Forecasting Pathogen Dynamics with Bayesian Model-Averaging: Application to Xylella fastidiosa

Candy Abboud, Eric Parent, Olivier Bonnefon, Samuel Soubeyrand

Forecasting pathogen dynamics with Bayesian model-averaging: Application to *Xylella fastidiosa*

2023

Candy Abboud[1,2,*], Eric Parent[3], Olivier Bonnefon[2], Samuel Soubeyrand[2,**]

[1] Current affiliation: College of Engineering and Technology, American University of the Middle East, Kuwait

[2] INRAE, BioSP, 84914 Avignon, France

[3] AgroParisTech, INRAE, UMR 518 Math. Info. Appli., Paris, France

[*] Corresponding author: `candy.abboud@aum.edu.kw`

[**] Secondary corresponding author: `samuel.soubeyrand@inrae.fr`

Abstract

Forecasting invasive-pathogen dynamics is paramount to anticipate eradication and containment strategies. Such predictions can be obtained using a model grounded on partial differential equations (PDE; often exploited to model invasions) and fitted to surveillance data. This framework allows the construction of phenomenological but concise models relying on mechanistic hypotheses and real observations. However, it may lead to models with overly rigid behavior and possible data-model mismatches. Hence, to avoid drawing a forecast grounded on a single PDE-based model that would be prone to errors, we propose to apply Bayesian model averaging (BMA), which allows us to account for both parameter and model uncertainties. Thus, we propose a set of different competing PDE-based models for representing the pathogen dynamics, we use an adaptive multiple importance sampling algorithm (AMIS) to estimate parameters of each competing model from surveillance data in a mechanistic-statistical framework, we evaluate the posterior probabilities of models by comparing different approaches proposed in the literature, and we apply BMA to draw posterior distributions of parameters and a posterior forecast of the pathogen dynamics. This approach is applied to predict the extent of *Xylella fastidiosa* in South Corsica, France, a phytopathogenic bacterium detected *in situ* in Europe less than 10 years ago (Italy 2013, France 2015). Separating data into training and validation sets, we show that the BMA forecast outperforms competing forecast approaches.

Keywords: Bayesian model-averaging; Importance sampling; Partial differential equations; Outbreak prediction; *Xylella fastidiosa*

1   Introduction

The emergence of exogenous pathogens in new territories may induce severe sanitary and socio-economical crises. Mainly, such crises are reinforced by the eventual long delay between pathogen

establishment in the new territory and its detection (Jones and Baker, 2004; Faria et al., 2014; Soubeyrand et al., 2018). As this delay increases, the potential for pathogen expansion and the resulting cost for pathogen eradication or containment generally increases. Hence, reconstructing the past dynamics of the pathogen (Boys et al., 2008; Roques et al., 2016; Soubeyrand and Roques, 2014) and predicting its future extent (Chapman et al., 2015; Peterson et al., 2003) are key steps for understanding the pathogen epidemiology, designing eradication or containment strategies and assessing their potential efficiency.

Partial differential equations (PDE) have been extensively used for modeling spatio-temporal population dynamics (Skellam, 1951; Shigesada et al., 1995; Turchin, 1998; Okubo and Levin, 2002). PDE can precisely be used for past dynamics reconstruction and future extent prediction, by exploiting their ability (i) to represent dynamics in a phenomenological and concise way, and (ii) to be fitted to data by attaching a probabilistic model of observations within a state-space modeling framework (the combination of (i) and (ii) corresponds to the so-called physical-statistical or mechanistic-statistical approach; Berliner, 2003; Wikle, 2003; Roques et al., 2011; Soubeyrand and Roques, 2014). To apply such an approach, the PDE is generally chosen to be parsimonious for identifiability reasons (i.e., the ability to estimate its parameters given the information contained in data). However, when left parsimonious, a PDE may not be proficient in describing all the processes and sources of variability involved in epidemiological dynamics. In addition, various structures of PDE are likely to be considered as candidate models for a given epidemic. When the goal of the study is to draw predictions, the use of a single model is prone to prediction error because this model may not have taken into account crucial drivers of the dynamics. This limitation can be circumvented by associating random terms to the PDE as proposed by Wikle (2003), or by considering a set of competing PDE-based models and applying either a model selection strategy (Burnham et al., 1995) or a model aggregation strategy (Hoeting et al., 1999).

