International Journal of Computer Applications (0975 - 8887)
Volume 174 - No.16, January 2021

A Spatio-temporal Stochastic Model for an Emerging Plant Disease Spread in a Heterogeneous Landscape

Clara Landry
CIRAD, UMR BGPI, F-97130 Capesterre-Belle-Eau,
Guadeloupe, France

Catherine Abadie
CIRAD, UMR BGPI, F-97130 Capesterre-Belle-Eau,
BGPI, Univ Montpellier, CIRAD, INRAE, Institut Agro,
Montpellier, France

François Bonnot
BGPI, Univ Montpellier, CIRAD, INRAE, Institut Agro,
Montpellier, France

Jean Vaillant
Université des Antilles, LAMIA,
F-97159 Campus de Fouillole, Guadeloupe

ABSTRACT

Black Leaf Streak Disease (BLSD) is the most severe foliar disease of banana and plantain. BLSD is caused by *Pseudocercospora fijiensis* , an ascomycete fungus which produces wind-borne spores responsible for its spatial dispersal. In order to evaluate the BLSD long-distance dispersal and to better understand the effect of environmental factors on its invasive spatial spread, a spatiotemporal study was set up during the recent BLSD invasion in the Martinique island (FWI). Disease detection was carried out from September 2010 to May 2012 and sampling squares were defined from a regular spatial grid built over the island. In this paper, we consider a stochastic model of spatio-temporal propagation of BLSD in an heterogeneous landscape and we present mathematical and computational results for this continuous-time model. Statistical inference of parameters is carried out from presence-absence data using a Bayesian framework based on a data augmentation method with respect to square first colonization times. Parameter posterior distribution calculations made possible the evaluation of the BLSD long-distance dispersal and land-cover influence on the disease propagation. Our results enabled the reenactment of the invasion.

General Terms

Sampling, biological invasion

Keywords

Stochastic process, Likelihood, MCMC, Bayesian inference, Data augmentation, BLSD

1. INTRODUCTION

Understanding biological pathogen invasion is a major contemporary issue: the occurrence of new diseases causes serious disruption of existing ecosystems, biodiversity and local economy [15, 2]. The evaluation of the spatiotemporal spread of a disease is necessary for effective and successful control strategies and requires a good knowledge on the long distance spread of the pathogen and on the effects of environmental factors on this dispersion [4, 12]. Interacting factors can be incorporated into the transmission model [3, 40, 41].

An epidemiological model for transmission in heterogeneous landscapes and Bayesian Markov chain Monte Carlo (MCMC) inference were used by [14] to estimate dispersal and life-cycle parameters. [26] presented a Bayesian Markov model for investigating environmental spread processes. Several authors used models based on spatial or spatio-temporal stochastic processes : for seed dispersal [25], spatio-temporal invasion of alien species [9], disease spread [8, 42]. One can refer to [11]for the theory of point processes and [10] for statistical analysis of spatio-temporal data. [22] showed the potential for integrating stochastic simulations into a framework for statistical modelling. [17, 18, 19, 9] discussed about bayesian methods for fitting model without missing or censored data and [32, 13] discussed the case with missing or censored data. However such missing data can be included as extra parameters in a Bayesian framework. [29] discussed about the use of data-augmented MCMC and presented a simulation-based Bayesian inference for epidemic models illustrated with data from an outbreak of smallpox. [30] addressed some shortcomings of partially observed diseases by means of a reconstruction framework. [31] developed a scheme for robust inference about transmission chains in the context of the Foot-and-Mouth Disease Virus. The framework presented leads to a Bayesian inference scheme able to reconstruct most likely transmission patterns and infection dates. [23] presented an auto logistic model with covariates for spatially correlated binary observations on a lattice but the time aspect was not taken into account. They adopted a Bayesian set-up with incomplete observations from sampled responses taken over the area of interest. Their aim was to improve predictions of probability of presence for such lattice binary data. [43] demonstrated how methods of MCMC based on a hybrid sampler combining Gibbs sampling within Metropolis-Hastings frameworks can be used to derive estimates of parameters and missing data for ecological process models. [7] used a continuous time discrete state space Markov process with spatially varying covariates to model colonization-dispersal of *Heracleum mantegazzianum*, an invasive alien weed. They developed a Bayesian approach to parameter inference taking into account uncertainty in colonization times.

Understanding invasive species dynamics is a real challenge [6]. To the best of our knowledge, there are few studies reporting the evaluation of the long distance dispersal of invasive species and the effects of environmental factors on invasion dynamics. Thus, our objective is twofold. We first focus on modeling the invasion process of Black Leaf Streak Disease (BLSD), a major foliar disease of bananas, in an heterogeneous landscape island scale. Then, we tackle the problem of performing statistical inference on this model with incomplete data in order to better understand the spatiotemporal spread of BLSD: evaluating the long distance dispersal and the effect of environmental factors on the dispersion during the invasive phase.
In this paper, we first introduce a spatiotemporal model based on a stochastic intensity taking into account past neighboring infections and environmental variables in section 2. Then, we present the data augmentation technique we developed within a Bayesian framework. In section 3, we present the presence-absence data available at the Martinique island scale over 83 survey weeks and we apply the method developed to the BLSD survey data to assess the long distance dispersal and the impact of land-cover type on the spatiotemporal dynamic and to propose a reenactment of the BLSD invasion.

