We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 6

Using Bayesian Inference to Investigate the Influence of Environmental Factors on a Phytoplasma Disease

Bernd Panassiti

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74637

Abstract

Phytoplasma diseases cause major economic damage on crops worldwide. To draw inferences from such a system, joint estimation of dependencies and high flexibility in the model structure are required. Using Bayesian inference, the aim of this chapter was to infer the apple proliferation (AP) disease epidemiology in South Tyrol, Italy. The data consisted of (1) presence/absence of the AP vector *Cacopsylla picta* collected in 44 orchards in 2014; (2) prevalence of the AP pathogen "*Candidatus* Phytoplasma mali" in the vector population; and (3) AP symptomatic trees visually assessed in 2015. Generalized linear mixed models evaluated in a Bayesian framework were used to test species-environment relationships. The model results indicated that the occurrence of the AP vector and symptomatic plants are positively influenced by elevation and temperature and negatively by management. Vector and pathogen predictions in the disease symptoms model correlated negatively or not at all with the prevalence of AP symptoms occurrence. In conclusion, the model results suggest that the presence/absence of the AP vector alone may not be the only cause for disease occurrence. Considering factors such as phytoplasma transmission via root-bridges and specific management strategies, may help to improve inference and finally to optimize the existing pest management.

Keywords: apple proliferation, Bayesian inference, habitat modeling, imperfect detection, latent infections, occupancy model, phytoplasma disease, pest insect, psyllid vector

1. Introduction

Phytoplasma-induced diseases occur in a range of economically important crops and are therefore major threats in agriculture worldwide [1]. Phytoplasma are cell wall-less plant pathogenic bacteria vectored by insects belonging to the order Hemiptera [2].

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

From an ecological perspective, phytoplasma diseases are complex biological systems. Complexity is linked to many sources of uncertainty which in most cases are difficult to measure. Among others, these sources of uncertainty include vector-pathogen-plant interactions, but also the presence of unknown vectors (e.g., *Reptalus panzeri* was recently confirmed as a new vector in the grapevine phytoplasma disease "bois noir" [3]) or the time (i.e., latent period) between pathogen infection and symptom expression (in case of plants) or the ability to retransmit the pathogen (in case of vectors).

Besides complexity, the statistical treatment of the inherent dependencies of such biological systems represents another challenge in the modeling process. Traditional statistical methods [such as generalized linear models (GLM), generalized linear mixed models (GLMM)] could be used in a step-wise approach. In a first step, the vector-environment relationship is identified (vector model). Second, using the results of the vector model, the pathogen-environment relationship is established (pathogen model). Finally, the results of both previous models are used to fit the plant disease model. However, this approach does not consider the dependencies between the responses at the same time. In contrast, methods that allow for combined dependencies such as structural equation modeling lack the flexibility in model specification [4]. One solution is using Bayesian inference which allows to jointly estimate the model parameters and at the same time offers high flexibility in defining the model structure.

1.1. Case study: apple proliferation

In this study, the phytoplasma disease apple proliferation (AP) was chosen as a modeling system. AP-specific disease symptoms on apple trees are the proliferation of auxiliary shots (formation of witches' broom) and enlarged stipulae. AP nonspecific disease symptoms include early leaf reddening, small, taste- and colorless fruits, chlorosis and premature bud break. The causal AP agent is '*Candidatus* Phytoplasma mali' [5]. In infected apple trees, the phytoplasma resides in phloem tubes and is transmitted by phloem-sucking insect vectors during feeding activity. In South Tyrol, Northern Italy, the most efficient AP vector is *Cacopsylla picta* (Hemiptera, Psyllidae) [6].

The aim of this chapter was to jointly infer the AP disease epidemiology in South Tyrol, Italy using Bayesian inference. Imperfect detection was accounted for in the vector and symptomatic plant models. The AP insect vector was modeled using an occupancy model. To account for detection bias during the vector sampling, information on sampling effort was used as a predictor in an additional Bernoulli process conditional on the AP vector's true presence or absence. Based on molecular analyses of AP prevalences in apple trees, I estimated the proportion of latent infected trees to account for imperfect detection of truly AP phytoplasma-infected apple trees.

