywords: Hellinger distance, model robustness, risk infection, sensitivity analysis, virus epidemiology.

1. Ordinary Text

Society is becoming increasingly concerned about environmental damage caused by agricultural activities. The sustainability of conventional agriculture is now being questioned, which is prompting traditional production systems to evolve toward production methods that can protect both environmental and human health (Van Bruggen, 1995; Bengtsson et al., 2005).

In recent decades, organic agriculture has grown rapidly in comparison with other agricultural systems. The adoption of these new agricultural practices has brought about the need to compare low-input and conventional systems to verify whether agroecosystem sustainability can be achieved (Bettiol et al., 2004). Despite the emergence of organic agriculture systems, the literature on their effects and interactions is scarce and insufficient, above all in the field of virology (Tomlinson, 1987). Diseases caused by viruses constitute a major threat to the large-scale production of crops worldwide, causing serious economic losses and undermining sustainability (Gallitelli, 2000). Assessing the risk of infection should therefore be a priority in the study of the epidemiology of such virus diseases.

The ecological and epidemiological factors that determine virus infections in vegetable crops are diverse and little is known about them. The sources and spread of viruses, together with certain agricultural and horticultural practices, have a strong influence on their prevalence (Hanssen et al., 2010). In this respect, studies on the risk of virus infections need to characterize the agroecosystem balance as well as understand the complex relationships between organisms (plants, pathogens, and vectors) and envi-

ronment (Serra et al., 1999).

The main scientific question addressed in this paper is the study and comparison of the risk of different virus infections in tomato and pepper plots characterized by their agroecosystem. Specifically, we focus on the detection and quantification of the effects associated with organic management. The agroecosystem of each plot is defined through a set of covariates containing information on its management conditions and altitude. Agroecosystems are dynamic entities (Finley et al., 2011) with complex sources of uncertainty and hierarchies. Following Thornley and France (2007), the estimation of the infection risk of different viruses within the same plot would require the modelling of not only a suitable set of covariates but also the inclusion of some probabilistic terms which connect the different observations of the same individidual.

The inclusion of dependence and/or correlation relationships among variables, response and/or covariates, is usually done by means of random effects whose stochastic nature adds much more probability to the structure of the model. Bayesian reasoning provides a natural environment for analysing them mainly because of the own conception of the Bayesian probability theory, which specifies all the uncertainties in the model through probabilistic elements (Loredo, 1990). Some applied papers that illustrate the benefits of hierarchical Bayesian models in biometrics scenarios are Alvares et al. (2016) in agriculture, Paradinas et al. (2015) in fisheries, Paciorek et al. (2009) in forestry, and Clark et al. (2007) in ecology.

A Bayesian binary correlated model under the generalized linear mixed models (GLMM) specification was considered to perform a regression analysis of the prevalence of the different viruses. Random effects were used to correlate the risk of infection of each virus in the same plot and quantify the intra-plot ability to be infected. Robustness in hierarchical Bayesian models is a major concern as it can be affected by an inappropriate choice of the hyperprior distributions for hyperparameters (Lambert et al., 2005;

Gelman, 2006; Roos and Held, 2011; Roos et al., 2015). To this effect, the sensitivity of the modelling was tested using several specifications for the hyperprior distribution of the random effects scale parameter. A general measure based on the Hellinger distance (Le Cam, 2012), with its calibration, was used to quantify discrepancies in the subsequent posterior marginal distribution of the common regression coefficients and hyperparameter.

The remainder of this article is organized as follows: Section 2 reviews the data and presents the formulation of the model. Section 3 reports and discusses the results with regard to multiple and single viral infections. Section 4 proposes several random effects specifications and analyses the robustness of the estimated models through a sensitivity measure based on the Hellinger distance. Some concluding remarks are given in Section 5.

2. Viruses data and statistical modelling

2.1. Data description

Globally, about 30 viruses are capable of affecting the most known horticultural crops. However, despite being able to infect a wide variety of species, they usually affect Solanaceae species, specially tomato (*Solanum lycopersicum*) and pepper (*Capsicum annuum L.*). These species are two of the most common vegetable crops grown in Spain whose production is being seriously limited by virus diseases. There has recently been a considerable increase in the cultivation of these vegetables under integrated systems such as organic agriculture. It is therefore essential to carry out subsequent virus prevalence studies in order to guarantee their sustainability.

A project under the auspices of the Valencian Institute Agricultural Research was conducted in the summer of 2012 in the Valencian region for this purpose. A total of 30 plots in tomato and pepper production were selected according to their system of production. Each plot was evaluated in terms of its agroecosystem characterization and

the presence or absence of three different viral infections in the crops: tomato mosaic virus (ToMV), cucumber mosaic virus (CMV) and tomato spotted wilt virus (TSWV). These viruses affect both tomato and pepper crops equally, are transmitted in different ways, and can cause substantial economic losses. The presence of each specific virus infection in a plot was assumed when the virus was detected in at least one of eight randomly-selected plants. The enzyme-linked immunosorbent assay (ELISA) technique (Clark et al., 1976) was used to detect each virus.

The assessment of the agroecosystem of each plot was determined by its management condition and altitude. Management condition was evaluated by classifying each plot as organic, non-organic with greenhouse structure, and non-organic with no greenhouse structure. These categories were defined according to the most representative agroecosystems in Spanish agriculture. Organic plots differ from the non-organic ones in many respects, but substantial differences are related to the use of agrochemicals and other external inputs with important influence in pest and disease prevalence. In fact, some purported drawbacks related to organic agriculture include an increasing incidence of pest damage and higher risks of pest outbreaks (Letorneau and Goldstein, 2001). All plots classified as organic complied with the current regulation and were certificated as such by the Organic Agriculture Committee of the Autonomous Government of València. The presence of greenhouse in non organic plots was also considered because is a frequent practice in non-organic systems. The use of covering protections suppose a physical barrier which is directly related to virus infection in the sense that denies insects (vector of virus transmission) acces to plants.

Of the total of 30 plots of our study, 18 were classified as organic and 12 as non-organic, 5 of them with greenhouse structure. For organic plots, the proportion of infected plants with ToMV, CMV, and TSWV was 0.222, 0.167, and 0.056, respectively. In the case of non-organic plots with greenhouse these proportions were 0.400, 0.200,

and 0.200, respectively, and 0.143, 0.286, and 0.286 for non-organic plots without greenhouse. The organic plots presented a lower proportion of plants infected by CMV and TSWV viruses, but the prevalence of ToMV was lowest in the non-organic plots with no greenhouse.

2.2. Statistical model

We consider a logit GLMM for correlated binary responses (Ntzoufras, 2009) to model the Bernoulli random variable Y_{ij} which describes the presence or absence of virus j ($j = 1$ corresponds to ToMV, $j = 2$ to CMV, and $j = 3$ to TSWV) in plot i,

$$\begin{aligned} (Y_{ij} \mid \theta_{ij}) &\sim \text{Bernoulli}(\theta_{ij}), \\ \text{logit}(\theta_{ij}) &= \boldsymbol{x}_i^{\top}\boldsymbol{\beta}_j + b_i,\ i = 1,\ldots,30, \end{aligned} \tag{1}$$

where θ_{ij} is the probability that virus j will be detected in plot i and represents risk of infection; $\boldsymbol{x}_i$ is the vector of covariates; $\boldsymbol{\beta}_j$ is the corresponding vector of the regression coefficients; and $(b_i \mid \sigma_b^2) \sim \text{N}(0, \sigma_b)$ is a normal random effect associated with plot i with mean zero and standard deviation σ_b. The three management conditions were coded in a sequence of two dummy variables (organic and non-organic, with and without greenhouse structure) to avoid overparameterization, with organic management as the reference category.

Random effects capture within-plot variability and correlate prevalence among all viruses so that each individual virus infection is determined by its own agroecosystem effect and an individual effect plot which denotes its ability to be infected. They also provided conditional independence among the prevalence of the three viruses as follows

$$P(Y_{ij} = y_j, j = 1,2,3 \mid \boldsymbol{\beta}, b_i, \boldsymbol{x}_i) = \prod_{j=1}^{3} P(Y_{ij} = y_j \mid \boldsymbol{\beta}_j, b_i, \boldsymbol{x}_i), \tag{2}$$

where $y_j \in \{0,1\}$, $j = 1,2,3$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_2)^{\top}$, and the conditional probability that plot

147 i will be infected with virus j can be expressed as

$$P(Y_{ij} = 1 \mid \boldsymbol{\beta}_j, b_i, \boldsymbol{x}_i) = \frac{\exp\{\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}_j + b_i\}}{1 + \exp\{\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}_j + b_i\}}, \tag{3}$$

148 The joint marginal distribution obtained integrating out the random effects in (4),

$$P(Y_{ij} = y_j, j = 1,2,3 \mid \boldsymbol{\beta}, \sigma_b, \boldsymbol{x}_i) = \int P(Y_{ij} = y_j, j = 1,2,3 \mid \boldsymbol{\beta}, b_i, \boldsymbol{x}_i)\, \mathrm{N}(b_i \mid 0, \sigma_b)\, \mathrm{d}b_i, \tag{4}$$

149 does not depend on the subject-specific random effects and can be interpreted as the com-
150 mon risk infection of a generic plot from the population with the same agroecosystem
151 and altitude.

