Non-stationary spatial model for the distribution of ***Xylella fastidiosa*** in Alicante.

Martina Cendoya[1], Ana Hubel[12], Antonio Vicent[1], David Conesa[2]

[1] Centre de Protecció Vegetal i Biotecnología, Institut Valencià d'Investigacions Agràries (IVIA). Valencia, Spain.
[2] Departament d'Estadística i Investigació Operativa, Universitat de València. Valencia, Spain.

E-mail for correspondence: `cendoya_mar@externos.gva.es`

Abstract: Describing the effect of climatic and spatial factors on the geographic distribution of the plant pathogenic bacterium *Xylella fastidiosa* has been the main aim since the moment that it was discovered its presence in Alicante (Spain). This work started with the analysis of the presence/absence data of the pathogen using Bayesian hierarchical models through the integrated nested Laplace approximation methodology and the stochastic partial differential equation approach. Spatial models usually assume stationarity, however, this may be not applicable when physical barriers are present in the study area. Taking into account the irregularities of the terrain and what this may entail in the spread of the disease, higher altitude areas have been considered as possible barriers in the area of interest. The results show that the spatial effect had a strong effect in the model and also that there was no great influence of the barriers due to their reduced extension. Future work will be focused in using these barriers models with theoretical phytosanitary barriers.

Keywords: *Xylella fastidiosa*; INLA; SPDE; Barriers.

1 Introduction

Species distribution models (SDMs) are useful tools to establish which conditions are potentially suitable for the expansion of populations, to evaluate the associations of biotic and abiotic factors with the geographic extent of the species, as well as to predict the species distribution in space and time. These types of models can be developed through different methodologies. However, in many cases, they ignore the spatial dependence which usually

exists among the geographical locations of the observations. This can lead to an overestimation of the parameters and to establish erroneus relationships between observations and covariates. Spatial Bayesian hierarchical models allow the inclusion of spatial autocorrelation.
Spatial models are usually based on the fact that the spatial correlation between observations only depends on the Euclidean distance between locations, i.e. that assumes that is stationary and isotropic. Nevertheless, this assumption may lead to a bias in the prediction of species distribution when there are dispersal barriers in the study area (Bakka et al., 2019, Martínez-Minaya et al., 2019).
Xylella fastidiosa was detected in 2017 in Alicante (Spain), affecting mainly almond trees, although it has also been detected in other plant species. The first interest in this study was to analyze the effects of climatic and spatial factors on the distribution of the pathogen. But taking into account that the study region had a variable topography, with areas at sea level and mountains over than 1500 m of altitude, the areas with the highest altitude were considered as physical barriers.

2 Data and modeling

Data were considered as continuous locations that occur within a defined spatial domain (geostatistical data). Presence/absence data of *X. fastidiosa* from the official surveys in 2018 in Alicante were analyzed using a Bayesian hierarchical model through the Integrated Nested Laplace Approximation methodology (Rue et al., 2009). The spatial effect was included using the Matérn covariance function, approximated as a solution to a stochastic partial differential equation (Lindgren et al., 2011).
The mean of the response variable Y_i was linked to a structural predictor which included the effect of covariates and spatial effect in an additive way:

$$g(\pi_i) = \beta_0 + \sum_{m=1}^{M} \beta_m x_{mi} + u(s_i),$$

where β_0 is the intercept, β_m are the coefficients of the covariates x_m, π_i is the probability of presence at location i, and u_i represents the spatial random effect.
Following Bakka et al. (2019), taking into account a non-stationary process, in the areas with barriers the correlation was eliminated by introducing a different Matérn field, with the same variance (σ) but a range (r) close to zero. Thus, $u(\boldsymbol{s})$ is the solution to a system of stochastic differential equations that includes the normal area with the area of the barriers. In this case, the areas with highest altitude were established as barriers (above 1065 m).
Climatic variables for Alicante were obtained from the WorldClim v.2 database (Fick and Hijmans, 2017). Due to the high linear correlation found

among the climatic variables, a selection of variables was made prior to the modeling process, where the collinearity was evaluated by means of the variance inflation factor (VIF). Once the climatic variables to be included in the model were pre-selected and taking into account the spatial effect, a model selection was made based on two criteria: the Watanabe Akaike information criterion (WAIC) (Watanabe, 2010), which indicates the goodness of fit of the model; and the logarithmic conditional predictive ordinate (LCPO) (Roos and Held, 2011), which evaluates the predictive capacity.