2. Materials and methods

2.1. Data

The AP vector *C. picta* and AP symptomatic apple trees were surveyed at 44 and 26 orchards in South Tyrol, Northern Italy, respectively. Prevalences of the AP phytoplasma '*Ca*.

http://dx.doi.org/10.5772/intechopen.74637

Phytoplasma mali' within the vectors were available from 28 orchards (**Figure 1**). Insect vectors were caught using the "beating tray" method [7, 8]. Depending on orchard size, between 20 and 200 apple trees were randomly selected for vector sampling which was carried out in 2014. Collected vectors were identified according to Ossiannilsson [9]. Insect vectors were then molecularly analyzed for the presence of the pathogen '*Ca*. Phytoplasma mali'. Phytoplasma detection based on a SYBR green real-time PCR was carried out as described in [6]. AP infection status of apple trees was assessed by trained and experienced professionals using visual inspection. Apple cultivars were Golden Delicious and Gala. The monitoring of AP symptoms started in 2013 and each year new AP symptomatic trees were recorded. In most cases, disease symptoms appear 1 year following an infection with '*Ca*. Phytoplasma mali' [10]. To take into account this latent period, monitoring data of AP symptoms from 2015 were used. Given the inspectors' skills to detect AP symptoms and independent molecular

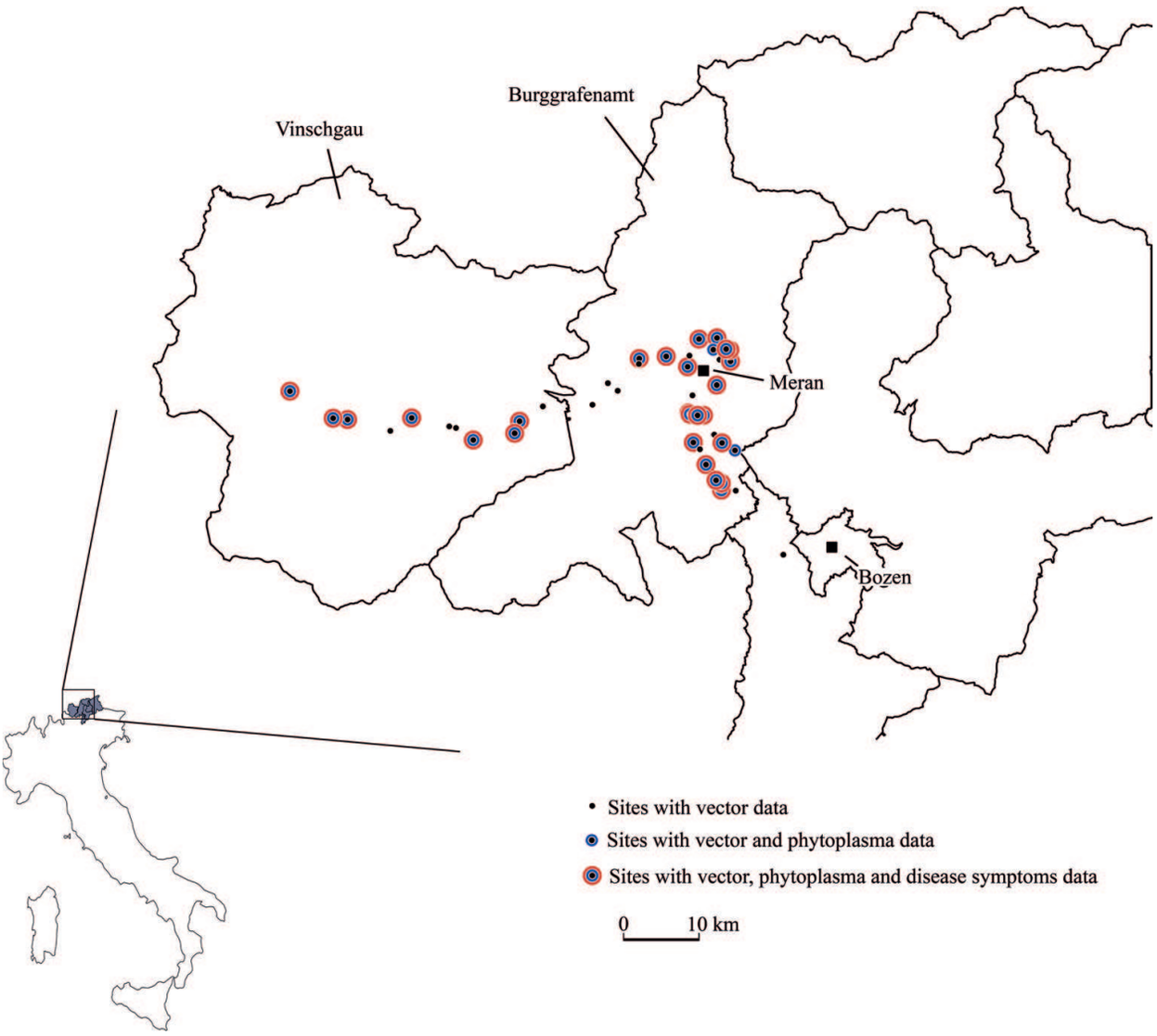


Figure 1. Sampling sites of vector, phytoplasma and plant disease symptoms of the apple proliferation epidemiology in South Tyrol, Northern Italy.

analysis of AP symptomatic trees, the false-positive rate can be assumed to be approximately zero. Symptomatic means that at least one specific AP symptom or a combination of at least two unspecific AP symptoms was present.

A summary of the final data set including the AP vector, the AP phytoplasma and AP symptoms of trees is provided in **Table 1**. Metric environmental predictors included elevation (m a.s.l.) and annual mean temperature (°C). Orchards were classified into integrated/not integrated management to account for different pest management strategies.

2.2. Modeling approach