Inference was carried out using Bayesian statistics. We therefore needed to elicit a prior distribution for the parameters and hyperparameters to complete the Bayesian model. We considered a prior independent default scenario with normal distributions centered at zero and a wide variance for the regression coefficients. As previously introduced, the specification of a hyperprior distribution for the random effects scale parameter is a challenging issue (Lambert et al., 2005; Gelman, 2006; Roos and Held, 2011; Roos et al., 2015). Section 4 contains a sensitivity analysis of the performance of various traditional hyperprior choices (gamma, uniform and half-normal) in our study. This analysis led us to choose the uniform distribution $\mathrm{Un}(\sigma_b \mid 0, 100)$ for the standard deviation of the random effects. Consequently

$$\begin{aligned} \pi(\boldsymbol{\beta}, \sigma_b) &= \textstyle\prod_{j=1}^{3} \prod_{k=0}^{3} \pi(\beta_{jk})\, \pi(\sigma_b) \\ &= \textstyle\prod_{j=1}^{3} \prod_{k=0}^{3} \mathrm{N}(\beta_{jk} \mid 0, \sigma^2 = 1000)\, \mathrm{Un}(\sigma_b \mid 0, 100) \end{aligned} \tag{5}$$

152 where $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \beta_{j2}, \beta_{j3})^{\mathsf{T}}$ are the regression coefficients associated with organic,
153 non-organic with and without greenhouse and altitude (in logarithmic scale) for virus j.

3. Results

The posterior distribution $\pi(\boldsymbol{\beta}, \sigma_b \mid \mathcal{D})$, where $\mathcal{D}$ denotes data, was approximated using Markov chain Monte Carlo (MCMC) simulation methods with WinBUGS Software (Lunn et al., 2000). Random effects models, and Bayesian categorical GLMs in particular, involve many computational difficulties (Albert and Chib, 1993). We fixed the number of iterations and the burn-in period with very large values to avoid strong correlation in the MCMCs samples and get a reliable sample of the posterior distribution. Specifically, simulation was run considering three Markov chains with 1 000 000 iterations and a burn-in period with 100 000. In addition, the chains were thinned by storing every 10th iteration in order to reduce autocorrelation in the saved sample and avoid computer memory problems.

Trace plots of the simulated values of the chains appear overlapping one another, indicating stabilization. Convergence of the chains to the posterior distribution was assessed using the potential scale reduction factor, $\hat{R}$, and the effective number of independent simulation draws, neff. In all cases, the $\hat{R}$ values were equal or close to 1 and neff > 100, thus indicating that the distribution of the simulated values between and within the three chains was practically identical, and that sufficient MCMC samples had been obtained, respectively (Gelman and Rubin, 1992).

3.1. Management conditions

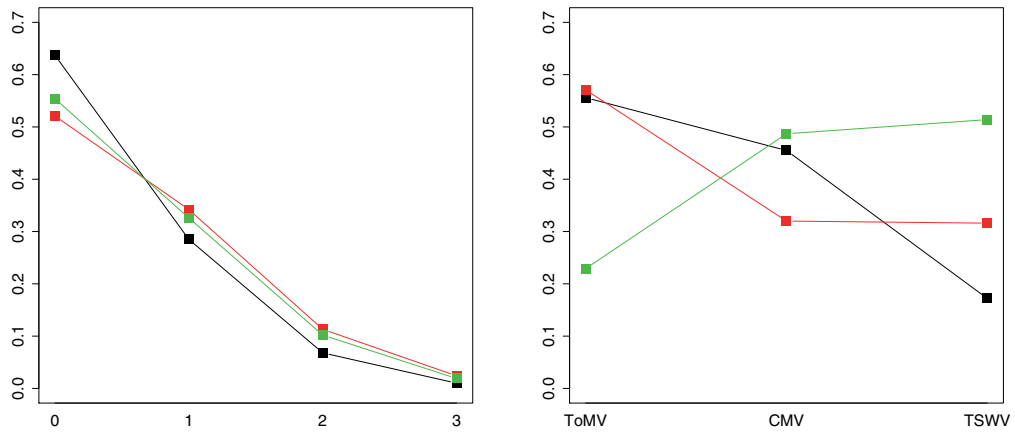
Multiple viral infections that may result in synergisms or antagonisms are frequently found in nature, with unpredictable pathological consequences. Synergistic interactions resulting from mixed infections with two or more viruses are common and well documented in plants (García-Cano et al., 2006). Viral synergism could affect various growth variables such as plant height, weight, and yield (Murphy and Bowen, 2006), and in extreme cases can lead to plant death.

The joint posterior distribution, $\pi(P(Y_{ij} = y_j, j = 1,2,3 \mid \boldsymbol{\beta}, \sigma_b, \boldsymbol{x}_i) \mid \mathcal{D})$, where $y_j \in \{0,1\}$, of the risk infection given in (4) for a generic plot at given altitude in each of the management systems is the basic tool for assessing such synergisms and antagonisms. This posterior distribution is also the starting point for the computation of relevant conditional or marginal inferences.

We begin by discussing some results about multiple viral infections with regard to plot management condition: the posterior distribution of the prevalence of the total number of viruses in a plot and the posterior distribution of the risk of a third infection in plots already infected with two of the viruses. Figure 1a shows the mean of the posterior distribution associated to the presence of 0, 1, 2 and 3 viruses in a generic plot *i* located at 76 meters of altitude (the sample mean) with regard to its management system. Most of the plots have no infections, but the organic ones present the highest rates for plots without infections. Non-organic plots, with and without greenhouse, behave similarly.

Figure 1b shows the posterior mean of the risk of a third infection in plots already infected with two of the viruses. Outcomes are also obtained for a generic plot *i* situated at 76 meters of altitude (the sample mean) with regard to its management system. For condition ToMV in the presence of CMV and TSWV, organic and non-organic with greenhouse plots behave similarly with probabilities around 0.6. This is not the case for non-organic with no greenhouse plots, with an estimated probability close to 0.2. CMV infection given ToMV and TSWV presents homogeneous results in all management systems, with a higher difference among estimated probabilities of 0.167. The pattern for the probability of a TSWV infection in plots already infected with ToMV and CMV seems to be different among the management conditions: non-organic with no greenhouse systems shows the highest probability (0.514), followed by non-organic with greenhouse plots (0.316), and organic (0.172), respectively. It is difficult to detect a general trend on conditional infections among the different agroecosystems analysed.

205 This is a very interesting subject and surely a new study with more data would be nec-
206 essary in order to better understand them.



207 (a) (b)

Figure 1: *(a) Probability (mean of the posterior distribution) for the presence of 0, 1, 2 and 3 viruses in organic (black), non organic-green (red) and non organic-non green (green) management systems. (b) Probability (mean of the posterior distribution) of the risk of a third infection in plots already infected with two of the viruses in organic (black), non organic-green (red) and non organic-non green (green) management systems.*

208 The marginal effect of the management conditions in each virus was assessed through
209 the marginal posterior distribution $\pi(P(Y_{ij} = 1 \mid \boldsymbol{\beta}, \sigma_b, \boldsymbol{x}_i) \mid \mathcal{D})$. Table 1 shows a descrip-
210 tive of the posterior distribution of the risk of infection for each virus and management
211 conditions for a generic plot situated at a height of 76 meters (the sample median). The
212 lowest risk of infection for a generic plot under organic management is for TSWV virus.
213 The most relevant differences among the management conditions were found for virus
214 ToMV. In contrast, virus CMV seemed the most stable. However, the organic effect was
215 weaker for ToMV risk, approximately about four times the one for TSWV virus. It is
216 important to mention the great uncertainty associated to all marginal posterior distribu-
217 tions in the analysis, mainly due to the combination of the reduced size of the sample and
218 the usual scarce information of binary data. To this effect, a bigger experiment would be
219 necessary for a more informative and objective study that allows to reach more precise

conclusions about the subject.

Table 1: *Summary of the posterior distribution of the risk of infection for each management condition and virus.*

Virus	Management	Mean	Sd	$Q_{2.5\%}$	$Q_{50\%}$	$Q_{97.5\%}$
ToMV	Organic	0.225	0.184	0.008	0.181	0.734
	Non-organic, greenhouse	0.311	0.252	0.006	0.248	0.900
	Non-organic, no greenhouse	0.100	0.147	0.000	0.041	0.553
CMV	Organic	0.169	0.161	0.004	0.124	0.634
	Non-organic, greenhouse	0.155	0.190	0.001	0.080	0.719
	Non-organic, no greenhouse	0.234	0.216	0.004	0.168	0.809
TSWV	Organic	0.057	0.093	0.000	0.026	0.309
	Non-organic, greenhouse	0.174	0.203	0.001	0.095	0.764
	Non organic, no greenhouse	0.253	0.223	0.005	0.189	0.831

Comparison of the three management systems was also quantified with the posterior distribution of the risk difference (*RD*) (Christensen et al., 2011). *RD* is an absolute and intuitive measure of association for quantifying difference between proportions associated to an outcome of interest in two groups. It is defined in $[-1,1]$ so that $RD = 0$ means no difference between groups, $-1 \leq RD < 0$ that risk is greater in group 2, and $0 < RD \leq 1$ the opposite.