As part of the model aggregation strategy, the Bayesian model averaging (BMA) approach has been proposed by Leamer (1978) to reduce and account for parameter and model uncertainties. This approach consists in performing a weighted average of candidate models in a Bayesian way and hence combining multiple predictions and multiple estimations of shared parameters (Raftery, 1996; Madigan and Raftery, 1994; Wintle et al., 2003). Theoretically, BMA provides better average predictive ability, as measured by a logarithmic scoring rule, than using any single model (Madigan and Raftery, 1994). Axiomatically, this result depends on the assumption that the data are generated in the following stages (Fletcher, 2018): (i) a model is selected at random from the set of candidate models using prior model probabilities, (ii) the parameter values for this model are generated using the relevant prior distribution, and (iii) the data are generated from the selected model and parameter values.

The BMA efficiency has been largely explored, in particular with respect to its theoretical properties (Rubin and Schenker, 1986; Madigan and Raftery, 1994), leave-one-out predictive performance (Madigan et al., 1995; Lamon and Clyde, 2000) and numerical performance (George and McCulloch, 1993; Viallefont et al., 2001). While BMA is an intuitively attractive solution to the problem of accounting for model uncertainty, it presents several difficulties related to its numerical implementation (Hoeting et al., 1999). By dint of some pioneering work implementing BMA (Madigan and Raftery, 1994; Raftery, 1996), this approach has then been applied in numerous study domains such as medicine (Oehler et al., 2009), ecology (Wintle et al., 2003), meteorology (Raftery et al., 2005), genetics (Yeung et al., 2005), economical and political sciences (Sidman et al., 2008), engineering and physical sciences (Parkinson and Liddle, 2013) and epidemiology (Viallefont et al., 2001). Despite ample literature on BMA and its usefulness, it has been marginally applied in the context of predictive epidemiology.

In this article, we investigate the application of BMA in the context of pathogen-dynamics prediction using PDE-based models and we want to test its efficiency on a real case study. The models are grounded on a family of reaction-diffusion equations, some of which include spatially heterogeneous diffusion and reproduction terms. Our aim is to compute, from post-introduction data, the BMA posterior distribution of a certain quantity of interest Δ, which is typically the introduction time or location of the pathogen or its future spatial extent. Following Abboud

et al. (2019), we apply to each model the Adaptive Multiple Importance Sampling algorithm (AMIS; Cornuet et al., 2012) for providing an empirical approximation, obtained via a weighted sample $\{\Delta_n, w_n\}_{n=1}^{N}$ of size N, of the posterior distribution of Δ given the specified model. Then, for drawing BMA posterior samples of Δ, we compute posterior probabilities of models using different approximations of the integrated likelihood that have been proposed in the literature. Namely, we compare an estimator of the integrated likelihood, which is easily obtained by averaging AMIS un-normalized importance weights (Bugallo et al., 2015), to estimators of the integrated likelihood grounded on information criteria (McElreath, 2018), as well as harmonic mean estimators (Raftery, 1996; Gelfand and Dey, 1994).

This approach is first tested on simulated data and then applied to make predictions concerning the dynamics of the phytopathogenic bacterium *Xylella fastidiosa* (Xf) in South Corsica, France. This quarantine pathogen in Europe has significantly impacted olive production in Puglia, Italy, and presents a drastic risk of environmental degradation due to its ability to reach a large variety of plant species. It is currently present in a large part of Corsica island and more marginally in Southern mainland France (Denancé et al., 2017a; Soubeyrand et al., 2018; Martinetti and Soubeyrand, 2019). Xf might cause a major sanitary crisis in France, as the one caused in Italy since 2013 where the socio-economical impacts are considerable due to the death and felling of numerous olive trees in Puglia. In the case of South Corsica, spatio-temporal and presence-absence post-introduction surveillance data were collected from an intensive surveillance plan implemented by governmental agencies after the first *in situ* detection of Xf in 2015 in the city of Propriano. These data covering about three years and a half are separated into a training set (nearly 2 years) used for fitting the models and a validation set (nearly 1.5 years) used for comparing the forecasts obtained with BMA, the best PDE-based model (i.e., the model with the largest posterior model probability computed in the BMA procedure), an ensemble approach (i.e., the equiprobable average of all the PDE-based models parameterized by the maximum likelihood estimates) and two data-oriented prediction strategies not grounded on PDE.

The paper is organized as follows: Data are briefly described in Section 2. The competing models coupling a partial differential equation and a Bernoulli observation process are presented in Section 3. The Bayesian model averaging technique is described in Section 4. The simulation study is presented in Section 5. Results obtained from surveillance data for Xf in South Corsica are detailed in Section 6 where we specifically focus on model comparison, parameter inference and out-of sample predictive performance. Finally, Section 7 provides a conclusion and a discussion of perspectives.