Bayesian inference was used to jointly estimate the dependencies of all responses (AP vector, AP phytoplasma prevalences of the vector, AP symptoms of apple trees) and the environment. To fit the model, all environmental predictors (except vector and phytoplasma predictions) were scaled and centered (i.e., mean subtracted and divided by the standard deviation) to allow a faster convergence of the model fitting algorithm. To decide whether to account for unimodal response-curves, in a pre-step, I fitted multivariate GLMs including quadratic terms of elevation and temperature [11]. As the unimodal response curves were not found to be ecologically sensible, only linear relationships were considered in subsequent analysis. Generalized linear models were developed using a binomial error distribution and a logit link function (GLMM; [12, 13]). The GLMMs were then evaluated in a Bayesian framework. As prior distributions for the fixed effects, zero-centered normal distributions were used. Except for the intercept, priors were defined to be mildly informative which results in a shrinkage effect similar to a ridge-regression [14].

The vector data set, as is common for ecological data, contained many zero values due to the rarity and detectability of the species. To account for imperfect detection, I used a site-occupancy model [15, 16]. These models rely on the "closure assumption" stating that the occupancy state remains unchanged between survey times. The occupancy model combines (1) an ecological process and (2) an observation process. The ecological process of the true occupancy state z (which is a latent or unobserved variable) can be described

	Min	Q1	Median	Q3	Max	Mean	sd	n	Missing
Observed vector	0	0	1	1	1	0.659	0.479	44	0
Vectors analyzed	1	1	2	3.25	35	4.29	7.91	28	0
Vectors Inf	0	0	0	1	8	0.929	1.98	28	0
Tree_total	390	950	1171	1764	3065	1365	613	26	0
Tree_inf	0	0	2.5	5	111	9.46	24.5	26	0

Table 1. Summary of vector, pathogen and disease symptoms of plants used to fit the apple proliferation (AP) joint model.

http://dx.doi.org/10.5772/intechopen.74637

using a Bernoulli distribution with the occupancy probability Ψ for each surveyed site (indexed with i):

$$z_i \sim \text{Bernoulli}(\Psi_i)$$

In the observation process, real observations (detections/nondetections) for each survey time (indexed by j) follow a Bernoulli distribution conditional on the true occupancy state z:

$$y_{ij} \sim \text{Bernoulli}(z_i\, p_{ij})$$

where p is defined as the detection probability at site i and survey time j given the site was actually occupied. The detection probability was modeled using a logistic regression and sampling effort as explanatory variable. Sampling effort was defined as the number of sampled trees in proportion to the total number of surveyed trees for AP symptoms.