2. MODEL AND INFERENCE

2.1 Model

We consider a spatiotemporal stochastic process of colonization which is assumed to be governed by its conditional intensity process. We denote by X the whole set of squares defined from a regular spatial grid built over the island, and by $\mathcal{F}_{t^-}$ the colonization history prior to time t so that the conditional intensity at any square i of X and at time t in R_+ conditional on $\mathcal{F}_{t^-}$ is defined as follows (e.g., [11]) :

$$\lambda_i(t) = \lim_{h \longmapsto 0^+} \frac{Pr(\text{square } i \text{ is colonized in } [t, t+h] \mid \mathcal{F}_{t^-})}{h} \qquad (1)$$

At time t, a susceptible square i receives spores transmitted from squares that are already infected before t. We assume that this square i has a suitability S_i associated with the rainfall level and the habitat receptivity to the infection ([20]). The effect of rainfall level is modeled by an exponential term $\exp(\gamma \mathcal{P}_i)$ where $\mathcal{P}_i$ is the mean rainfall level in square i, and $\gamma \in (-\infty, \infty)$ is a parameter to be estimated : positive (negative) values of this parameter implies that locations with high rainfall levels are (not) preferred by the pathogen. The effect of habitat receptivity is modeled by a term $\sum_{k=0}^{5} r_k.\mathcal{H}_{i,k}$ where $\mathcal{H}_{i,k}$ is the proportion of square i belonging to landscape category k, $k \in \{0, ..., 5\}$ and $r_k \in [0, \infty)$ is a parameter representing the receptivity for colonization of landscape category k. We set $r_0 = 0$ to reflect that the volcano/sea (habitat 0) is uncolonizable due to the absence of host, and $r_1, ..., r_5$ representing respectively five different landscape compositions.
Let $d_{i,j}$ be the distance between squares i and j. If j is infected before t and i susceptible at time t, the rate of transmission from square j to square i is assumed to be equal to $S_i f_\beta(d_{i,j})$ where $S_i = exp(\gamma \mathcal{P}_i) \sum_{k=0}^{5} r_k \mathcal{H}_{i,k}$ and $f_\beta(d_{i,j}) = \beta^2 d_{i,j}^{-2\beta}$ stands for the isotropic power-law dispersal kernel with decay parameter β, [7].

Therefore, the stochastic intensity for a susceptible square i at time t conditionally to the history of the process prior to time t is :

$$\lambda_i(t) = \exp(\gamma \mathcal{P}_i) \sum_{k=0}^{5} (r_k \mathcal{H}_{i,k}) \sum_{j \in C_{t^-}} f_\beta(d_{i,j}) \qquad (2)$$

where C_{t^-} is the set of colonized squares before time t.
It is worth noticing that if square i is colonized at time t, then

$$\lambda_i(t) = 0.$$

2.2 Likelihood

Denoting by θ the parameter vector, we have $\theta = (\beta, \gamma, r_1, \cdots, r_5)$. We face the problem of inferring θ from incomplete data consisting of intervals in which first dates of colonization are detected or not for squares included in the survey over the time interval $[0, T]$. If we denote by τ_i the unobserved colonization date for square i, then the complete data set is

$$\tau = \{\min(\tau_i, T), i \in X\}$$

and the likelihood function of the complete data set is (see for example [10], p 349)

$$L(\theta; \tau) = \big(\prod_{\substack{i \in X \\ \tau_i \leq T}} \lambda_i(\tau_i) \big) \times \exp\Big(-\sum_{i \in X} \int_0^T \lambda_i(t) dt \Big) \qquad (3)$$

From (2) and (3), we can write $L(\theta; \tau) =$

$$\prod_{i \in C_T} \left[\left(S_i \sum_{j \in C_{\tau_i^-}} f_\beta(d_{i,j}) \right) \times \exp\left\{ -S_i \sum_{j \in C_{\tau_i^-}} f_\beta(d_{i,j})(\tau_i - \tau_j) \right\} \right]$$

$$\times \prod_{i \notin C_T} \exp\left\{ -S_i \sum_{j \in C_{T^-}} f_\beta(d_{i,j})(T - \tau_j) \right\} \qquad (4)$$

where $S_i = \exp(\gamma \mathcal{P}_i) \sum_{k=0}^{5} (r_k \mathcal{H}_{i,k})$ is the suitability for colonization in square i.
The first product on the right-hand side of (4) is the contribution from the squares colonized before the last observation date T whereas the second one corresponds to the probability for the remaining squares not being colonized at this date.
We have to tackle the problem of inferring θ from incomplete data consisting of intervals in which dates of colonization occur. The complete data set τ is not available and the incomplete data likelihood involves integrals which are analytically and numerically intractable (6). If we denote by $\mathcal{D}$ the observed incomplete set of data, then the joint posterior distribution for θ and τ is

$$P(\theta, \tau | \mathcal{D}) = \frac{P(\mathcal{D}, \tau | \theta) \pi(\theta)}{P(\mathcal{D})} \qquad (5)$$

where $\pi(.)$ is the prior distribution for θ and $P(\mathcal{D})$ considered as a normalizing constant. $\pi(.)$ reflects the knowledge about the model parameters prior to the observations. Therefore, the posterior distribution for θ is obtained by marginalizing over τ :

$$P(\theta | \mathcal{D}) = \int_H P(\theta, \tau | \mathcal{D}) d\tau \qquad (6)$$

Table 1. Notations used in the likelihood and the MCMC algorithm.