2   Surveillance data with presence-absence records

In this article, we analyze spatio-temporal presence-absence data collected in South Corsica, France, and informing if sampled plants are positive or negative to Xf based on a molecular test. Data have been collected since the first detection of the bacterium in the study region in July 2015. Between July 2015 and January 2019, approximately 9500 plants were sampled, among which 900 have been diagnosed as infected with a real-time polymerase chain reaction (real-time PCR) technique (Denancé et al., 2017b). GPS coordinates, sampling dates and sanitary statuses (healthy/infected) are available for all the sampled plants. Spatial locations and sanitary statuses at the sampling times are shown in Figure 1, left.

As for other bacteria, the growth and mortality of Xf are affected by various environmental variables such as habitability of the environment, nutrients, climatic conditions and availability of dissemination means (typically, insect vectors). In this study, to account for spatial heterogeneity in the diffusion and the reproduction regimes of the epidemics, we use temperature data to divide the spatial domain denoted by Ω into two complementary and non-overlapping sub-domains, and different diffusion and growth terms are applied to the two sub-domains. We exploit a freely available database (PVGIS © European Communities, 2001–2008) providing, in particular, monthly averages of the daily minimum temperature reconstructed over a grid

with 1×1km spatial resolution (Huld et al., 2006); these monthly averages correspond to the period 1995-2003, which is between the probable introduction around 1960 of Xf in South Corsica (Soubeyrand et al., 2018) and the observation window starting in 2015. We use these data to build the average of the daily minimum temperature over January and February, say $T(\mathbf{x})$ for any location $\mathbf{x}$; see Figure 1, right. Average daily minimum temperature in Winter is known to be a crucial factor for the presence or abundance of Xf (Anas et al., 2008; Martinetti and Soubeyrand, 2019).

Remark. In this work we divide the spatial domain based on a static assessment of winter temperatures. This allows us to take into account a relatively large spatial heterogeneity of the environment since the winter temperature, as we defined it, ranges from -0.7°C in the coldest area of South Corsica to 6.8°C in its warmest area. Ideally, we should also take into account the annual variability of winter temperatures and eventual other environmental variables in the dynamic process, as proposed by Botella et al. (2022) for instance. Such data may be obtained from a data base such as WorldClim (`https://worldclim.org/`) at the 1×1km resolution that we consider here.

3 Competing models

Here, a set of models based on parabolic partial differential equations is used to describe pathogen dynamics at large spatial scales. As explained in the introduction, these models have been extensively used to represent population dynamics in a phenomenological and concise way, and can be fitted to data in a hierarchical modeling setting incorporating a probabilistic model of observations. In this section, we propose a family of mechanistic models and we present the model for the observation process.

3.1 Family of mechanistic models

We introduce a discrete family $\boldsymbol{\mathcal{M}} = \{\mathcal{M}_i(\tilde{T}) : 0 \leq i \leq I; \tilde{T} \in \mathcal{T}\}$ of models governing the probability $u(t, \mathbf{x})$ of a host located at site $\mathbf{x} = (x_1, y_1) \in \Omega$ to be infected (i.e., sick because of the pathogen under consideration) at time t, where $I \in \mathbb{N}^*$ and $\mathcal{T}$ is a finite collection of real values. The label i refers to a model structure, i.e. a specific form for the parabolic PDE. The label $\tilde{T}$ refers in our application to a temperature threshold, which splits the spatial domain into two sub-domains where diffusion and growth terms may be different. The generic form of models in family $\boldsymbol{\mathcal{M}}$ satisfies:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \boldsymbol{\Delta}(D_{i\tilde{T}}(\mathbf{x})u) + f_{i\tilde{T}}(u), & t \geq \tau_0, \ \mathbf{x} \in \Omega, \\ \boldsymbol{\nabla}(D_{i\tilde{T}}(\mathbf{x})u(t,\mathbf{x})).n(\mathbf{x}) = 0, & t \geq \tau_0, \ \mathbf{x} \in \partial\Omega, \\ u(\tau_0, \mathbf{x}) = u_0(\mathbf{x}), & \mathbf{x} \in \Omega, \end{cases} \quad (1)$$

where the first line is the reaction-diffusion equation, the second line gives boundary conditions, the third line gives initial conditions, $\boldsymbol{\Delta} = \dfrac{\partial^2}{\partial x_1^2} + \dfrac{\partial^2}{\partial x_2^2}$ is the 2-dimensional diffusion operator of Laplace, and $\boldsymbol{\nabla} = \dfrac{\partial}{\partial x_1} + \dfrac{\partial}{\partial x_2}$ is the 2-dimensional gradient operator.