Field surveys on the prevalence of plant diseases caused by plant-pathogenic bacteria are often based on visual diagnosis of disease symptoms [17–19]. Given trained and experienced plant inspectors, the false-positive rate can be assumed to be close to zero. The false negative rate is also often considered very small because latent infections are mostly ignored. Based on molecular analyses latent infections for the AP disease were found to be 2.32 and 10.48% depending on age of the apple trees [20]. To account for imperfect detection caused by latent infections, an informative beta prior was used for the detection probability p with parameters $a = 2$ and $b = 80$. The specified prior distribution has a mean value of 0.02 and a 95% quantile of 0.06. Hence, in the observation process of the AP disease symptoms model, AP symptoms detections/nondetections were drawn from a Binomial distribution as follows:

$$y_i \sim \text{Binomial}(N_i, z_i(1 - p_i))$$

where N is the total number of survey trees for each site.

MCMC sampling was carried out by the STAN software (RStan version 2.12.1), which uses the No-U-Turn sampler (NUTS) [21, 22]. Model specifications included three chains with 3000 iterations each and considered a chain to be converged when the potential scale reduction statistic, $\hat{R} < 1.05$ [23]. To access model, fit posterior predictive checks were applied on each model separately using the DHARMa package [24]. The DHARMa package calculates scaled residuals (Bayesian p-values) by comparing observations simulated from the fitted model with observed values. All statistical analyses were carried out in the R statistical environment (version 3.2.2; [25]).

The RStan code for the joint model is available in Appendix A.

3. Results

The marginal posterior distributions of the parameter of interest of the AP joint model are shown in **Figure 2**. For the AP vector *C. picta*, I found that the occurrence of the AP vector was positively correlated with elevation and temperature. The opposite was found for integrated pest management measures which negatively affected the vector occurrence probabilities. The sampling effort represented by the amount of sampled trees seemingly did not affect the detection of the AP vector. Unlike elevation, temperature and pest management, which did not affect the prevalences of the pathogen '*Ca*. Phytoplasma mali' within the AP vector, an assumably positive relationship between AP vector and its phytoplasma infection rates is indicated in **Figure 2**. The 80% credible interval (CrI), however, also indicates that a high uncertainty is associated with the true value of this parameter.

As the vector model, AP symptom occurrences on apple trees were likewise positively correlated with elevation and temperature and negatively with integrated pest management measures. Moreover, the model estimated a negative correlation between the AP vector and AP symptoms. No relationship between phytoplasma infection rates within the AP vector and AP symptoms was found.

Regarding the model performance, the potential scale reduction statistic, $\hat{R}$, for each parameter was close to 1 (not shown). Hence, I found no indication of non-convergence of the three chains.