Notation	Definition
τ_i	time of colonization of the i^{th} square
C_t	set of squares colonized up to time t
C_{t-}	set of squares colonized prior to time t
N	number of squares in Martinique
$\lambda_i(t)$	rate of colonization of a square i in time t conditionally to the history of the process prior to time t
θ	$(\beta, \gamma, r_1, ..., r_5)$, the parameter vector
θ^*	infered parameter vector
H	the space of unobserved colonization times
$\pi(.)$	prior distribution of parameter
$q_k(.,.)$	proposal distribution for the kth element of θ
$\mathcal{E}(a)$	exponential distribution with parameter a

where H is the space of unobserved colonization times.
In next section, we focus on MCMC techniques [37] for sampling from the posterior $P(\theta, \tau|\mathcal{D})$ and then from $P(\theta|\mathcal{D})$. For any single parameter, the marginal posterior is then estimated by the sampled values of this parameter only. The notations used are summarized in Table 1.

2.3 Bayesian inference on parameters

By using MCMC techniques and considering the unknown colonization times as augmented variables in the MCMC algorithm, we can sample from the posterior distribution of θ. We generalize the sampling method proposed by [7, 5] by updating the whole set of first colonization times at each iteration by means of a reversible-jump hybrid sampler. This updating method reconstructs realizations of the colonization times consistent with the data and the model. Each iteration of this algorithm includes also updating of all model parameters. We assume that the prior distributions for the parameters are independent and uniform over a fixed interval. To sample from the posterior distribution $P(\theta, \tau|\mathcal{D})$, the algorithm starts with an initial vector $\tau^{(0)}$ of colonization times consistent with the observation $\mathcal{D}$. Then, at the mth step, the vector $\tau^{(m+1)}$ of colonization times consistent with $\mathcal{D}$ is obtained from a sampler conditional on $\tau^{(m)}$ and $\theta^{(m)}$ the parameter vector sampled at step m. At step m, our proposal samples a vector $\tau^{(m+1)}$ according to a truncated multivariate Gaussian law centered on $\tau^{(m)}$ with an adaptive variance matrix. The algorithm is as follows:

(1) Start with $\tau^{(0)}$ consistent with the data and iterate the following procedure:

(2) Assign values to $\theta^{(0)}$;

(3) Set $m = 0$;

(4) Repeat
   - —Draw $\tau^{(m+1)}$ consistent with $\mathcal{D}$ using the data augmentation sampler with $\tau^{(m)}$ and $\theta^{(m)}$;
   - —Sample $\theta^{(m+1)}$ using Metropolis-Hastings with $\tau^{(m+1)}$ and $\theta^{(m)}$;
   - —Set $m = m + 1$;
   - —Store every $\tau^{(m)}$ and $\theta^{(m)}$; after an initial burn-in period;

(5) End repeat.

For $\tau^{(0)}$, each unknown colonization time is drawn according to the uniform distribution on a interval consistent with $\mathcal{D}$.

$$\theta^{(m+1)} = (\beta^{(m+1)}, \gamma^{(m+1)}, r_1^{(m+1)}, r_2^{(m+1)}, r_3^{(m+1)}, r_4^{(m+1)}, r_5^{(m+1)})$$

and the kth element $\theta_k^{(m+1)}$ of $\theta^{(m+1)}$ is drawn according to the acceptance probability:

$$\min\left(1, \frac{P\Big((\theta_1^{(m+1)}, \cdots, \theta_{k-1}^{(m+1)}, \theta_k^{\star}, \theta_{k+1}^{(m)}, \cdots \theta_7^{(m)}); \tau^{(m+1)}\Big) q_k(\theta_k^{\star}, \theta_k^{(m)})}{P\Big((\theta_1^{(m+1)}, \cdots, \theta_{k-1}^{(m+1)}, \theta_k^{(m)}, \theta_{k+1}^{(m)}, \cdots \theta_7^{(m)}); \tau^{(m+1)}\Big) q_k(\theta_k^{(m)}, \theta_k^{\star})}\right) \tag{7}$$

where q_k is the proposal for θ_k.

2.4 Method evaluation using artificial data

The stochastic model presented in section 2.1 is easily simulated as a time continuous Markov process when the covariates vary spatially only. We use the fact that the waiting time between two consecutive events follows, conditionally to the past history, an exponential distribution with parameter depending on the set of squares colonized prior to the last event. Using a spatio-temporal point process with conditional intensity $\lambda_x(t)$ for square x in X at time t given by equation 2, the waiting time of the next infection is distributed according to the exponential distribution $\mathcal{E}(a)$ with $a = \sum_{x \in X} \lambda_x(t)$. It is worth noticing that $\lambda_x(t) = 0$ if x is already infected at time t.
The simulation states in two steps, the first step consists in the calculus of the waiting time of the next infection ϵ :

$$\epsilon = \inf(\tau_x) \sim \mathcal{E}(\sum_{x \in X} \lambda_x(t)) \tag{8}$$

The second step is to establish the square position concerned by the future infection in time $(t + \epsilon)$. The square j is retained with probability p_j such that :

$$p_j = \frac{\lambda_j(t)}{\sum_{j \in X} \lambda_j(t)}. \tag{9}$$

The evaluation of the MCMC algorithm was carried out using independent artificial data sets on 500 squares. The posterior densities of the 7 parameters are presented in Fig 1.
The Gelman-Rubin criterion [16] was used to confirm the chain convergence and set at the value 1.