Figure 2 shows, for each virus, the posterior mean and 95% credible interval of the *RD* between organic and non-organic, with and without greenhouse, generic plots. Information provided by this graphic reaffirms the results in Table 1. Note that the differences between organic management conditions and the two non-organic conditions are clear in the case of TSWV infection: both posterior distributions are highly concentrated on the negative *RD* values with associated posterior probabilities 0.764 and 0.910 when comparing organic and non-organic with and without greenhouse management, respectively. For CMV infections, the results are less clear, with posterior probabilities of 0.395 and 0.611, respectively. In the case of ToMV infection, there are few differences between

organic and non-organic with greenhouse conditions (posterior probability of a negative difference is 0.620), but a relevant probability, 0.84, that the risk of infection will be greater in organic than in non-organic without greenhouse.

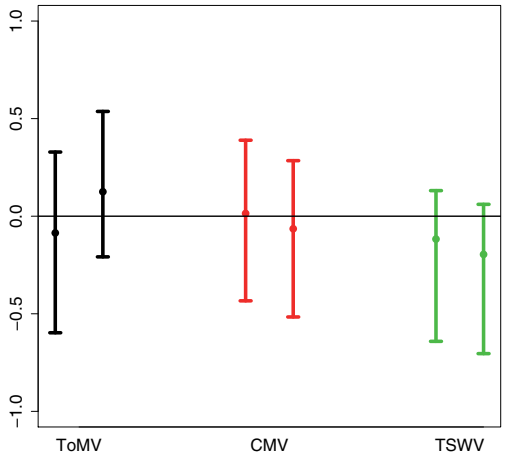
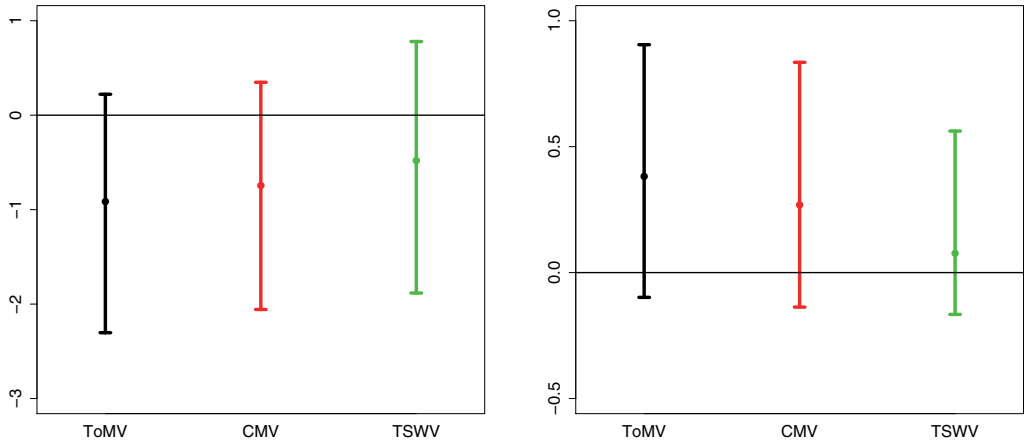


Figure 2: *Posterior mean and 95% credible interval of the RD between organic system in relation to non organic-green (left) and non organic-no green (right) system for ToMV, CMV and TSWV infections.*

3.2. Altitude condition effect

Plot altitude is a relevant epidemiological information due to its important role in shaping insect vector distributions and virus survival. The effect of altitude on the risk of infection is clearly negative in all viruses and therefore we can expect a decrease of the risk of infection as altitude increases. Figure 3a shows the posterior distribution of the regression coefficient associated to altitude for each virus: -0.914, -0.745 and -0.480 are, respectively, the subsequent posterior mean of the coefficient for virus ToMV, CMV, and TSWV, with posterior probabilities 0.940, 0.904, and 0.768 associated to their negative values. Note that virus ToMV is the most negatively associated with altitude. Figure 3b shows the posterior distribution of the *RD* between two generic organic plots with altitudes of 16 and 604 m, the lowest and highest values of the organic plots in the sample. These graphics are in line with the previous comments and also indicate the less

254 important role of altitude in the risk of a TSWV infection in organic crops.



255 (a) (b)

Figure 3: *For virus ToMV (in black), CMV (in red), and TSWV (in green): posterior mean and 95% credible interval of the regression coefficient associated to the altitude (in logarithmic scale) (a), and posterior distribution of the RD between a typical organic plot at altitudes 16 and 604 m (b).*

256 *3.3. Individual random effects*

257 Random effects for each plot capture the ability to be infected of individual plots, thus
258 correlating the risk of infection among the viruses of each plot. Since each individual
259 random effect is responsible for the differences in the estimation of the risk between
260 plots managed under similar agroecosystem conditions, quantifying their contribution to
261 the analysis in terms of factors and covariates is highly relevant to our understanding of
262 the weight of the common and individual elements in the model.

263 The mean of the posterior distribution of the standard deviation, σ_b, of the plot ran-
264 dom effect is 0.968 with a 95% credible interval [0.046, 2.671]. In addition, we assessed
265 the contribution of the random effect associated to each plot towards the conditional
266 posterior distribution of the risk of infection $\pi(P(Y_{ij} = 1 \mid \boldsymbol{\beta}, b_i, \boldsymbol{x}_i) \mid \mathcal{D})$. It was esti-
267 mated individually for the three viruses at the altitude of 76 meters with the purpose of

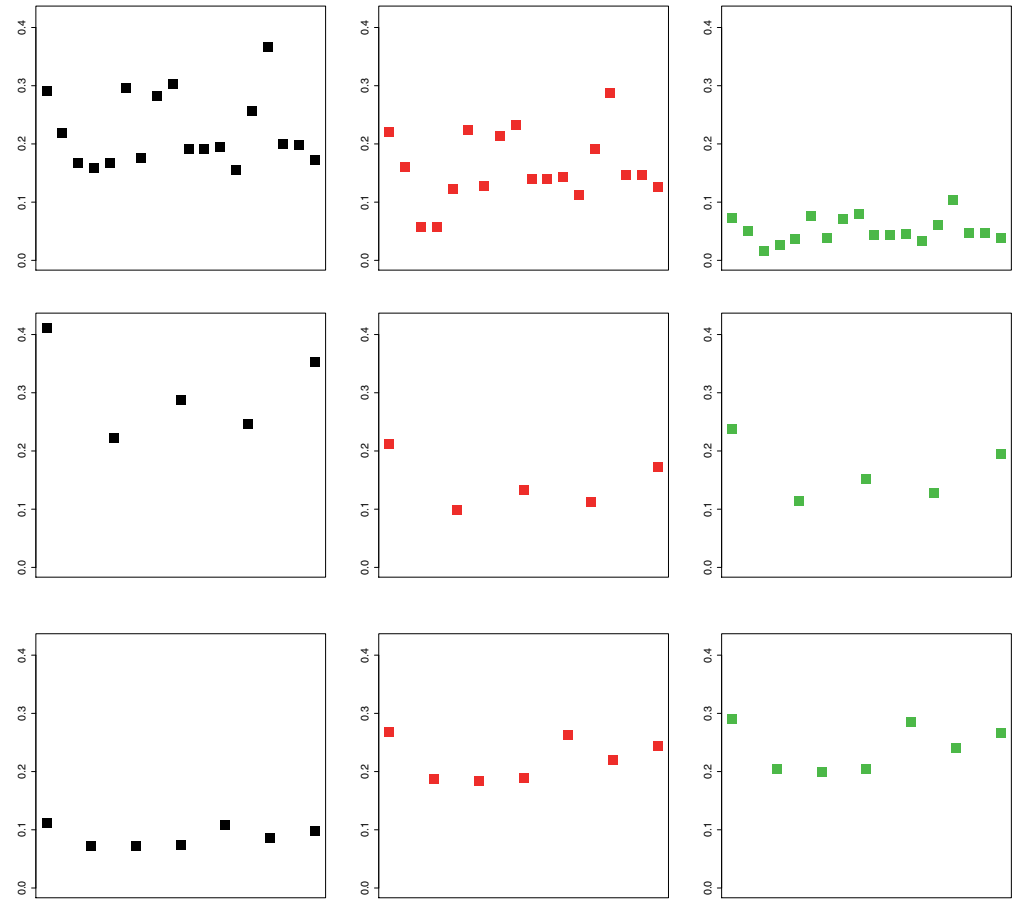
assessing differences in risk infection among individuals that share the specification of the vector of covariates $\boldsymbol{x}_i$, that is to say, plots that were managed under the same system. Figure 4 shows a mosaic of subfigures in which each one displays the posterior expectation of the risk of infection for each plot grouped according to management condition (rows) and the type of virus infection (columns).