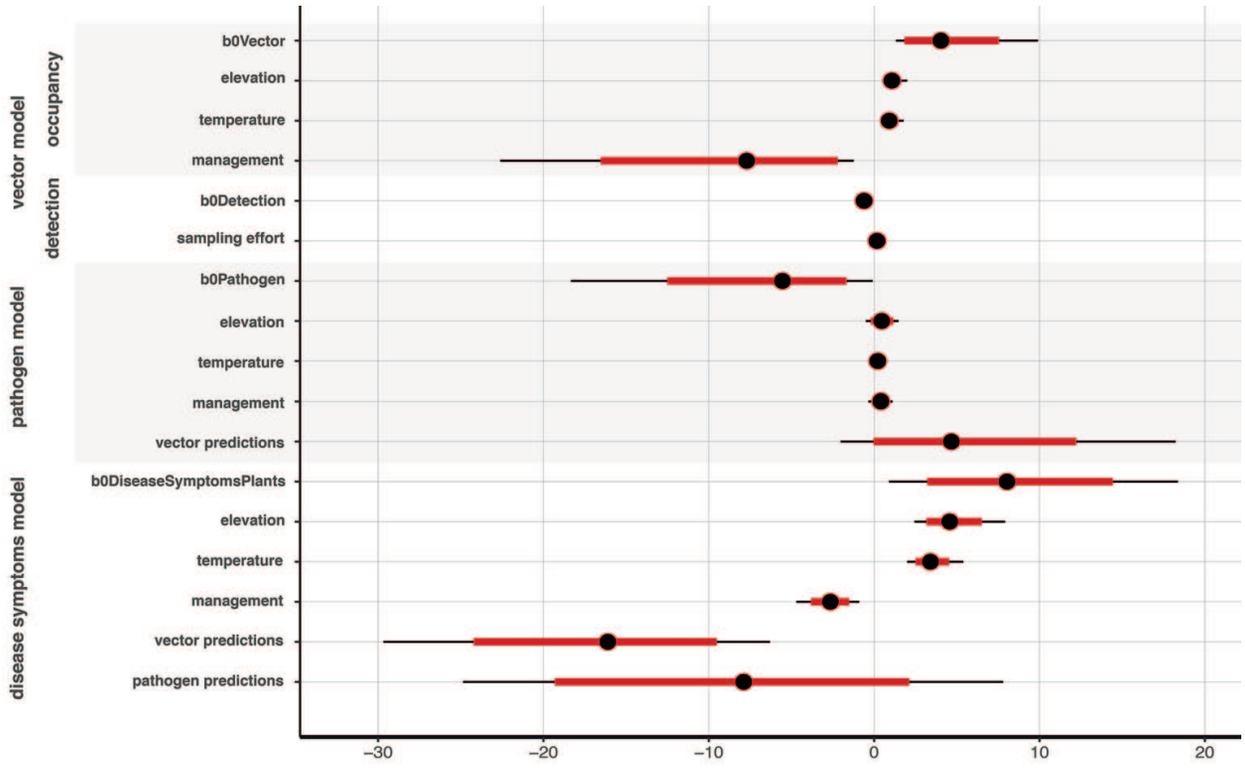


Figure 2. Marginal posterior distributions of all environmental predictors for the vector, pathogen and disease symptoms of plants models. Black intervals denote the 95% credible intervals, and red bands indicate the 80% credible intervals. For each predictor, black bullets denote the medians of the three MCMC chains.