The diffusion coefficient $D_{i\tilde{T}}(\mathbf{x})$ may be spatially heterogeneous and is defined as a spatial regularization of $d_{i\tilde{T}}(\mathbf{x}) = \sum_{k=1}^{2} D_{i\tilde{T}k} \mathbb{1}(\mathbf{x} \in \Omega_{\tilde{T}k})$, $\forall i \leq I, \forall \tilde{T} \in \mathcal{T}$, where $\mathbf{x} \mapsto \mathbb{1}(\mathbf{x} \in \Omega_{\tilde{T}k})$ is the indicator function taking the value 1 if $\mathbf{x} \in \Omega_{\tilde{T}k}$ and 0 otherwise, and the sub-domains $\Omega_{\tilde{T}1}$ and $\Omega_{\tilde{T}2}$ are defined by thresholding the spatial function T, with the threshold value $\tilde{T}$ such that: $\Omega_{\tilde{T}1} = \{\mathbf{x} \in \Omega : T(\mathbf{x}) > \tilde{T}\}$ and $\Omega_{\tilde{T}2} = \{\mathbf{x} \in \Omega : T(\mathbf{x}) \leq \tilde{T}\}$. In our application, T is a measure of temperature in winter, $\Omega_{\tilde{T}1}$ is the warm region of Ω, and $\Omega_{\tilde{T}2}$ the cold one. If $D_{i\tilde{T}1} = D_{i\tilde{T}2}$, then the diffusion coefficient is spatially homogeneous. The spatial regularization is required for the existence and the uniqueness of a classic solution $u(t, \mathbf{x})$ of Equation (1); see

Roques (2013). Thus $D_{i\tilde{T}}$ is defined as:

$$D_{i\tilde{T}}(\mathbf{x}) = 1(x \in \Omega) \int_{\mathbb{R}^2} \phi(\mathbf{x} - \mathbf{y}) d_{i\tilde{T}}(\mathbf{y}) d\mathbf{y}, \ \forall \mathbf{x} \in \Omega, \tag{2}$$

where ϕ is the normal regularization kernel: $\phi(\mathbf{x}) = \dfrac{1}{2\pi\mathcal{V}} \exp\left(-\dfrac{\|\mathbf{x}\|^2}{2\mathcal{V}}\right)$, and the transition speed $\mathcal{V}$ has to be tuned to approach more or less the piecewise constant function $d_{i\tilde{T}}$. In other words, $D_{i\tilde{T}}$ locally smoothes the step function $d_{i\tilde{T}}$ around the transition border between the two subdomains with a bandwidth determined by $\mathcal{V}$.

The reproduction term may also be either spatially heterogeneous or not. In the homogeneous case, we consider a logistic growth

$$f_{i\tilde{T}}(u) = bu\left(1 - \frac{u}{K}\right), \tag{3}$$

where b is the intrinsic growth rate of the epidemic and $K \in (0, 1]$ is a plateau for the probability of infection (i.e., an analog to the carrying capacity of the environment). With such a reproduction term, the growth is nearly exponential when u is small, and the growth rate is around zero when u reaches the plateau K, which gives the maximum prevalence of the pathogen in the environment. In the heterogeneous case, the growth is logistic in $\Omega_{\tilde{T}1}$ (i.e., warmest areas) and negative in $\Omega_{\tilde{T}2}$ (i.e., coldest areas), mimicking a source-sink dynamics:

$$f_{i\tilde{T}}(u) = bu\left(1 - \frac{u}{K}\right)\mathbb{1}(\mathbf{x} \in \Omega_{\tilde{T}1}) - \alpha u \mathbb{1}(\mathbf{x} \in \Omega_{\tilde{T}2}), \forall \tilde{T} \in \mathcal{T}, \tag{4}$$

where α is the decrease rate of the infection in $\Omega_{\tilde{T}2}$. Thus, the pathogen is affected by two contradictory forces in the cold region: it is forced to decline with the rate α but may be at the same time reinforced by the sources of infection located in the warm region.