We can distinguish a certain stability in risk infection regarding individuals belonging to non-organic no greenhouse systems (row 3) with maximum differences among individuals of 0.039, 0.084 and 0.090 for ToMV, CMV and TSWV respectively. Non-organic with greenhouse plots (row 2) are less similar with maximum differences in risk infection no greater than 0.190 (ToMV). Organic plots showed the most remarkable differences among their individuals, with maximum differences of 0.211 for ToMV and 0.231 for CMV. In contrast TSWV showed the opposite behaviour with a slight maximum difference of 0.087. In conclusion, we suspect the strong relevance of the common elements in the model (fixed effects) in the case of non-organic and no greenhouse plots regardless of virus infection. On the other hand, in the case of organic plots the weight of the common elements effect in the model was not so evident considering that not all viruses exhibited a similar tendency: ToMV and CMV risk infection varied considerably among individuals, but this was not the case with TSWV.

4. Sensitivity analysis

Bayesian GLMMs are a particular class of models for which the estimation process can be seriously affected by the elicitation of prior distributions for the random effects scale parameter (standard deviation, σ_b, or a one-to-one transformation of it, variance σ_b^2 or precision $\tau_b = 1/\sigma_b^2$). Special attention is required in studies where the number of groups is small, σ_b is close to zero, and/or the number of groups is large compared to the number of observations in each group (Box and Tiao, 1992; Gelman, 2006; Roos and Held,

293 2011). This latter situation is the case of our study, with $I = 30$ plots and only three
294 observations in each of them. An additional element that aggravates the situation is the
295 sparsity of the data due to its categorical, binary condition. We conducted a sensitivity
296 analysis of the posterior distribution to the specification of several prior hyperdistribu-
297 tions for the random effects scale parameter. This analysis was based on the methodol-
298 ogy developed in McCulloch (1989), Roos and Held (2011), and Roos et al. (2015) re-
299 garding the stability of the marginal posterior distribution of the regression coefficients
300 of the model and the relative changes in the subsequent marginal posterior distributions
301 of the random effects scale parameter.



(a) ToMV (b) CMV (c) TSWV

Figure 4: *Posterior mean of the conditional posterior distributions associated to management systems organic (row 1), non organic and greenhouse (row 2) and non organic and non greenhouse (row 3) for viruses ToMV (column1), CMV (column 2) and TSWV (column 3) obtained from a fixed altitude value of 76 m.*

4.1. Hyperprior distributions

For the random effects scale parameter, different hyperprior distributions were specified for τ_b within the family of gamma, and for σ_b within uniform and half-normal distributions

- Gamma: $\text{Ga}(0.001, 0.001)$, $\text{Ga}(0.005, 0.005)$, and $\text{Ga}(0.05, 0.05)$ (Ga1, Ga2, and Ga3, respectively),
- Uniform: $\text{Un}(0, 100)$, $\text{Un}(0, 55.63)$, and $\text{Un}(0, 7.92)$ (Un1, Un2, and Un3), and
- Half-normal: $\text{HN}(10)$, $\text{HN}(3.0387)$, and $\text{HN}(0.3965)$ (HN1, HN2, and HN3).

Gamma distributions were parameterized in terms of a shape and a rate parameter, and half-normal distributions were set according its standard deviation. Hyperdistributions Ga1, Un1, and HN1 were considered the default choices due to their "noninformative" nature and their common use in Bayesian applications. In addition, two other hyperparameter specifications within each family of hyperdistributions were contemplated to assess the effect of small and medium perturbations in the hyperparameter specifications on posterior inferences. These hyperprior distributions were set following the criterion of the Hellinger distance (Le Cam, 2012). This is a symmetric and invariant measure of discrepancy between two probability distributions taking values between 0 and 1, where the value 0 represents no divergence and 1, full divergence (See Appendix 1).

Hyperparameter values were assessed considering two reference Hellinger distance values, a small and a medium perturbation. This computation was based on the analytical expression of the Hellinger distance between gamma, uniform and half-normal distributions (see Appendix 1). Small perturbation was associated to a Hellinger distance of 0.541 and medium to 0.848. Consequently, Ga2, Un2, and HN2 hyperparameteres were determined to obtain a Hellinger distance of 0.541 in relation to hyperdistributions Ga1, Un1, and HN1, respectively. Hyperparameter values for Ga3, Un3, and HN3 were selected because of their Hellinger distance, 0.848, to hyperpriors Ga1, Un1, and HN1, respectively.

Focusing on gamma hyperdistributions, Ga1 exhibits the widest range of uncertainty with a variance of 1000. It is frequently used in many of the examples provided with the WinBUGS software (Lunn et al., 2012) and shows a uniform shape for most of the range with a spike of probability density near zero. Ga2 and Ga3 share this shape, although they show lower range coverage as a consequence of their fewer variance values, 200 and 20. Hyperprior Un1 is recommended by Spiegelhalter et al. (2004) in their book on clinical trials. It is a very generous distribution allowing for a great space of values due to its variance of 833.3. Un2 and Un3 display variance values of 257.84 and 5.23, and as such they are very different from the non-null density range. The half-normal default option, HN1, is a choice used in Thompson et al. (1997) and Roos and Held (2011). It exhibits a variance of 36.3 giving a low probability to values greater than this. HN2 and HN3 are more informative versions, especially HN3 with a variance value of 0.06.

We conducted nine independent inferential processes with the same data and the same marginal prior distribution $\pi(\boldsymbol{\beta})$ for the regression coefficients as in (5) but varying marginal hyperprior distribution according to the specifications previously presented.

4.2. Sensitivity of the regression coefficients

We discuss sensitivity of the marginal posterior distributions of the regression coefficients derived from the inferential processes described above. Discrepancies among the estimates of posterior marginal distributions were the result of alterations in the hyperprior values. Hellinger distances between posterior marginal distributions approximated by MCMC methods were computed via expression (A.1) in Appendix 1 and implemented by means of the function `HDistNoSize` in the `R` package `bmk` (Krachey and Boone, 2012). Furthermore, to facilitate interpretation these values were calibrated with regard to a normal distribution with variance 1 (see Appendix 2 for more details about calibration).

Table 2 shows the calibration of the Hellinger distance between the posterior marginal

distribution of the different coefficients of regression computed from the hyperpriors considered. In none of the comparisons the discrepancies observed were greater than the differences between the normal distributions $\mathrm{N}(0, 1)$ and $\mathrm{N}(0.284, 1)$, which reveals that Hellinger values are in general close to zero (see Table 4 in Appendix 2 where a calibration of the normal mean related to its subsequent Hellinger distance is displayed). Uniform distributions have the smallest discrepancies despite the existing differences among hyperpriors Un1, Un2, and Un3. The behaviour of half-normal distributions was similar to that of the uniform distributions in the case of hyperpriors HN1 and HN2. Nevertheless, inference from hyperprior HN3 exhibited the greatest discrepancies, surely due to its informative nature. Gamma showed greater discrepancies than uniform hyperpriors in all cases, although in none of the scenarios did these differences exceed those obtained from hyperprior HN3. Thus, the above comments enable us to conclude that our assumptions on the choice of hyperparameter prior distribution influences the estimates of the regression coefficients only to a minor extent.

We now discuss the effect of the different hyperpriors considered on the posterior distribution of each regression coefficient. Figure (5) is a mosaic of subfigures. Each subfigure displays the posterior mean of the regression coefficients of the different inferential processes conducted. The order of the points corresponds to the order in which hyperpriors are presented (Ga1, Ga2, Ga3; Un1, Un2, Un3; and HN1, HN2, HN3). A great similarity can repeatedly be seen, in practically all coefficients and viruses, between results from hyperpriors HN1 and HN2, and also those from the uniform hyperpriors. As expected, results from HN3 are very different, most likely due to its informative characteristics. Finally, posterior means from the analyses based on the gamma hyperpriors vary the most, indicating a greater sensitivity to parameter specification.