3. APPLICATION: REGIONAL SCALE DYNAMICS OF BLSD

The Black Leaf Streak Disease (BLSD) or Black Sigatoka of bananas, is caused by the devastating ascomycete fungal plant pathogen *Pseudocercospora fijiensis* . It is one of the most important food security threat at global scale [34] because of its massive damage to banana crops and its rapid worldwide expansion [38]. In the Caribbean Islands, the first detection was in Cuba in 1990 and one of the most recent introduction was in Martinique (French West Indies). The first report of BLSD on Martinique island in September 2010 indicates that the disease was introduced passively by spores, from nearby islands where BLSD was yet present, during August 2010 [24]. Now some genetic studies suggest that the disease was introduced through two successive and independent modes (natural and anthropogenic) ([20]). Indeed, a surveillance network was set up since 2009 and allowed to check monthly the banana production with respect to BLSD.

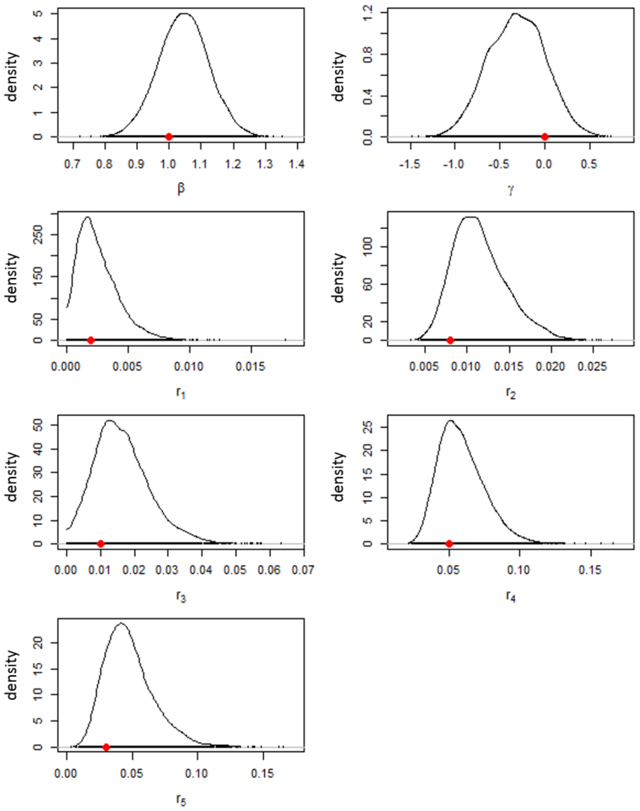

Fig. 1. Bayesian posterior densities for each of the seven model parameters (β, γ, r_1, r_2, r_3, r_4, r_5) obtained from 10000 runs of the algorithm using artificial data. The red dot indicates the real parameter value used in the data simulation. We used a burn-in period of 1000 iterations.

BLSD is specific to Musa hosts (bananas and plantain). Banana susceptibility to BLSD and plantation density are also important in the invasion process [20, 21]. The disease causes many lesions on leaves which become large necrotic surface reducing the photosynthetic surface of the plant. On susceptible cultivars, BLSD may destroy (by necrosis) 40% to 100% of total leaf surface [27] which causes important yield losses, from 20 to 80% in the absence of fungicide treatments [20].

Pseudocercospora fijiensis displays both an asexual and sexual reproduction emitting respectively and successively two types of spores (conidia and ascospores). Both spores play an important part in the fungus dispersal, but only ascospores are considered as responsible for the long-distance dispersal [20]. The long dispersal distance of *Pseudocercospora fijiensis* spores was studied at different spatial scales. [36] proposed the estimation of an anisotropic dispersal function using a 2 km square experiment zone and observed the disease symptoms across the experiment (1 km from the infection source). At a larger spatial scale, [1] detected the presence of the pathogen 6 km away from the infectious source. The velocity of an epidemic invasion depends on climatic conditions and landscape composition [33]. Indeed, rainfall and wind are the two natural dispersal modes of BLSD.

3.1 Data

Presence-absence data for *Pseudocercospora fijiensis* were collected in Martinique island within the framework of a disease surveillance network on banana over the whole island coordinated by the DAAF (Direction de l'Alimentation, de l'Agriculture et de la Forêt) and the FREDON Martinique (Fédération REgionale de Défense contre les Organismes Nuisibles de la Martinique). Martinique island surface area is about 1228 km^2. A regular grid with cell size of 1 km x 1 km was placed over Martinique providing 1236 observation units. As mentioned in paragraph 2.1, we refer to such an observation unit as a square. Presence-absence observations were carried out (Fig. 2) within different time windows (time between two consecutive observations of the same square) and surrounding environmental variables were measured. After the detection of BLSD on banana tree in a square, no more observations were carried out on it since once a square is colonized by the pathogen it remains infected.