In the application, we will consider three model structures:

- $\mathcal{M}_0(\tilde{T})$, under which $D_{i\tilde{T}1} = D_{i\tilde{T}2}$ and $f_{i\tilde{T}}$ satisfies Equation (4), i.e. homogeneous diffusion but heterogeneous growth, like in Abboud et al. (2019);

- $\mathcal{M}_1(\tilde{T})$, under which $D_{i\tilde{T}1} \neq D_{i\tilde{T}2}$ and $f_{i\tilde{T}}$ satisfies Equation (4), i.e. heterogeneous diffusion and growth;

- $\mathcal{M}_2(\tilde{T})$, under which $D_{i\tilde{T}1} \neq D_{i\tilde{T}2}$ and $f_{i\tilde{T}}$ satisfies Equation (3), i.e. heterogeneous diffusion, but homogeneous growth.

The second equation in System (1) corresponds to the homogeneous Neumann condition on the boundary $\partial\Omega$ of Ω (i.e. with reflection on the boundary). This condition is formalized by setting that the gradient of the spatial function $x \mapsto D_{i\tilde{T}}(\mathbf{x})u(t, \mathbf{x})$ is orthogonal to the outward normal vector $n(x)$ at point x on the boundary $\partial\Omega$, for all $t \geq \tau_0$. Thus, physically, there is neither outward nor inward flux from and to Ω.

The spatial function u_0 models the introduction of the pathogen in the study domain at time $\tau_0 \in \mathbb{R}$. Following Abboud et al. (2019), the introduction represents the initial phase of the outbreak corresponding to the arrival of the pathogen and its local establishment. Thus, u_0 is not expressed as a Dirac function but as a kernel function centered around the central point of the introduction $\tilde{\mathbf{x}}_0 = (\tilde{x}_0, \tilde{y}_0) \in \Omega$. More precisely, the probability of a host at $\mathbf{x}$ to be infected at τ_0 satisfies:

$$u_0(\mathbf{x}) = P_0 \exp\left(-\frac{\|\mathbf{x} - \tilde{\mathbf{x}}_0\|^2}{2\sigma^2}\right), \tag{5}$$

where P_0 is the infection probability at $(\tau_0, \tilde{\mathbf{x}}_0)$, $\sigma^2 = \frac{r_0^2}{q}$, q is the 0.95-quantile of the χ^2 distribution with two degrees of freedom, and r_0 is the *radius* of the kernel. Thus, at τ_0, if we neglect border effects, 95% of the infected plants are located within the ball with center $\tilde{\mathbf{x}}_0$ and radius r_0.

With these initial and boundary conditions, the equation system (1) is well-posed (Evans, 1998) and, by constraining P_0 in $[0, K]$, the principle of parabolic comparison (Protter, MH and Weinberger, HF, 1967) implies that the solution of (1) remains in the interval $[0, K]$.

3.2 Probabilistic model of the observation process

Let $t_j \in \mathbb{R}$ denote the sampling time of host $j \in \{1, \ldots, J\}$, $J \in \mathbb{N}^*$, $\mathbf{x}_j \in \Omega$ its location and $Y_j \in \{0, 1\}$ its observed sanitary status at t_j (1 for infected, 0 for healthy). Given u, $\mathcal{M}_i(\tilde{T})$ and $\{(t_j, \mathbf{x}_j) : 1 \leq j \leq J\}$, the sanitary statuses Y_j, $j \in \{1, \ldots, J\}$, are independent random variables following Bernoulli distributions with success probabilities $u(t_j, \mathbf{x}_j)$:

$$Y_j \mid u, \mathcal{M}_i(\tilde{T}), \{(t_j, \mathbf{x}_j) : 1 \leq j \leq J\} \underset{\text{indep.}}{\sim} \text{Bernoulli}(u(t_j, \mathbf{x}_j)), \qquad (6)$$

where u depends on the model $\mathcal{M}_i(\tilde{T})$ and its vector of parameters $\Theta_{i\tilde{T}}$.

Remark. This data model was proposed in Abboud et al. (2019) for its simplicity. It could be refined to account for sampling errors classically encountered in epidemiology, e.g. false-positive and false-negative observations, as well as spatial and temporal dependencies not accounted for in the process model.