Table 2: *Calibration of the Hellinger distance between the posterior marginal distribution of the coefficients of regression associated to organic (β_o), non-organic with greenhouse ($\beta_{no\text{-}g}$), non-organic without greenhouse ($\beta_{no\text{-}ng}$) and altitude in logarithmic scale (β_{alt}) computed from hyperprior distributions Ga1 and Ga2, Ga1 and Ga3, Un1 and Un2, Un1 and Un3, HN1 and HN2, and HN1 and HN3.*

Virus	Coeff.	(Ga1,Ga2)	(Ga1,Ga3)	(Un1,Un2)	(Un1,Un3)	(HN1,HN2)	(HN1,HN3)
ToMV	β_{o}	0.038	0.084	0.024	0.022	0.034	0.236
	$\beta_{\text{no-g}}$	0.032	0.068	0.019	0.019	0.035	0.197
	$\beta_{\text{no-ng}}$	0.020	0.042	0.018	0.020	0.024	0.124
	β_{alt}	0.043	0.099	0.022	0.024	0.039	**0.284**
CMV	β_{o}	0.033	0.068	0.023	0.021	0.034	0.201
	$\beta_{\text{no-g}}$	0.029	0.056	0.021	0.019	0.025	0.148
	$\beta_{\text{no-ng}}$	0.029	0.060	0.019	0.020	0.027	0.171
	β_{alt}	0.037	0.085	0.023	0.023	0.038	0.249
TSWV	β_{o}	0.022	0.052	0.019	0.021	0.030	0.144
	$\beta_{\text{no-g}}$	0.024	0.043	0.021	0.020	0.025	0.108
	$\beta_{\text{no-ng}}$	0.023	0.048	0.020	0.019	0.025	0.139
	β_{alt}	0.028	0.069	0.020	0.019	0.034	0.193

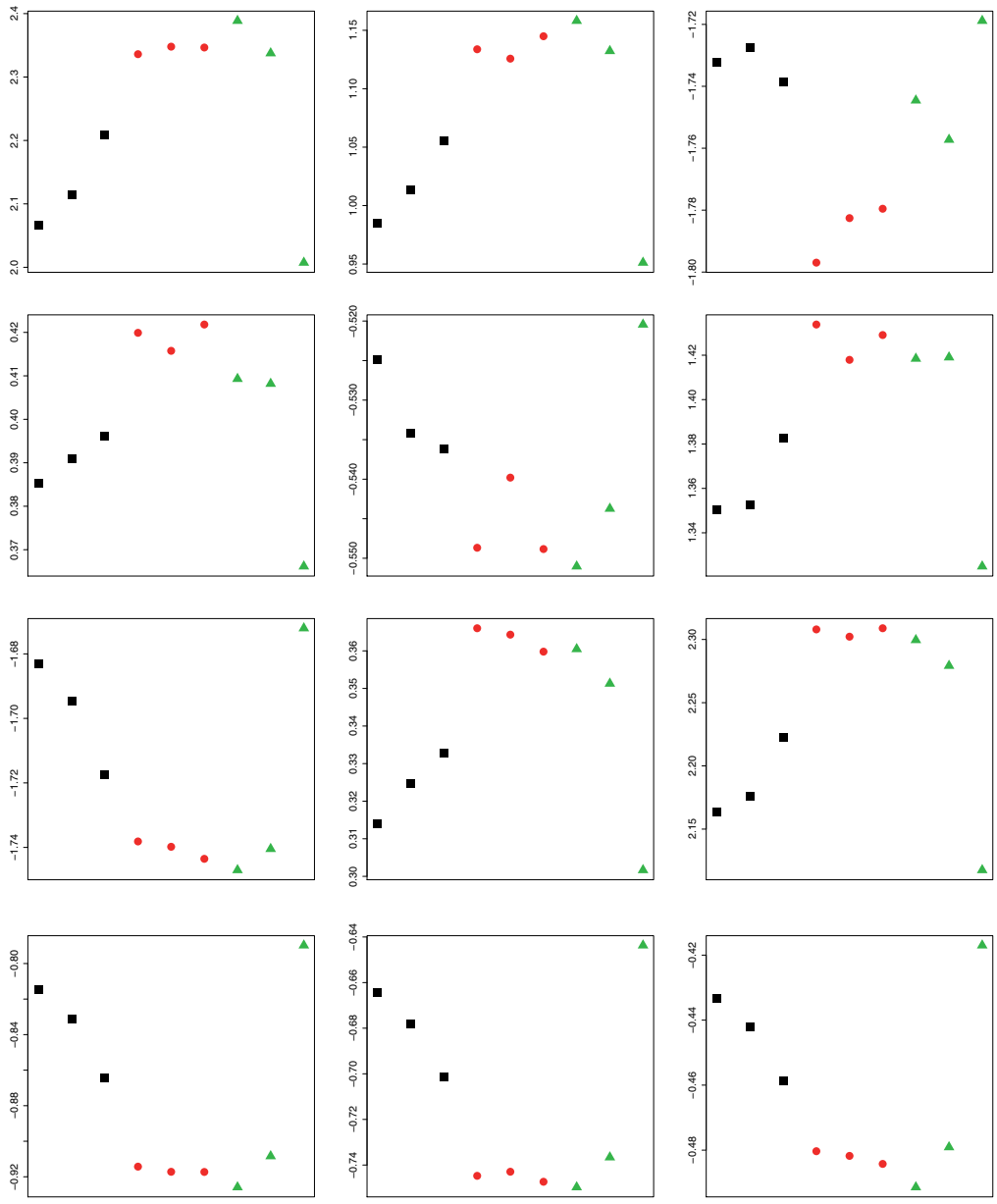
4.3. Sensitivity of the variability of the random effects

We now discuss and assess the sensitivity of the random effects scale parameter corresponding to the inferential processes described in Subsection 4.1. Figure 6 shows the posterior marginal distribution (mean and 95% credible intervals) of the standard deviation of the random effects. It is worth noting that in the case of the gamma hyperpriors, the posterior marginal distribution $\pi(\sigma_b \mid \mathcal{D})$ is computed from the joint posterior $\pi(\boldsymbol{\beta}, \tau_b \mid \mathcal{D})$, which is based on the prior $\pi(\boldsymbol{\beta}, \tau_b)$. The results from the uniform hyperdistribution are stable, since the subsequent marginal posterior distributions are virtually indistinguishable. The opposite occurs for results from the gamma hyperpriors, with very different posterior distributions greatly influenced by the spike near zero of the subsequent hyperprior. The half-normal distribution also exhibits a sensitive performance, with the posterior distributions from HN1 and HN2 practically equal to those from the

395 uniform distribution. As previously noted, the exception is for the posterior distribution
396 from the informative HN3.

Finally, we used a sensitivity measure developed in Roos and Held (2011) to evaluate the relative change in the posterior marginal distribution of the random effects scale parameter with regard to subsequent change in the prior distribution. Changes in both prior and posterior distributions are assessed through the ratio between two Hellinger metrics in the form

$$S(\pi_1, \pi_2) = \frac{H(\pi_1(\theta \mid \mathcal{D}), \pi_2(\theta \mid \mathcal{D}))}{H(\pi_1(\theta), \pi_2(\theta))},$$



(a) ToMV (b) CMV (c) TSWV

Figure 5: *Posterior mean of the regression coefficients associated to plot categories organic (row 1), non organic and greenhouse (row 2), non organic and non greenhouse (row 3), and covariate altitude in logarithmic scale (row 4) for viruses ToMV (column 1), CMV (column 2), and TSWV (column 3) obtained from the full inferential process based on G1, G2 and G3 (black), Un1, Un2 and Un3 (red) and HN1, HN2 and HN3 (green) hyperpriors.*

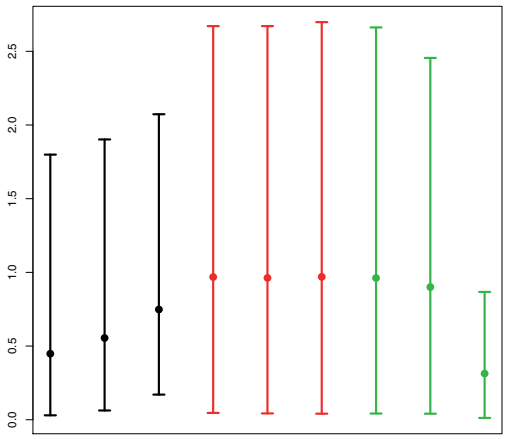


Figure 6: *Posterior mean and 95% credible interval for σ_b obtained from hyperpriors Ga1, Ga2, and Ga3 in black, Un1, Un2, and Un3 in red, and HN1, HN2, and HN3 in green.*

where $\pi_1(\theta \mid \mathcal{D})$ and $\pi_2(\theta \mid \mathcal{D})$ are the subsequent posterior distributions from $\pi_1(\theta)$ and $\pi_2(\theta)$. Note that $S(\pi_1, \pi_2)$ only depends on the Hellinger distance, and consequently, because of its invariancy to any one-to-one transformations we can parameterize the prior and posteriors in terms of τ_b or σ_b.

As expected, sensitivity values with gamma hyperpriors are very relevant,$S(\text{Ga1,Ga2}) = 0.274$ and $S(\text{Ga1,Ga3}) = 0.477$, with calibrated values 0.267 and 0.436 respectively. Thus, considering a Hellinger priors difference such as that reported between the normal distributions $\text{N}(0,1)$ and $\text{N}(1,1)$, their corresponding Hellinger posteriors difference should be understood as equal to that generated between the pair $\text{N}(0,1)$ and $\text{N}(0.267,1)$ in the case of hyperpriors Ga1 and Ga2, $\text{N}(0,1)$ and $\text{N}(0.436,1)$ in the case of Ga1 and Ga3 (see Appendix 2 for more details of calibration). In contrast, sensitivity values associated to uniform hyperpriors are near zero, $S(\text{Un1,Un2}) = 0.017$, $S(\text{Un1,Un3}) = 0.010$, with calibrated values 0.017 and 0.010, despite the Hellinger distance between their corresponding priors being identical in gamma choices. In the case of the half-normal hyperpriors, the sensitivity associated to HN1 and HN2 is small (0.071 and calibrated

value 0.069) but relevant when comparing HN1 and HN3 (S(HN1,HN3) $= 0.588$ and calibrated value 0.576).