Detection rates were spatially and temporally heterogeneous : the time window length was comprised between 1 and 82 weeks with an average of 21 weeks, in some squares only the date of the first detection was available, and in other squares only the last date of non-detection was available. The whole observation period was over 83 weeks (from September 2010 to May 2012). Fig. 3 shows the spatial locations of the fungus at three different dates: beginning of the study with 37 squares contaminated, middle (week 36) and end (week 83) of the study.

Five land-cover categories were defined : isolated banana from partially resistant varieties, private garden with few plants of partially resistant varieties, creole garden (small-scale farming with banana and other productions such as vegetables, fruits), Cavendish plots specialized in production for export on a larger scale and plantain plots possibly extending over several hectares. Banana cultivated in plots are very susceptible to BLSD and disease controls are used to maintain good production. On Cavendish plots fungicide chemical treatments are preferably used whereas on plantain plots manual control, such as defoliating necrotic leaves, are privileged [35]. The disease evaluation of each visual positive detection on a banana tree was confirmed with laboratory diagnostic. As lesions were analyzed on infected leaf samples using molecular markers of high sensitivity levels, there are no false-positive results.

Rainfall data were collected by Météo France on 19 weather stations located from north to south of the island. The average weekly rainfall was calculated over the 83 weeks and used as a covariate.

As environmental variables such as land-cover and humidity induced by rainfall are assumed to affect the spreading process, each of the 1236 squares covering Martinique was associated with a vector of covariates : the proportions of each landscape type and the rainfall level. The proportions of each landscape type were calculated from geo-referenced administrative surveys on Cavendish and plantain plots made from years 2010 to 2012 and expert opinions regarding the distribution of urban and rural zones of the island.

3.2 Bayesian analysis of the real data

The prior distributions chosen for the model parameters reflect the lack of information about these parameters : $\beta \sim U[0,10]$, $\gamma \sim U[-10,10]$ and $r_k \sim U[0,10]$ for each k in $\{1,\cdots,7\}$.

We used Normal distributions centered on the current parameter for each proposal distribution. After tuning, the standard deviation used for $\beta,\gamma,r_1,r_2,r_3,r_4$ and r_5 are respectively $0.01, 0.5, 0.001, 0.01, 0.01, 0.01$ and 0.01. Fig. 4 gives the posterior distributions of the land-cover parameters and γ, whereas the

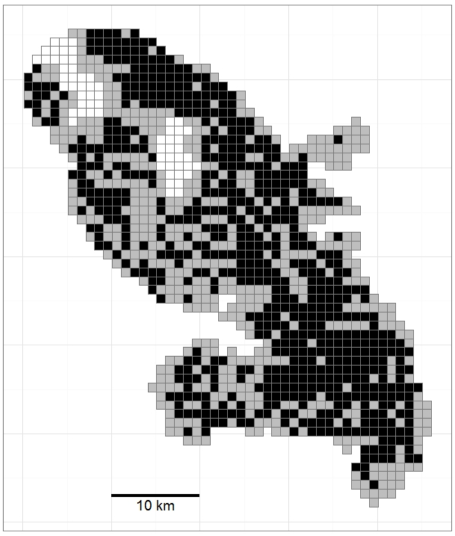

Fig. 2. Set of squares considered over the 83 weeks of survey. Black squares correspond to squares with observations, grey squares are non-observed squares and white squares are uncolonizable (high mountain and volcano).

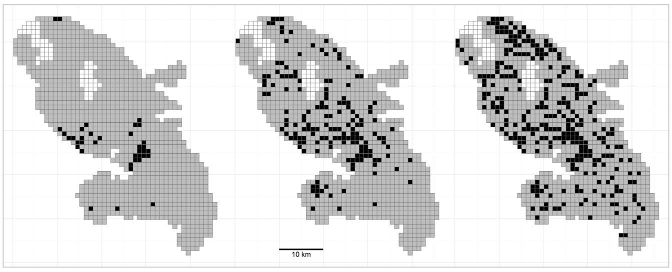

Fig. 3. Presence-absence maps for *Pseudocercospora fijiensis* (left to right), at the beginning of the survey (week 1), at the middle of the survey (week 36), and at the end of the survey (week 83). Black cells are squares with disease detection, grey squares are non informed or non infected squares and white squares are uncolonizable (high mountain and volcano).

posterior distribution of β is shown in Fig. 5. The posterior means and standard errors are given in table 2. From these parameter estimates, several numerical experiments were carried out.

Table 2. Bayesian estimates of the parameter vector $\theta = (\beta, \gamma, r_1, r_2, r_3, r_4, r_5)$.