3 Results and discussion

Based on the linear correlation and the value of the VIF, the pre-selected climatic variables were: mean diurnal range (*bio2*), mean temperature of wettest quarter (*bio8*) and precipitation of wettest month (*bio13*). The combination of these three climatic covariates and the spatial effect resulted in 16 models to evaluate. According to WAIC and LCPO criteria, the one that included the covariate *bio13* and the spatial effect was selected as the best model.

The probability that the posterior distribution of the parameter for *bio13* was less than zero was 0.94, therefore, it was considered relevant in the model. The effect of this covariate on the model would imply that areas with higher precipitation in the wettest month would have lower probability of the presence of *X. fastidiosa*.

Figure 1 shows the mean and standard deviation of the predictive posterior distribution. Athough the covariate *bio13* was considered relevant in the model, a strong influence of the spatial component in the model was observed. In this way, the highest probability of the presence of *X. fastidiosa* was found in those areas where the spatial effect had higher values.

Due to the small extent of the barriers considered in the study area of *X. fastidiosa*, they did not have a major impact on the spatial component, nevertheless, it was observed a smoothing effect around the areas with higher altitude. In the study of species distributions, the elements that are barriers for dispersal cannot be ignored, since it would be wrongly assumed that the species can be found in areas where it would be actually impossible to be present.

Acknowledgments: We thank Generalitat Valenciana for providing the survey data. The present work has received funding from the European Unions Horizon 2020 research and innovation programme under Grant Agreement No. 727987 - XF-ACTORS; from Project E-RTA 2017-00004-C06-01 FEDER INIA-AEI Ministerio de Ciencia, Innovacin y Universidades and Organizacin Interprofesional del Aceite de Oliva Espaol, Spain; and finally from grants MTM2016-77501-P TEC2016-81900-REDT from the

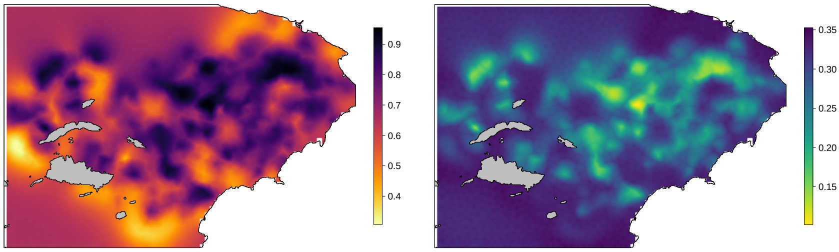

FIGURE 1. Mean (left) and standard deviation (right) of the posterior predictive distribution of the probability of presence of *Xylella fastidiosa*.

Spanish Ministry of Science, Innovation and Universities State Research Agency. MC held an IVIA grant partially funded by the European Social Fund Comunitat Valenciana 2014-2020.

References

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D., and Rue, H. (2019). Non-stationary Gaussian models with physical barriers. *Spatial Statistics*, **29**, 268 – 288.

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302 – 4315.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**, 423 – 498.

Martínez-Minaya, J., Conesa, D., Bakka, H., and Pennino, M. G. (2019). Dealing with physical barriers in bottlenose dolphin (Tursiops truncatus)distribution. *Ecological Modelling*, **406**, 44 – 49.

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

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**, 319 – 392.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, **11**, 3571 – 3594.