4 Bayesian model averaging

4.1 Principle

Briefly, the BMA consists in estimating the expectation of the posterior distributions of the variable of interest Δ provided under all the competing models and weighted by the posterior model probabilities (Raftery, 1996; Hoeting et al., 1999). In the modeling setting introduced above, Δ is typically a vector of shared parameters such as the introduction point $(\tilde{\mathbf{x}}_0, \tau_0)$, the temperature threshold $\tilde{T}$ or the spatial probability of infection u over a future period. Using Gelfand's bracket notation for probability distributions (Gelfand and Smith, 1990), the BMA posterior distribution of Δ given training data Y satisfies:

$$[\Delta|Y] = \sum_{0 \leq i \leq I, \tilde{T} \in \mathcal{T}} [\Delta|Y, \mathcal{M}_i(\tilde{T})] \times [\mathcal{M}_i(\tilde{T})|Y]. \qquad (7)$$

The posterior model probability of $\mathcal{M}_i(\tilde{T})$ is:

$$[\mathcal{M}_i(\tilde{T})|Y] = \frac{[Y|\mathcal{M}_i(\tilde{T})] \times [\mathcal{M}_i(\tilde{T})]}{\sum\limits_{0 \leq i' \leq I, \tilde{T}' \in \mathcal{T}} [Y|\mathcal{M}_{i'}(\tilde{T}')] \times [\mathcal{M}_{i'}(\tilde{T}')]}. \qquad (8)$$

The integrated likelihood $[Y|\mathcal{M}_i(\tilde{T})]$ of $\mathcal{M}_i(\tilde{T})$, which may be a complex integral depending on the dimension of the unknowns and eventual dependencies, satisfies:

$$[Y|\mathcal{M}_i(\tilde{T})] = \int [Y|\Theta, \mathcal{M}_i(\tilde{T})] \times [\Theta|\mathcal{M}_i(\tilde{T})] d\Theta. \qquad (9)$$

where Θ is the vector of parameters of $\mathcal{M}_i(\tilde{T})$, $[Y|\Theta, \mathcal{M}_i(\tilde{T})]$ is the likelihood under $\mathcal{M}_i(\tilde{T})$, $[\Theta|\mathcal{M}_i(\tilde{T})]$ is the prior distribution of Θ under $\mathcal{M}_i(\tilde{T})$, and $[\mathcal{M}_i(\tilde{T})]$ is the prior probability of $\mathcal{M}_i(\tilde{T})$. The posterior mean of Δ is a weighted average of the posterior means under the competing models:

$$\mathbb{E}[\Delta|Y] = \sum_{0 \leq i \leq I, \tilde{T} \in \mathcal{T}} [\mathcal{M}_i(\tilde{T})|Y] \times \mathbb{E}[\Delta|\mathcal{M}_i(\tilde{T}), Y]. \qquad (10)$$

The posterior variance is expressed as follows:

$$\mathbb{V}[\Delta|Y] = \sum_{0 \le i \le I, \tilde{T} \in \mathcal{T}} [\mathcal{M}_i(\tilde{T})|Y] \times \left(\mathbb{V}[\Delta|\mathcal{M}_i(\tilde{T}), Y] + \left(\mathbb{E}[\Delta|\mathcal{M}_i(\tilde{T}), Y] - \mathbb{E}[\Delta|Y] \right)^2 \right). \quad (11)$$

4.2   Implementation

The BMA-posterior distribution of Δ is computed from the following two-step process: First, we compute the posterior distribution of Δ given model $\mathcal{M}_i(\tilde{T})$ and training data Y for all $i \in \{1, \dots, I\}$; Second, we compute the posterior model probabilities.

4.2.1   Approximation of the posterior distribution of Δ given a model $\mathcal{M}_i(\tilde{T})$

For any mechanistic-statistical model defined in Section 3, neither the posterior distribution nor the likelihood can be expressed analytically, especially because they depend on the solution u of the PDE that cannot be written in a closed form. Thus, following Abboud et al. (2019), we use the adaptive multiple importance sampling algorithm (AMIS; Cornuet et al., 2012), that consists of successively generating parameter vectors under an adaptive proposal distribution and assigning / updating weights for the parameter vectors. To design efficient importance sampling algorithms, the auxiliary proposal distribution should be chosen as close as possible to the posterior distribution. However, the posterior distribution being unknown, the crucial choice of the proposal is a difficult task (Gelman et al., 1996). The main aim of AMIS is to overcome this difficulty by tuning the coefficients of the proposal distribution picked from a parametric family of distributions, generally the Gaussian one, at the end of each iteration. In this framework, at each iteration, new coefficient values for the proposal distribution are determined using the current weighted posterior sample, then the posterior sample is augmented by generating new replicates from the newly tuned proposal distribution and the weights of the cumulative posterior sample are updated. The AMIS algorithm provides a weighted posterior sample $\{(\Delta_m^l, w_m^l) : 1 \le m \le M, 1 \le l \le L\}$ of size ML, which forms an empirical approximation of the posterior distribution $[\Delta|Y, \mathcal{M}_i(\tilde{T})]$ (m stands for the iteration; l stands for the replicate generated at iteration m). Conditions leading to the convergence in probability of the posterior mean of any function (integrable with respect to the posterior distribution) of the parameters are described in Cornuet et al. (2012) and Marin et al. (2019), and are satisfied in our case (Abboud et al., 2019).