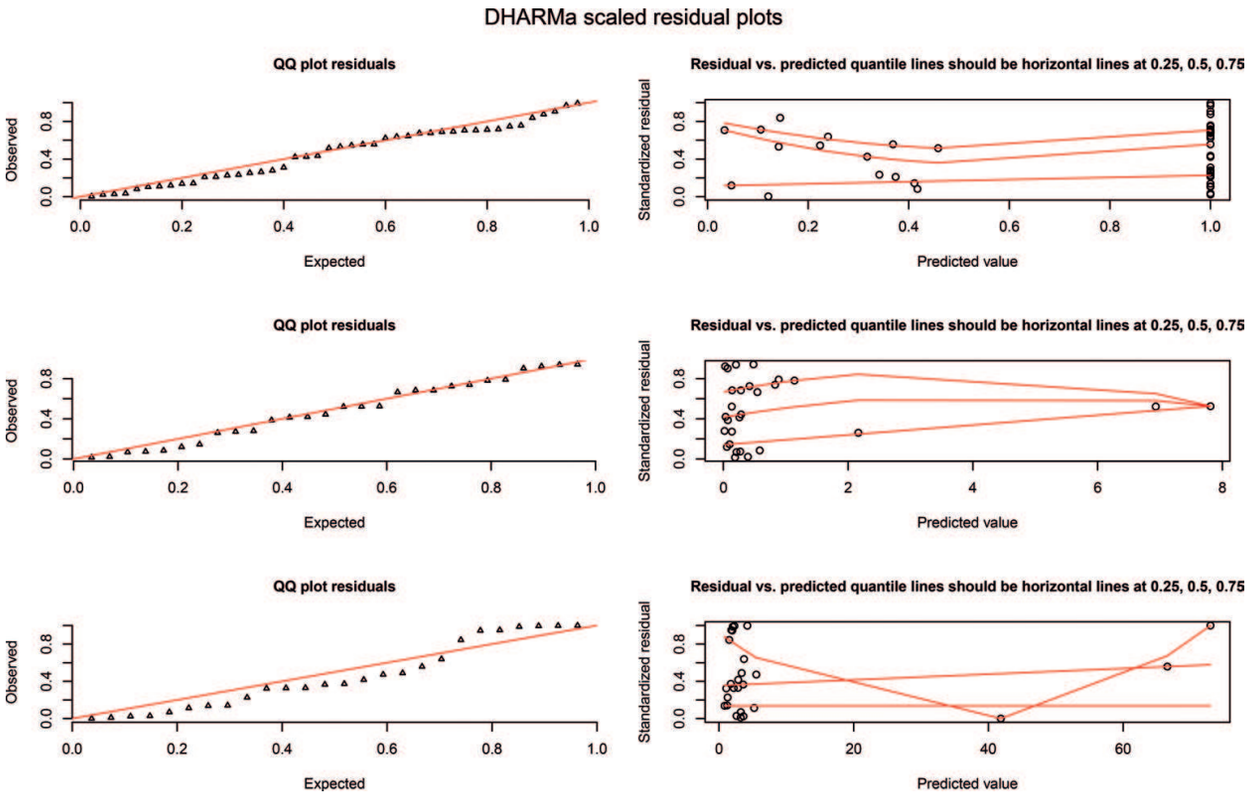


Figure 3. Residual diagnosis for the vector (a), pathogen (b) and disease symptoms of plants (c) models. Residual diagnosis is based on scaled residuals. For each model, qq-plots are shown on the left, and scaled residuals *versus* predicted values are plotted on the right.

Figure 3 shows the results of the residual diagnosis. The plots show no serious violations of distributional assumptions. To confirm the overall uniformity of the scaled residuals, I applied one-sample Kolmogorov–Smirnov tests, which were not significant for all three models.

4. Discussion

The modeling case study presented in this chapter illustrated the use of Bayesian inference to jointly investigate the influence of environment on the occurrence of the AP vector *C. picta*, the prevalences of the AP pathogen ('*Ca*. Phytoplasma mali') within the vector and the occurrence of AP disease symptoms on apple trees.

4.1. Influence of environment on apple proliferation epidemiology

Using the 80% credible interval, I found that AP vector and AP symptoms on apple trees were positively associated with elevation and temperature and negatively with integrated pest management. While having similar ecological requirements, the joint model indicated a negative relationship between vector and symptoms. Elevation, temperature and integrated pest management did not affect AP phytoplasma prevalences within the vector. No correlation was found between prevalences of AP phytoplasma and symptoms.

'*Ca*. Phytoplasma mali'-infection rates of *C. picta* in South Tyrol are usually higher than those of *C. melanoneura,* another AP vector (0.6% compared to 11.6%, [26]). Moreover, *C. picta* is assumed to be the more effective AP vector because it was shown to be able to vertically transmit the pathogen to its offspring [6]. Therefore, the finding that the vector is not correlated with AP symptoms is unexpected but agrees with the vector-symptomatic plant relationship of "bois noir," a phytoplasma disease on grapevines [17]. The authors argued that this discrepancy may be explained by acknowledging the fact that vector's presence alone is not responsible for disease occurrence rather it is important to define the pathogen prevalence in the vector population. In this study, however, no correlation was also found between pathogen predictions and AP symptoms. Intuitively, one would expect more AP symptoms given a high infection rate of the vectors but the marginal posterior distribution of the phytoplasma prevalence in the vector population is associated with a large credible interval and does not allow an interpretation of the true parameter value. The lack of positive correlation between infected vectors and AP symptoms hint at other infection sources not considered in this study. For example, recently, a phytoplasma transmission via root-bridges between apple trees was hypothesized [27].

Even though the joint model did not identify a clear correlation between the predictor variable integrated pest management and pathogen occurrences, overall, it seems that integrated pest management is an important environmental driver, negatively influencing vector, and disease symptom occurrences. But it is also possible that the AP responses are influenced by different management measures. For example, the presence/absence of the vectors may be influenced by application time, quantity and type of insecticides, while new disease incidences in plants may also relate to different levels in the effort of uprooting AP symptomatic trees, thereby eliminating sources of new vector infections or root transmissions to adjacent trees [28]. Hence, in a follow-up study, it would be worth to further investigate which specific management measure leads to a decrease in the responses to optimize insect pest management strategies.