4.4. Sensitivity of the risk of plot infection

The risk of plot infection was considered the most appropriate measure to describe results in Section 3 due to its great relevance in agronomic studies. In this sense, the analysis of the variability of the estimates from different modelling prior scenarios could be an important issue, mainly as a measure of confidence and reliability. As it was defined in (4), its posterior estimation will depend on the covariates, regression coefficients and random effects, which show different patterns regarding sensitivity. We carried out a sensitivity analysis for that on a similar basis as that for Section 3: the posterior distribution of the risk infection was calculated for a generic plot situated at altitude 76 meters (the sample median) for each virus and management conditions within each hyperprior scenario.

Table 3 shows the calibration of the Hellinger distance between the posterior distribution of the risk of plot infection for each management condition and virus. Similarly to the particular behaviour of the regression coefficients, the estimation of the risk of plot infection seems to be weakly influenced by the different hyperprior assumptions. In any case, the discrepancies observed between all the comparisons were not greater than the difference between the normal distribution $\mathrm{N}(0,1)$ and $\mathrm{N}(0.583,1)$, which reveals that Hellinger values are in general close to zero. It is worth noting that the Hellinger distance between normal distributions $\mathrm{N}(0,1)$ and $\mathrm{N}(1,1)$ is 0.343 (see again Table 4 in Appendix 2). In a similar manner, the uniform distributions had the smallest discrepancies together with half-normal distributions HN1 and HN2. However, as we expected inferences from HN3 exhibited the greatest discrepancies. Gamma hyperpriors showed substantial discrepancies, above all between Ga1 and Ga3, although these differences did not exceed those obtained from hyperprior HN3. Thus, these outcomes seem to in-

dicate that the particular choice of a hyperprior distribution influences the estimation of the risk infection weakly but in a major extent that in the case of the estimates of the regression coefficients.

Table 3: *Calibration of the Hellinger distance between the posterior marginal distribution of the risk infection computed from hyperprior distributions Ga1 and Ga2, Ga1 and Ga3, Un1 and Un2, Un1 and Un3, HN1 and HN2, and HN1 and HN3.*

Virus	Management	(Ga1,Ga2)	(Ga1,Ga3)	(Un1,Un2)	(Un1,Un3)	(HN1,HN2)	(HN1,HN3)
ToMV	Organic	0.087	0.234	0.011	0.014	0.041	**0.583**
	Non-organic, greenhouse	0.051	0.139	0.011	0.011	0.029	0.355
	Non-organic, no greenhouse	0.041	0.100	0.015	0.016	0.031	0.268
CMV	Organic	0.079	0.213	0.015	0.014	0.041	0.536
	Non-organic, greenhouse	0.039	0.107	0.012	0.010	0.028	0.285
	Non-organic, no greenhouse	0.053	0.142	0.009	0.012	0.028	0.369
TSWV	Organic	0.049	0.128	0.026	0.025	0.037	0.323
	Non-organic, greenhouse	0.040	0.103	0.014	0.009	0.029	0.280
	Non-organic, no greenhouse	0.053	0.142	0.013	0.011	0.030	0.380

There are not so many discrepancies among the posterior means of the risk of a plot infection from the different hyperprior scenarios but there are many in the posterior variabilities (see Table 4). We accounted for variability in terms of standard deviation because it is a measure which describes the grade of uncertainty of the quantity of interest but mainly due to its direct agronomic interpretation. A great similarity in the posterior standard deviation values is repeatedly appreciated in results derived from Un1, Un2, Un3, HN2 and HN2 scenarios. The HN3 value was the most different. However, estimates corresponding to Ga1, Ga2 and Ga3 vary the most, especially in the case of Ga1.

Table 4: *Posterior standard deviation of the risk of a plot infection from the full inferential process based on Ga1, Ga2, Ga3, Un1, Un2, Un3, HN1, HN2 and HN3 hyperpriors.*

Virus	Management	Ga1	Ga2	Ga3	Un1	Un2	Un3	HN1	HN2	HN3
ToMV	Organic	0.136	0.146	0.161	0.184	0.184	0.184	0.183	0.178	0.118
	Non-organic, greenhouse	0.217	0.224	0.235	0.252	0.252	0.253	0.251	0.248	0.206
	Non-organic, no greenhouse	0.118	0.123	0.131	0.147	0.147	0.147	0.147	0.142	0.109
CMV	Organic	0.119	0.127	0.140	0.161	0.161	0.162	0.161	0.156	0.102
	Non-organic, greenhouse	0.161	0.166	0.175	0.190	0.190	0.190	0.189	0.186	0.151
	Non-organic, no greenhouse	0.179	0.186	0.198	0.216	0.216	0.216	0.215	0.211	0.166
TSWV	Organic	0.066	0.071	0.078	0.092	0.093	0.093	0.092	0.088	0.057
	Non-organic, greenhouse	0.172	0.178	0.187	0.203	0.202	0.202	0.201	0.198	0.162
	Non-organic, no greenhouse	0.185	0.192	0.204	0.223	0.223	0.224	0.222	0.218	0.172

In this sense, the posterior standard deviation for risk of a plot infection exhibits a considerable sensitivity to hyperparameter specification. For instance, the risk of a ToMV infection of a generic plot in an organic management system was estimated from 0.028 to 0.553 with 95% probability according to Ga1 scenario, but the subsequent interval in the Un1 scenario was [0.008,0.734].

5. Discussion

In this paper we have proposed a Bayesian correlated model (GLMM) to study and compare the risk of several virus infections in tomato and pepper plots under different agroecosystem conditions. First, we estimated several models, maintaining model specification but varying prior scenario default in accordance with different hyperprior distributions for the random effects scale parameter. Next, we conducted a sensitivity analysis to select the most stable model, in which effects of management conditions, altitude and random individual effects were assessed by estimating different derived quantities considered to be agronomically relevant.

Regarding the model covariates effect, the risk of plot infection was the quantity chosen to analyse agronomic outcomes. The risk of plot infection was estimated in the framework of mixed infections (with more than one virus) as well as in single infections

(with only one virus). All the quantities applied for a "generic" plot of the population of each one of the agroecosystems considered. In the case of single infections, risk difference was also used to quantify differences among agroecosystems. Individual random effects were evaluated by assessing differences in the estimation of the risk of infection among plots managed under similar agroecosystem conditions. This enables the evaluation of the contribution of the common and of the individual elements in the model, and therefore the explanatory capacity of covariates.

In the case of mixed infections, organic agroecosystems exhibited lower prevalence for a three viruses joint infection. Non organic plots, independently of the presence of a greenhouse structure, showed a similar behaviour. Single infections were generally less prevalent or similar in organic systems than in conventional (non-organic with and without greenhouse), while TSWV and CMV infections were less prevalent under organic management; ToMV infection showed a slightly different behaviour pattern possibly as a consequence of the way it is transmitted (mechanical transmission). Altitude effect was clearly negative in all viruses but displayed considerable variability among the three viruses. Random effects behaviour was very regular in individuals belonging to non-organic with greenhouse and non-organic with no greenhouse considering that individual effects did not generate great differences among plots' risk infection estimates. Organic individuals exhibited more variable results in this aspect, but in general we can assume that all the fixed effects included in the model have a good explanatory capacity.

Sensitivity analysis was based on the methodology developed by Roos and Held (2011) and Roos et al. (2015). Hellinger distance and sensitivity measure, together with their corresponding calibration, allowed us to assess discrepancies in the estimation of the fixed effects (regression coefficients), the random effects standard deviation σ_b as well as the "generic" risk of infection among the prior scenarios tested. The evaluation of the posterior mean of the regression coefficients, the graphical characterization of the

499 marginal posterior distribution of σ_b and the assessment of the standard deviation of the
500 posterior distribution of the risk of plot infection among the several modelling scenarios
501 completed the analysis. The outcomes obtained exhibited an insensitive behaviour of
502 the fixed effects to hyperprior alterations with Hellinger values very close to zero and
503 to each other. Only visual analysis of posterior means enabled us to detect a certain
504 instability among inferences obtained from models under gamma hyperdistributions.

505 The estimation of σ_b showed a highly sensitive behaviour: gamma hyperpriors re-
506 peatedly exhibited the most relevant differences showing the greatest sensitivity values
507 and the most divergent posterior distributions. In the case of risk infection estimation, in
508 spite of all the Hellinger distances were around zero, gamma hyperdistributions showed
509 interesting differences in terms of the standard deviation of the posterior distribution of
510 the risk of plot infection. We therefore agree with Browne and Draper (2006), Roos et al.
511 (2015), Roos and Held (2011), Gelman (2006), and Lunn et al. (2009) that gamma hy-
512 perpriors in hierarchical models lack robustness and a sensitivity analysis must be carried
513 out in the Bayesian hierarchical framework to assess reliability of the performance. Fur-
514 thermore, we also conclude that the "noninformative" nature of a hyperprior does not
515 guarantee its impartiality in the inference process.