Parameter	Posterior Mean	Posterior standard error	Quantile 2.5%	Quantile 97.5%
β	1.991	0.655	1.069	1.325
γ	-2.124	1.454	-4.574	0.861
r_1	5.58e-03	2.98e-03	6.07e-04	0.012
r_2	1.303e-02	3.684e-03	6.383e-03	0.021
r_3	5.746e-03	4.132e-03	2.815e-04	0.015
r_4	4.418e-03	1.292e-03	2.25e-03	0.0072
r_5	2.473e-03	1.793e-03	1.77e-04	0.0069

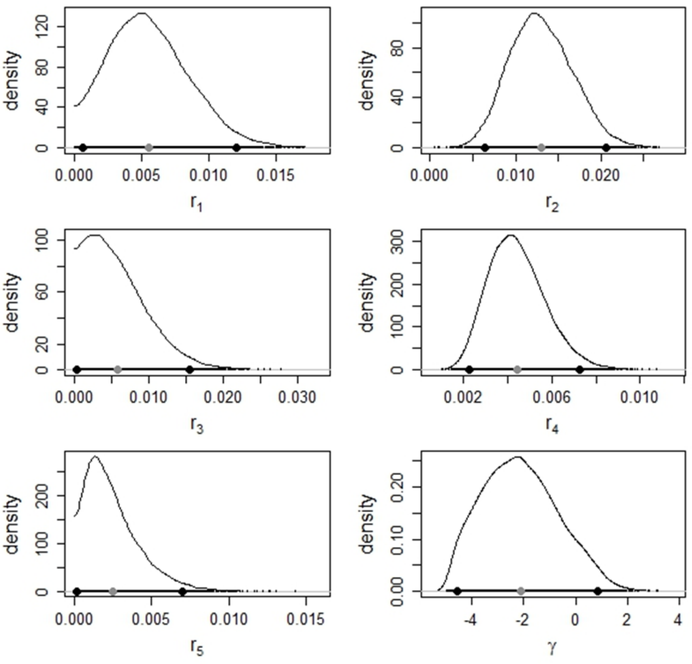

Fig. 4. Bayesian posterior density of the rainfall parameter γ, along with the ones for the land-cover parameters r_i representing the receptivity for colonization of landscape category i ranging from 1 to 5: isolated banana, private garden, creole garden, Cavendish plots and plantain plots. Grey points correspond to the posterior mean value, black points correspond (left to right) to the 2.5% and 97.5% quantiles.

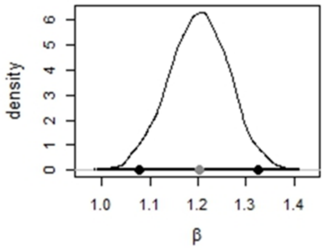

Fig. 5. Bayesian posterior density of β, the decay parameter of the isotropic power-law dispersal kernel. Grey points correspond to the posterior mean value, black points correspond (left to right) to the 2.5% and 97.5% quantiles.

3.3 Spatiotemporal prediction

New infected squares appear to be clustered around previous ones due to the pathogen spread characteristics as shown in Fig 3. The invasion process seems to be related to the landscape structure : the southern part of the island is less infected by the disease, probably because of the dry vegetation of this region whereas the northern part is mostly covered by the tropical rain-forest. Based on the proposed model, the simulated predictions of spatio-temporal spread of infections provide very similar results to that observed, as in Fig. 6 which shows artificial data generated from the parameter Bayesian estimates.

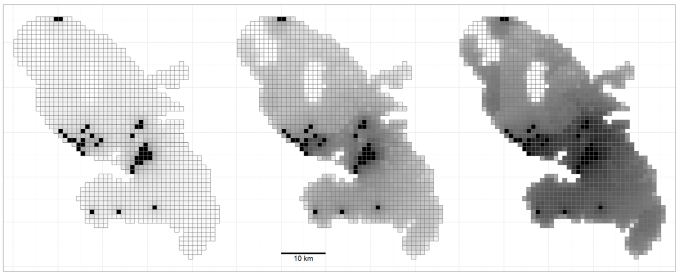

Fig. 6. Probability of colonization map (left to right), at the beginning of the survey (week 10), at the middle of the survey (week 36) and at the end of the survey (week 83), obtained from 10000 model simulations using the estimated posterior density of the parameter vector. The darker the square, the higher the probability

4. DISCUSSION

In this study, we set out to model the invasion process of an emerging plant disease in a heterogeneous landscape. To achieve this goal, we had to address the problem of performing statistical inference on the model with incomplete data. This approach was applied to the incidence data obtained during the invasion of Black Leaf Streak Disease of bananas in Martinique.

We modeled successfully the BLSD invasion through a spatiotemproal stochastic process. The island landscape was divided in squares of 1 km x 1 km, which were characterized by two environmental factors (rainfall level and land-cover type).

A first sensitivity analysis was realized using variance-based method, the extended Fourier Amplitude Sensitivity Test (eFAST) [39]. It shows that the land-cover is less influential than the rainfall level and dispersal parameters. The sensitivity analysis had to be performed on averaged outputs because of the stochasticity of the model. [28] proposed the use of meta-models in this context to obtain valuable indexes.

We used the Bayesian paradigm for fitting a spatio-temporal stochastic process to square data with unobserved first colonization times and subject to censoring. Bayesian statistical inference could be performed using a data augmentation method which reconstructs the first colonization time for each square at each iteration of MCMC algorithm. This data reconstruction method is powerful and original in the epidemiology science. Whereas some short-distance function were available for BLSD [36], the inference of the dispersal parameter allowed to get for the first time a long-distance function for this disease. This function would be useful to develop spatial epidemiological model to evaluate control methods at landscape scale (such as varieties mixtures).