4.2. Advantages of Bayesian inference

Besides jointly estimating the disease system, Bayesian inference allows high flexibility in the model specifications. Models can be easily extended to include detection probabilities, overdispersion or zero-inflation [29–31]. The present joint model could be further extended by including AP symptoms detection probabilities depending on the cultivar and observed symptoms. The high flexibility is also important when data is collected for purposes different than statistical inference and prediction. For example, if vector data was collected to determine the first appearance of the vector in the orchard (to timely optimize the application of insecticides), vector prediction probabilities need to be constraint by probabilities of the true flight period of the pest insect.

Some parameter estimates in this study were associated with large credible intervals, meaning high uncertainty. One solution would be to use a higher number of observations, which is not always feasible in ecological studies. Another possibility is to include informative priors derived from the literature or previous analysis as illustrated for the informative beta prior to account for imperfect AP symptom detection due to latent infections. Priors play an essential role in every Bayesian analysis. For the environmental parameter estimates included in this chapter, no prior information from previous analysis was available. However, the identified relationships could be used to define prior distributions in future studies.

Finally, the results of a Bayesian inference (posterior distributions) can be summarized using, for example, credible intervals which allow an intuitive interpretation of the parameter estimates associated with well-defined uncertainties. Given chain convergence and successful posterior predictive checks, Bayesian credible intervals are also appropriate for small data sets [32]. This is especially true in observational studies on animal and plant populations where data collection is often time- and cost-consuming.

5. Conclusion

In summary, the results of the AP joint model suggested that the presence of the AP vector is not necessarily positively correlated with disease occurrence. Instead, other factors such as phytoplasma transmission via root-bridges or specific management strategies should be additionally considered in future studies. In case of the AP disease system, Bayesian inference allowed to jointly fit combined dependencies which are common to phytoplasma epidemiological diseases. Unlike maximum likelihood methods, posterior distributions for all quantities of interest are obtained which could be further summarized using credible intervals and allowed intuitive interpretation of the results. The provided example of a joint Bayesian modeling framework can be used as a basis to infer species-environment relationships of phytoplasma disease systems.

Acknowledgements

The work was performed as part of the project APPLClust and was funded by the Autonomous Province of Bozen/Bolzano (Italy) and the South Tyrolean Apple Consortium. The author would like to thank Stefanie Fischnaller, Martin Parth, Manuel Messner, Robert Stocker, Christine Kerschbamer and Katrin Janik for providing data on insect vectors, phytoplasma prevalences and occurrences of disease symptoms of apple trees.

Appendix A. Supplementary data

Supplementary data associated with this chapter is available online: https://doi.org/10.6084/m9.figshare.5896789

Author details

Bernd Panassiti

Address all correspondence to: bernd.panassiti@gmail.com

Laimburg Research Center, Auer, Italy

References

[1] Bertaccini A, Duduk B, Paltrinieri S, Contaldo N. Phytoplasmas and phytoplasma diseases: A severe threat to agriculture. American Journal of Plant Sciences. 2014;**5**:1763-1788. DOI: 10.4236/ajps.2014.512191

[2] Alma A, Tedeschi R, Lessio F, Picciau L, Gonella E, Ferracini C. Insect vectors of plant pathogenic Mollicutes in the euro-Mediterranean region. Phytopathogenic Mollicutes. 2015;**5**:53-73

[3] Cvrković T, Jović J, Mitrović M, Krstić O, Toševski I. Experimental and molecular evidence of *Reptalus panzeri* as a natural vector of bois noir. Plant Pathology. 2013;**63**:42-53. DOI: 10.1111/ppa.12080

[4] Austin M. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling. 2007;**200**:1-19

[5] Seemüller E, Schneider B. Taxonomic description of '*Candidatus* Phytoplasma mali' sp. nov., '*Candidatus* Phytoplasma pyri' sp. nov. and '*Candidatus* Phytoplasma prunorum' sp. nov., the causal agents of apple proliferation, pear decline and European stone fruit yellows, respectively. International Journal of Systematic and Evolutionary Microbiology. 2004;**54**:1231-1240

[6] Mittelberger C, Obkircher L, Oettl S, Oppedisano T, Pedrazzoli F, Panassiti B, Kerschbamer C, Anfora G, Janik K. The insect vector *Cacopsylla picta* vertically transmits the bacterium '*Candidatus* Phytoplasma mali' to its progeny. Plant Pathology. 2016;**66**:1015-1021. DOI: 10.1111/ppa.12653