516 Appendix 1. The Hellinger distance

The Helliger distance (Le Cam, 2012) is a symmetric and invariant to any one-to-one transformation measure of discrepancy between two probability distributions, f and g, defined as follows

$$H(f,g) = \sqrt{\frac{1}{2}\int_{-\infty}^{+\infty}\left(\sqrt{f(u)}-\sqrt{g(u)}\right)^2 \mathrm{d}u},$$

517 where $0 \leq H(f,g) \leq 1$, 0 represents no divergence, and 1 full divergence.

518 The Hellinger distances between two gamma, uniform and half-truncated distribu-

519 tions are

- for gamma densities $\text{Ga}(\alpha_1,\beta_1)$ and $\text{Ga}(\alpha_2,\beta_2)$

$$H^2(\text{Ga}(\alpha_1,\beta_1),\text{Ga}(\alpha_2,\beta_2)) = 1-\Gamma\left(\frac{\alpha_1+\alpha_2}{2}\right)\sqrt{\frac{\beta_1^{\alpha_1}\beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)(\frac{\beta_1+\beta_2}{2})\alpha_1+\alpha_2}}$$

- for uniform densities $\text{Un}(0,\eta_1)$ and $\text{Un}(0,\eta_2)$, with $\eta_1 \leq \eta_2$

$$H^2(\text{Un}(0,\eta_1),\text{Un}(0,\eta_2)) = 1-\left(\frac{\eta_1}{\sqrt{\eta_1\eta_2}}\right)$$

- for half-normal densities $\text{HN}(0,\sigma_1^2)$ and $\text{HN}(0,\sigma_2^2)$

$$H^2(\text{HN}(0,\sigma_1),\text{HN}(0,\sigma_2)) = 1-\frac{\frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}^{1/4}}{\sqrt{\frac{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}{2}}}$$

520 In the case of posterior distributions $\pi_1(\boldsymbol{\theta} \mid \mathcal{D})$ and $\pi_2(\boldsymbol{\theta} \mid \mathcal{D})$, the Hellinger distance
521 can be approximated numerically at a finite set of K integration points as follows

$$H^2(\pi_1(\boldsymbol{\theta} \mid \mathcal{D}),\pi_2(\boldsymbol{\theta} \mid \mathcal{D})) = \frac{1}{2}\sum_{k=1}^{K}\left(\sqrt{\pi_1(\boldsymbol{\theta} \mid \mathcal{D})(k)}-\sqrt{\pi_2(\boldsymbol{\theta} \mid \mathcal{D})(k)}\right)^2\Delta_k, \quad \text{(A.1)}$$

522 where the weights Δ_k are provided by the trapezoidal rule.

523

524 Appendix 2. Calibration

The Hellinger distance can be calibrated to evaluate the importance of the observed discrepancies by means of a reference parameter. Calibration was undertaken with respect to the normal distribution with variance one. The Hellinger distance between densities

$\mathrm{N}(0,1)$ and $\mathrm{N}(\mu,1)$ is

$$H(\mathrm{N}(0,1),\mathrm{N}(\mu,1)) = \sqrt{1-\exp(-\mu^2/8)},$$

and consequently

$$\mu = \sqrt{-8\log(1-H^2(\mathrm{N}(0,1),\mathrm{N}(\mu,1)))}$$

Table A.2.1 shows a range of calibrated values μ with its subsequent Hellinger distance, $H(\mathrm{N}(0,1),\mathrm{N}(\mu,1))$.

Table A.2.1: *Calibration of the Hellinger distance.*

μ	$H(N(0,1),N(\mu,1))$
0	0
1	0.343
2	0.627
3	0.822
4	0.930
5	0.978
6	0.994
7	0.999
8	0.999
9	0.999
10	1

The sensitivity measure introduced previously can also be calibrated. Calibration of the sensitivity value obtained, *s*, has been obtained following the subsequent equation:

$$C(s,\mu') = \mu(s \times H(\mathrm{N}(0,1),\mathrm{N}(\mu',1))) \tag{A.2}$$

Interpretation of calibration can be conditioned by the choice of μ', so that for a value $\mu' = 1$, the value of *s*, would be comparable with the Hellinger distance obtained between two normal priors, $\mathrm{N}(0,1)$ and $\mathrm{N}(\mu' = 1,1)$ and the subsequent normal posteriors,

534 $\mathrm{N}(0,1)$ and $\mathrm{N}(C(s,\mu'=1),1)$. It is important to note that if $s>1$ then $C(s,\mu')>\mu'$; if
535 $s<1$ then $C(s,\mu')<\mu'$; and if $s=1$ then $C(s,\mu')=\mu'$.

536 Acknowledgements

537 We are very grateful to Josep Roselló, José Serra and MÂ[a] José Muñoz for providing
538 the data and useful comments. We thank Marie Savage very much for reviewing the
539 English and editing the manuscript. Elena Lázaro research was funded by the Spanish
540 Ministry of Education, Culture and Sports, Grant FPU 2013/02042. The work of Carmen
541 Armero was partially supported by Grant MTM2016-77501-P from the Spanish Ministry
542 of Economy and Competitiveness. The work of Luis Rubio was partially funded by the
543 INIA Grant RTA2013-00047-C02. We wish to acknowledge two anonymous referees
544 and the Associate Editor for their valuable comments that substantially improved the
545 original version of the paper.

546 References

547 Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data.
548 *Journal of the American Statistical Association*, 88, 669–679.

549 Alvares, D., Armero, C., Forte, C., and Rubio, L. (2016). Exploring Bayesian models to evaluate
550 control procedures for plant disease. *Statistics and Operations Research Transactions,*
551 *SORT*, 40, 139–152.

552 Bengtsson, J., Ahnström, J., and Weibull, A. C. (2005). The effects of organic agriculture on
553 biodiversity and abundance: a meta-analysis. *Journal of Applied Ecology*, 42, 261–269.

554 Bettiol, W., Ghini, R., Galvão, J. A. H., and Siloto, R. C. (2004). Organic and conventional
555 tomato cropping systems. *Scientia Agricola*, 61, 253–259.

556 Box, G. E., and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. Hoboken: John
557 Wiley & Sons.

558 Browne, W. J. and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods
559 for fitting multilevel models. *Bayesian Analysis*, 1, 473–514.

560 Clark, M. F., Adams, A. N. and Barbara, D. J. (1976). The detection of plant viruses by enzyme-
561 linked immunosorbent assay (ELISA). In *X International Symposium on Fruit Tree Virus*
562 *Diseases* 67, 43–50.

563 Clark, J. S., Wolosin, M., Dietze, M., Ibanez, I., Ladeau, S., Welsh, M. and Kleoppel, B. (2007).
564 Tree growth inference and prediction from diameter censuses and ring widths. *Ecological*
565 *Applications*, 17, 1942–1953.

566 Christensen, R., Johnson, W., Branscum, A. and Hanson, T. E. (2011). *Bayesian Ideas and*
567 *Data Analysis: An Introduction for Scientists and Statisticians*. Boca Raton: Chapman &
568 Hall/CRC Press.

569 Finley, A. O., Banerjee, S., and Basso, B. (2011). Improving Crop Model Inference Through
570 Bayesian Melding with Spatially-Varying Parameters. *Journal of Agricultural, Biological*
571 *and Environmental Statistics*, 16, 453–474.

572 Gallitelli, D. (2000). The ecology of Cucumber mosaic virus and sustainable agriculture. *Virus*
573 *Research*, 71, 9–21.

574 García-Cano, E., Resende, R. O., Fernández-Muñoz, R. and Moriones, E. (2006). Synergis-
575 tic interaction between Tomato chlorosis virus and Tomato spotted wilt virus results in
576 breakdown of resistance in tomato. *Phytopathology*, 96, 1263–1269.

577 Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment
578 on article by Browne and Draper). *Bayesian Analysis*, 1, 515–534.

579 Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation using Multiple Se-
580 quences. *Statistical Science*, 7, 457–511.

581 Hanssen, I. M., Lapidot, M. and Thomma, B. P. (2010). Emerging viral diseases of tomato crops.
582 *Molecular Plant-Microbe Interactions*, 23, 539–548.

583 Jeffreys, H. (1998). *The Theory of Probability*. Third edition. New York: Oxford University
584 Press.

585 Krachey, M. and Boone, E.L. (2012). `bmk`: MCMC diagnostics package. R package version
586 1.0.
587 *http://CRAN.R-project.org/package=bmk*.

588 Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R. and Jones, D.R. (2005). How vague is
589 vague? A simulation study of the impact of the use of vague prior distributions in MCMC
590 using WinBUGS. *Statistics in Medicine*, 24, 2401–2428.

591 Le Cam, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-
592 Verlag.

593 Letourneau, D.K. and Goldstein, B. (2001). Pest damage and arthropod community structure in
594 organic vs. conventional tomato production in California. *Journal of Applied Ecology*,
595 38, 557–570.