   We implemented the AMIS algorithm in the `R` statistical software, with calls to the software `Freefem++` for numerically solving the PDE in the mechanistic-statistical environment (MSE; `https://informatique-mia.inrae.fr/mse/`). We performed parallel computation to compensate for the non-negligible time of PDE resolution. With $(M, L) = (50, 10^4)$ (which ensured the stabilization of the posterior samples using the deviation measure described by Abboud et al., 2019) and the use of 100 cluster cores (the cluster being composed of 40-cores nodes Xeon(R) 2.2 Ghz, 228 Go RAM), the estimation procedure for one model takes approximately 1.75 days in average. Unlike the MCMC algorithm that is often used in the mechanistic-statistical framework (Soubeyrand and Roques, 2014; Lanzarone et al., 2017), AMIS, as a purely Monte Carlo algorithm, can be easily parallelized, its tuning parameters are automatically adapted at each iteration, and all the samples generated throughout the algorithm are recycled thanks to the update of weights at each iteration. The AMIS algorithm provides at each iteration an assessment of the posterior distribution of parameters, which is expected to be stable after a burn-in period and to converge to the true posterior distribution.

4.2.2   Computation of the integrated likelihood

Computing the posterior model probability requires the evaluation of the integrated likelihood as shown in Equation (8). Various methods to estimate the integrated likelihood (that is not analytically tractable in our case) are encountered in the literature. In this article, we briefly

assess several of these methods with respect to their impact on BMA predictions. Nine integrated likelihood estimators, described in Online Resource 1, Section S1, are considered: the un-normalized weight estimator denoted by UWE (Bugallo et al., 2015), the estimator based on an information criterion (McElreath, 2018) of which we consider six different specifications denoted by WAIC_1, WAIC_2, BIC, DIC_1, DIC_2 and IC, and the harmonic mean estimator (Newton and Raftery, 1994; Gelfand and Dey, 1994) of which we consider two specifications denoted by HME_1 and HME_2.

4.2.3 Priors and posterior samples

For the applications, we assume as a prior knowledge that the models are equally weighted. Because several model structures with different sets of parameters are considered, the prior distribution of Θ partly depends on the model structure. These distributions combine vague uniform and Dirac distributions (Dirac distributions are considered for r_0 and p_0 for identifiability issues) and are provided in Online Resource 1, Section S2. AMIS is then applied to obtain a weighted posterior sample of size $ML = 5 \times 10^5$ for each candidate model; see Section 4.2.1. Posterior model probabilities, empirical approximations of BMA posterior distributions and other posterior quantities (including predicted infection maps) were approximated by sampling with replacement 10^4 models$\times$parameters with respect to model and parameter weights.

5 Application to simulated data

We perform a simulation study to illustrate the heterogeneity in the predictive performance of the inference approach (by focusing on the posterior model probabilities) when one considers the different methods presented in Section 4.2.2 for computing the integrated likelihood.

The simulation study is carried out by generating three different data sets $\{\mathcal{O}^{(g)} : g = 1, 2, 3\} = \{(t_j, \mathbf{x}_j, Y_j^{(g)}) : 1 \leq j \leq J\}$ from two different generative models. The simulations are performed by using characteristics of real data analyzed in the next section: we use the same spatial domain Ω, observations locations $\mathbf{x}_j$ and observation times t_j and, for most of the parameters, we use values close to parameter estimates obtained in the next section. The two models that we consider are $\mathcal{M}_0(5.5)$, in which the diffusion is homogeneous ($D_1 = D_2$), and $\mathcal{M}_1(5.5)$, in which the diffusion is heterogeneous. For the latter model, we consider two cases: $D_2 = 0.9D_1$ and $D_2 = 0.1D_1$. Table S1 in Online Resource 1 summarizes model specifications and provides parameter values. Figures S1–S3 in Online Resource 1 display the posterior map of the introduction location and the marginal posterior distributions of the other parameters when the true model is fitted to data. We can observe that the true parameter values are all in their respective 95% credible intervals.