[7] Horton DR. Monitoring of pear psylla for pest management decisions and research. Integrated Pest Management Reviews. 1999;**4**:1-20

[8] Muther J, Vogt H. Sampling methods in orchard trials: A comparison between beating and inventory sampling. IOBC WPRS Bulletin. 2003;**26**:67-72

[9] Ossiannilsson F. The Psylloidea (Homoptera) of Fennoscandia and Demark. Leiden. New York: E.J. Brill; 1992

[10] Unterthurner M, Baric S. Sechs Jahre Erfahrungen in einer Modellanlage. Obst- und Weinbau. 2011;**3**:77-78

[11] Austin MP. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. Ecological Modelling. 2002;**157**:101-118

[12] Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Cambridge; New York: Cambridge University Press; 2007

[13] Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, White JS. Generalized linear mixed models: A practical guide for ecology and evolution. Trends in Ecology & Evolution. 2009;**24**:127-135

[14] Park T, Casella G. The Bayesian lasso. Journal of the American Statistical Association. 2008;**103**:681-686

[15] MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA. Estimating site occupancy rates when detection probabilities are less than 1. Ecology. 2002;**83**:2248-2255

[16] MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. Occupancy Estimation and Modelling. Inferring Patterns and Dynamics of Species Occurrence. Boston: Elsevier; 2006

[17] Panassiti B, Hartig F, Breuer M, Biedermann R. Bayesian inference of environmental and biotic factors determining the occurrence of the grapevine disease 'bois noir'. Ecosphere. 2015;**6**:art143, 1-13. DOI: 10.1890/ES14-00439.1

[18] Parry M, Gibson GJ, Parnell S, Gottwald TR, Irey MS, Gast TC, Gilligan CA. Bayesian inference for an emerging arboreal epidemic in the presence of control. Proceedings of the National Academy of Sciences. 2014;**111**:6258-6262. DOI: 10.1073/pnas.1310997111

[19] Thébaud Gl SN, Chadœuf Jl DA, Gr L. Identifying risk factors for european stone fruit yellows from a survey. Phytopathology. 2006;**96**:890-899

[20] Baric S, Kerschbamer C, Dalla VJ. Detection of latent apple proliferation infection in two differently aged apple orchards in South Tyrol (northern Italy). Bulletin of Insectology. 2007;**60**:265-266

[21] Hoffman MD, Gelman A. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning research. 2014;**15**:1593-1623

[22] Stan Development Team. Stan modeling language users guide and reference manual; 2017

[23] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. London: Chapman & Hall; 2014

[24] Hartig F. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models; 2016

[25] Development Core Team R. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2011

[26] Baric S, Öttl S, Dalla Via J. Infection rates of natural psyllid populations with '*Candidatus* Phytoplasma mali' in South Tyrol (Northern Italy). In: 21st International Conference on Virus and other Graft Transmissible Diseases of Fruit Crops. Neustadt, Germany: Julius Kühn-Institut; 2009. pp. 189-192

[27] Baric S. Molecular tools applied to the advancement of fruit growing in South Tyrol: A review. Erwerbs-Obstbau. 2012;**54**:125-135. DOI: 10.1007/s10341-012-0170-y

[28] Baric S, Kerschbamer C, Vigl J, Via JD. Translocation of apple proliferation phytoplasma via natural root grafts – A case study. European Journal of Plant Pathology. 2007;**12**:207-211

[29] Wikle CK. Hierarchical bayesian models for predicting the spread of ecological processes. Ecology. 2003;**84**:1382-1394. DOI: 10.1890/0012-9658(2003)084[1382,HBMFPT]2.0.CO;2

[30] Clark JS. Why environmental scientists are becoming Bayesians. Ecology Letters. 2005;**8**:2-14. DOI: 10.1111/j.1461-0248.2004.00702.x

[31] Zuur AF, Saveliev AA, Ieno EN. Zero Inflated Models and Generalized Linear Mixed Models with R. Newburgh: Highland Statistics Ltd.; 2012

[32] Dunson DB. Commentary: Practical advantages of Bayesian analysis of epidemiologic data. American Journal of Epidemiology. 2001;**153**:1222-1226

IntechOpen