596 Loredo, T.J. (1990). From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics
597 (1990). In *Maximum Entropy and Bayesian Methods* (P.F. Fougère eds), 81-142. Dor-
598 drecht: Kluwer Academic.

599 Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2012). *The BUGS Book: A*
600 *practical Introduction to Bayesian Analysis*. Boca Raton: Chapman & Hall/CRC Press.

601 Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – A Bayesian mod-
602 elling framework: Concepts, Structure and Extensibility. *Statistics and Computing*, 10,
603 325–337.

604 Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS Project: Evolution,
605 Critique and Future Directions. *Statistics in Medicine*, 28, 3049–3067.

606 McCulloch, R.E. (1989). Local Model Influence. *Journal of the American Statistical Associa-*
607 *tion*, 84, 473–478.

608 Murphy, J.F. and Bowen, K.L. (2006). Synergistic disease in pepper caused by the mixed infec-
609 tion of Cucumber mosaic virus and Pepper mottle virus. *Phytopathology*, 96, 240–247.

610 Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Hoboken: John Wiley & Sons.

611 Paciorek, C.J. and McLachlan, J.S. (2009). Mapping Ancient Forests: Bayesian Inference for
612 Spatio-Temporal Trends in Forest Composition Using the Fossil Pollen Proxy Record.
613 *Journal of the American Statistical Association*, 104, 608–622.

614 Paradinas, I., Conesa, D., Pennino, MG., Muñoz, F., Fernández, AM., López-Quílez, A. and
615 Bellido, JM. (2015). Bayesian spatio-temporal approach to identifying fish nurseries by
616 validating persistence areas. *Marine Ecology Progress Series*, 52, 245–255.

617 Roos, M. and Held, L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models
618 for binary data. *Bayesian Analysis*, 6, 259–278.

619 Roos, M., Martins, T. G., Held, L. and Rue, H. (2015). Sensitivity analysis for Bayesian hierar-
620 chical models. *Bayesian Analysis*, 10, 321–349.

621 Serra, J., Ocon, C., Jiménez, A., Arnau, J., Malagón, J. and Porcuna, J.L. (1999). Epidemiología
622 de las virosis en la Comunidad Valenciana: el caso del “virus de la cuchara” del tomate.
623 *Comunidad Valenciana Agraria*, 14, 47–53.

624 Spiegelhalter, D.J., Abrams, K.R. and Myles, J.P. (2004). *Bayesian Approaches to Clinical trials*
625 *and Health-Care evaluation*. Chichester: John Wiley & Sons.

626 Thompson, S.G., Smith, T.C. and Sharp, S.J. (1997). Investigating underlying risk as a source of
627 heterogeneity in meta-analysis. *Statistics in Medicine*, 16, 2741–2758.

628 Thornley, J.H. and France, J. (2007). *Mathematical Models in Agriculture: Quantitative Methods*
629 *for the Plant, Animal and Ecological Sciences*. Oxon: Cabi.

630 Tomlinson, J.A. (1987). Epidemiology and control of virus diseases of vegetables. *Annals of*
631 *Applied Biology*, 110, 661–681.

632 Van Bruggen, A.H. (1995). Plant disease severity in high-input compared to reduced-input and
633 organic farming systems. *Plant Disease*, 79, 976–984.

References

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:669–679.

Alvares, D., Armero, C., Forte, C., and Rubio, L. (2016). Exploring Bayesian models to evaluate control procedures for plant disease. *Statistics and Operations Research Transactions, SORT* 40(1), 139–152.

Bengtsson, J., Ahnström, J., and Weibull, A. C. (2005). The effects of organic agriculture on biodiversity and abundance: a meta-analysis. *Journal of Applied Ecology* 42(2), 261–269.

Bettiol, W., Ghini, R., Galvão, J. A. H., and Siloto, R. C. (2004). Organic and conventional tomato cropping systems. *Scientia Agricola* 61(3), 253–259.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian Computation and Stochastic Systems. *Statistical Science* 10(1), 3–41.

Box, G. E., and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. Hoboken: John Wiley & Sons.

Brooks, S. (1998). Markov chain Monte Carlo method and its application *Journal of the Royal Statistical Society: Series D (The Statistician)* 47(1), 69–100.

Browne, W. J., and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* 1(3), 473–514.

Clark, M. F., Adams, A. N., and Barbara, D. J. (1976). The detection of plant viruses by enzyme-linked immunosorbent assay (ELISA). In *X International Symposium on Fruit Tree Virus Diseases* 67, 43–50.

Clark, J. S., Wolosin, M., Dietze, M., Ibanez, I., Ladeau, S., Welsh, M., and Kleoppel, B. (2009). Tree growth inference and prediction from diameter censuses and ring widths. *Ecological Applications* 17, 1942–1953.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2011). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. Boca Raton: Chapman & Hall/CRC Press.

Finley, A. O., Banerjee, S., and Basso, B. (2011). Improving Crop Model Inference Through Bayesian Melding with Spatially-Varying Parameters. *Journal of Agricultural, Biological, and Environmental Statistics* 16(4), 453-474.

Gallitelli, D. (2000). The ecology of Cucumber mosaic virus and sustainable agriculture. *Virus Research* 71(1), 9–21.

García-Cano, E., Resende, R. O., Fernández-Muñoz, R., and Moriones, E. (2006). Synergistic interaction between Tomato chlorosis virus and Tomato spotted wilt virus results in breakdown of resistance in tomato. *Phytopathology* 96(11), 1263–1269.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 1(3), 515–534.

Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation using Multiple Sequences. *Statistical Science* 7, 457–511.

Hanssen, I. M., Lapidot, M., and Thomma, B. P. (2010). Emerging viral diseases of tomato crops. *Molecular Plant-Microbe Interactions* 23(5), 539–548.

Jeffreys, H. (1998). *The Theory of Probability*. Third edition. New York: Oxford University Press.

Krachey, M. and Boone, E. L. (2012). `bmk`: MCMC diagnostics package. R package version 1.0. *http://CRAN.R-project.org/package=bmk*.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., and Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* 24(15), 2401–2428.

Le Cam, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag.

Letourneau, D. K., and Goldstein, B. (2001). Pest damage and arthropod community structure in organic vs. conventional tomato production in California. *Journal of Applied Ecology* 38(3), 557–570.

Loredo, T. J. (1990). From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics (1990). In *Maximum Entropy and Bayesian Methods* (P. F. Fougère eds), 81-142. Dordrecht: Kluwer Academic.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012). *The BUGS Book: A practical Introduction to Bayesian Analysis*. Boca Raton: Chapman & Hall/CRC Press.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, Structure and Extensibility. *Statistics and Computing* 10(4), 325–337.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The BUGS Project: Evolution, Critique and Future Directions. *Statistics in Medicine* 28(25), 3049–3067.

McCulloch, R. E. (1989). Local Model Influence. *Journal of the American Statistical Association*84(406), 473–478.

Murphy, J. F., and Bowen, K. L. (2006). Synergistic disease in pepper caused by the mixed infection of Cucumber mosaic virus and Pepper mottle virus. *Phytopathology* 96(3), 240–247.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Hoboken: John Wiley & Sons.

Paciorek, C. J. and McLachlan, J. S. (2009). Mapping Ancient Forests: Bayesian Inference for Spatio-Temporal Trends in Forest Composition Using the Fossil Pollen Proxy Record. *Journal of the American Statistical Association* 104, 608–622.

Paradinas, I., Conesa, D., Pennino, MG., Muñoz, F., Fernández, AM., López-Quílez, A., and Bellido, JM. (2015). Bayesian spatio-temporal approach to identifying fish nurseries by validating persistence areas. *Marine Ecology Progress Series* 52, 245–255.

Roos, M. and Held, L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis* 6(2), 259–278.

Roos, M., Martins, T. G., Held, L., and Rue, H. (2015). Sensitivity analysis for Bayesian hierarchical models. *Bayesian Analysis* 10(2), 321–349.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71(2), 319–392.

Serra, J., Ocon, C., Jiménez, A., Arnau, J., Malagón, J., and Porcuna, J. L. (1999). Epidemiología de las virosis en la Comunidad Valenciana: el caso del “virus de la cuchara” del tomate. *Comunidad Valenciana Agraria* 14, 47–53.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical trials and Health-Care evaluation*. Chichester: John Wiley & Sons.

Thompson, S. G., Smith, T. C. and Sharp, S. J. (1997). Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* 16, 2741–2758.

Thornley, J. H. and France, J. (2007). *Mathematical Models in Agriculture: Quantitative Methods for the Plant, Animal and Ecological Sciences*. Oxon: Cabi.

Tomlinson, J. A. (1987). Epidemiology and control of virus diseases of vegetables. *Annals of Applied Biology* 110(3), 661–681.

Van Bruggen, A. H. (1995). Plant disease severity in high-input compared to reduced-input and organic farming systems. *Plant Disease* 79(10), 976–984.

Zeger, S. L., and Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach.*Journal of the American statistical association* 86(413), 79–86.