Numerical simulations can be performed to evaluate where future colonization events are likely to occur using the Bayesian parameter estimates obtained from our inferential routine. We could simulate square colonization beyond the end of our survey data and for unobserved squares. These predictive data generation can be useful in a control strategy framework. Numerical simulations can be also carried out in case where the covariates are time dependent but this is more computer-time consuming.

Acknowledgements

We would like to thank Elea Ksiazek, Sophie Veran, David Pleydell and Virginie Ravigné for scientific discussions about the study. We thank also DAAF and FREDON Martinique for conducting the experiments over all the island on 83 weeks, Météo France for valuable rainfall data, IGN (Institut National de l'information Géographique et forestière) for providing georeferenced maps and Jean Iotti for valuable comments on island landscape description. We are particularly grateful to Raphael Pasquier of University of Antilles (C3I) for his assistance in computing and numerical experiments. Clara Landry PhD studies were supported by a doctoral fellowship from CIRAD, and financially supported by FEDER Guadeloupe through a research project (convention 1/1.4-30305 project/*Gestion des risques en santé animale et végétale*) and a cooperation project through the Interreg IV Caraibes fundings (convention num. 31409/CABARé). This research project is part of the MALIN program funded by the European Union and the Guadeloupe Regional Council under the European Research and Development Funds (ERDF) 2014-2020 program (Grant 2018-FED-1084).

5. REFERENCES

[1] A.F. Amil, S.P. Heaney, C. Stanger, and M. W. Shaw. Dynamics of QoI sensitivity in *Mycosphaerella fijiensis* in Costa Rica during 2000 to 2003. *Phytopathology*, 2007.

[2] Pamela K. Anderson, Andrew A. Cunningham, Nikkita G. Patel, Francisco J. Morales, Paul R. Epstein, and Peter Daszak. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, 19(10):535 – 544, 2004.

[3] D.E. Aylor. Spread of plant disease on a continental scale: role of aerial dispersal of pathogens. *Ecology*, 84(8):1989–1997, 2003.

[4] J.K.J Brown and M.S. Hovmoller. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, 297:537–541, 2002.

[5] Patrick E. Brown, Florencia Chimard, Alexander Remorov, Jeffrey S. Rosenthal, and Xin Wang. Statistical inference and computational efficiency for spatial infectious disease models with plantation data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2013.

[6] M. L. Cain, G.M. Brook, and E. S. Allan. Long-distance seed dispersal in plant populations. *American Journal of Botany*, 87(9):1217–1227, 2000.

[7] Stephen Catterall, Alex R. Cook, Glenn Marion, Adam Butler, and Philip E. Hulme. Accounting for uncertainty in colonisation times: a novel approach to modelling the spatio-temporal dynamics of alien invasions using distribution data. *Ecography*, 35(10):901–911, 2012.

[8] A. Cook, G. Marion, A. Butler, and G.J. Gibson. Bayesian inference for the spatio-temporal invasion of alien species. *Bulletin of Mathematical Biology*, 69:2005–2025, 2007.

[9] A. Cook, W. Otten, G. Marion, G.J. Gibson, and C.A. Gilligan. Estimation of multiple transmission rates for epidemics in heterogeneous population. *PNAS*, 104(51):20392–20397, 2007.

[10] Noel Cressie and Christopher K. Wikle. *Statistics for Spatio-Temporal Data*. Wiley, 2011.

[11] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer, 2nd Ed., New York, 2003.

[12] M.L. Desprez-Loustau, C. Robin, M. Buee, R. Courtecuisse, J. Garbaye, F. Suffert, I. Sache, and Rizzo D.M. The fungal dimension of biological invasions. *Trends in Ecology & Evolution*, 22(9):472–80, 2007.

[13] J.A.N. Filipe, W. Otten, G.J. Gibson, and C.A. Gilligan. Inferring the dynamics of a spatial epidemic from time-series data. *Bulletin of Mathematical Biologyy*, 66:373–391, 2004.

[14] Joao A. N. Filipe, Richard C. Cobb, Ross K. Meentemeyer, Christopher A. Lee, Yana S. Valachovic, Alex R. Cook, David M. Rizzo, and Christopher A. Gilligan. Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: Sudden oak death in northern californian forests. *PLoS Comput Biol*, 8(1):e1002328, January 2012.

[15] Matthew C. Fisher, Daniel. A. Henk, Cheryl J. Briggs, John S. Brownstein, Lawrence C. Madoff, Sarah L. McCraw, and Sarah J. Gurr. Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484(7393):186–194, April 2012.

[16] Andrew Gelman, Donald B Rubin, et al. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[17] G.J. Gibson, A. Kleczkowski, and C.A. Gilligan. Bayesian analysis of biotanical epidemics using stochastic compartmental models. *PNAS*, 101(33):12120–12124, 2004.

[18] G.J. Gibson, W. Otten, J.A.N. Filipe, A. Cook, G. Marion, and C.A. Gilligan. Bayesian estimation for percolation models of disease spread in plant population. *Stat Comput*, 16(51):391–402, 2006.

[19] G.J. Gibson, G. Streftaris, and S. Zachary. Generalised data augmentation and posterior inferences. *Journal of statistical planning and inference*, 141:156–171, 2010.

[20] M. Guzmán, L. Pérez-Vicente, J. Carlier, C. Abadie, L. de Lapeyre de Bellaire, F. Carreel, D.H. Marin, R.A. Romero, F. Gauhl, C. Pasberg-Gauhl, and D.R.. Jones. *Handbook of diseases of banana, abaca and enset*, chapter Black leaf streak. CABI Publishing [New York], 2018.

[21] F. M. Hamelin, F. Castella, V. Doli, Marçais B., V. Ravigné, and M.A. Lewis. Mate finding, sexual spore production, and the spread of fungal plant parasites. *Bulletin of Mathematical Biology*, 78(4):695–712, 2016.

[22] F. Hartig, J.M. Calabrese, B. Reineking, T. Wiegand, and A. Huth. Statistical inference for stochastic simulation models - theory and application. *Ecology Letters*, 14:816–827, 2011.

[23] Jennifer A. Hoeting, Molly Leecaster, and David Bowden. An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics*, 5(1):102–114, 2000.

[24] R Ioos, J. Hubert, C. Abadie, D. Duféal, G. Opdebeeck, and J. Iotti. First report of black sigatoka disease in banana caused by mycosphaerella fijiensis on Martinique island. *Plant Disease*, 95(3):359–359, 2011.

[25] D. J. Levey, J. J. Tewksbury, and B. M. Bolker. Modelling long-distance seed dispersal in heterogeneous landscapes. *Journal of Ecology*, 96:599–608, 2008.

[26] Jed A. Long, Colin Robertson, Farouk S. Nathoo, and Trisalyn A. Nelson. A bayesian space time model for discrete spread processes on a lattice. *Spatial and Spatio-temporal Epidemiology*, 3(2):151–162, 2012.

[27] Douglas H. Marin, Ronald A. Romero, Mauricio Guzman, and Turner B. Sutton. Black sigatoka: An increasing threat to banana cultivation. *Plant Disease*, 87(3):208–222, March 2003.

[28] A. Marrel, B. Ioos, S. Da Veiga, and M. Ribatet. Global sensitivity analysis of stochastic computer models with joint metamodels. *Stat Comput*, 22:833–847, 2012.

[29] Trevelyan J. McKinley, Joshua V. Ross, Rob Deardon, and Alex R. Cook. Simulation-based bayesian inference for epidemic models. *Computational Statistics and Data Analysis*, 71:434 – 447, 2014.

[30] N. Mollentze, L.H. Nel, Sunny Townsend, Kevin le Roux, Katie Hampson, DT Haydon, and Samuel Soubeyrand. A bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. *Proceedings of the Royal Society of London. Series B: Biological Sciences.*, 281(1782), 2014.

[31] Marco J. Morelli, Gaell Thébaud, Joel Chadoeuf, Donald P. King, Daniel T. Haydon, and Samuel Soubeyrand. A bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. *PLoS Comput Biol*, 8(11), 11 2012.

[32] P.D. O'Neill. A tutorial introduction to bayesian inference for stochastic epidemics models using Markov chain mote carlo methods. *Mathematical Biosciences*, 180:103–114, 2002.

[33] S. Parnell, T. R. Gottwald, C. A. Gilligan, N. J. Cunniffe, and F. van den Bosch. The effect of landscape pattern on the optimal eradication zone of an invading epidemic. *Phytopathology*, 100(7):638–644, 2010.

[34] Elizabeth Pennisi. Armed and dangerous. *Science*, 327(5967):804–5, 2010.

[35] Randy C. Ploetz, Gert H.J. Kema, and Li-Jun Ma. Impact of diseases on export and smallholder production of banana. *Annual Review of Phytopathology*, 53(1):269–288, 2015.

[36] A. Rieux, S Soubeyrand, F. Bonnot, E. K. Klein, J. E. Ngando, A. Mehl, V Ravigné, Carl, and L. de Lapeyre de Bellaire. Long-distance wind-dispersal of spores in a fungal plant pathogen: estimation of anisotropic dispersal kernels from an extensive field experiment. *Plos One*, 2014.

[37] C.P. Robert and G Casella. *Monte Carlo Statistical Methods*. second ed. Springer, 2010.

[38] S. Robert, V. Ravigné, M-F. Zapater, C. Abadie, and J. Carlier. Contrasting introduction scenarios among continents in the worldwide invasion of the banana fungal pathogen mycosphaerella fijiensis. *Molecular Ecology*, pages 11–12, 2012.

[39] A. Saltelli, S. Tarantola, and K.P.S. Chan. A quantitative model independent method for global sensitivity analysis of model output. *Technometrics*, 41:39–56, 1999.

[40] Z. Shi-Mai and Y. Luo. Long-distance spread and interregional epidemics of wheat stripe rust in China. *The American Phytopathological Society, Plant Disease*, 90(8):980–988, 2008.

[41] S. Soubeyrand, A.L. Laine, I. Hanski, and A. Penttinen. Spatiotemporal structure of host-pathogen interactions in a metapopulation. *The American Naturalist*, 174(3), 2009.

[42] J Vaillant, G Puggioni, L.A. Waller, and J Daugrois. A spatiotemporal analysis of the spread of suger cane yellow leaf virus. *Journal of time series analysis*, 2010.

[43] David M. Walker, F. Javier Perez-Barberia, and Glenn Marion. Stochastic modelling of ecological processesusing hybrid gibbs samplers. *Ecological Modelling*, 198:40 – 52, 2006.