The relevancy of each approach for computing posterior model probabilities is assessed by fitting the two models $\mathcal{M}_0(5.5)$ and $\mathcal{M}_1(5.5)$ to the three simulated data sets and by checking, for each data set, whether the true model has the largest posterior probability. Both candidate models $\mathcal{M}_0(5.5)$ and $\mathcal{M}_1(5.5)$ have equal prior weights. Posterior model probabilities provided in Table 1 are computed with each of the nine estimators of the integrated likelihood listed in Section 4.2.2. Only posterior model probabilities estimated from UWE, HME_2, WAIC_1 and WAIC_2 are consistent in the sense that the true model has the largest posterior probability.

In the real case study presented below, we use the estimator grounded on un-normalized weights (UWE) because it is considered as one of the most efficient approaches for assessing the integrated likelihood, displaying desirable convergence properties (Ford and Gregory, 2007), and because, from a practical viewpoint, it is easily obtained from AMIS output.

6 Application to *Xylella fastidiosa* data

8

6.1 Model Comparison

In the analysis of *Xylella fastidiosa* (Xf) data in South Corsica, we first compare the 27 models under consideration (the three model shapes presented in Section 3 times nine temperature thresholds, from 4°C to 6°C every 0.25°C). The models are fitted to training data consisting of all the samples from July 2015 to April 2017 (subsequent data from May 2017 to January 2019 are used for assessing predictive performance of diverse forecast approaches in Section 6.3). The stabilization of AMIS assessed as in Abboud et al. (2019) is satisfactory for all models (see Online Resource 1, Figure S4 for an example). Posterior model probabilities are provided in Online Resource 1, Table S2, for the computation based on UWE as well as for the computations based on the other approximation approaches. Regarding the best model, UWE and HME_2 lead to $\mathcal{M}_0(4.75)$ (homogeneous diffusion but heterogeneous growth with a temperature threshold equal to 4.75°C), DIC_1 and IC lead to $\mathcal{M}_0(5.75)$ and the others lead to $\mathcal{M}_1(5.75)$ (homogeneous diffusion and growth). Thereafter, we only analyze results obtained with UWE, as explained in Section 5. Overall, models with the largest posterior probabilities computed from UWE are $\mathcal{M}_0$ and $\mathcal{M}_1$ with temperature thresholds from 4.75°C to 5.5°C.

6.2 Inference from BMA

BMA posterior distributions of shared parameters (the temperature threshold, the introduction time and the introduction location) are displayed in Figures 2 and 3. Figure 2 illustrates an advantage of BMA since one obtains a posterior distribution for the temperature threshold instead of the unique value obtained by Abboud et al. (2019) with a model selection approach. The introduction of Xf tends to be relatively ancient but uncertain (posterior mean: 1956; posterior median: 1954; posterior SD: 15 years; 95%-posterior interval: [1933;1988]). The inference of the introduction date provided by the best model is slightly less variable (posterior mean: 1958; posterior median: 1957; posterior SD: 14 years; 95%-posterior interval: [1934;1986]). A larger difference in estimation uncertainty arises for the introduction location, which is more uncertain in the BMA inference than, for instance, in the inference resulting from the model selected by Abboud et al. (2019, Fig. 6). The BMA estimation possibly better reflects the uncertainty about the introduction location than a single-model estimation.

6.3 Forecast performance

We assess forecast performance by considering different in-sample predictors trained over the period from July 2015 to April 2017 (8152 observations) and evaluated over the period from May 2017 to January 2019 (1523 observations). We consider five in-sample predictors:

- Best model: the posterior mean of u computed and averaged over the period May 2017–January 2019, which is provided by the best model selected from and fitted to training data (the best model is the model with the highest posterior model probability given by Equation (8));

- BMA: the posterior mean of u computed and averaged over the period May 2017–January 2019, which is provided by BMA applied to training data;

- Ensemble: the mean of the 27 maximum likelihood estimates (MLE) of u computed and averaged over the period May 2017–January 2019 obtained under the 27 models, where the MLE of u corresponds for each model to the parameter vector simulated in the AMIS algorithm leading to the largest likelihood value;

- Climatology: the so-called climatology forecast (Mason, 2004), which is simply the mean of $\{Y_j : t_j \in [\text{July 2015–April 2017}]\}$;

- Kernel smoother: the spatial kernel smoother of $\{(\mathbf{x}_j, Y_j) : t_j \in [\text{July 2015–April 2017}]\}$, using the Epanechnikov kernel, which is proportional to $d \mapsto (1 - d^2)\mathbb{1}(|d| \leq 1)$, where d is the geographical distance scaled by a bandwidth value ranging from 2.5 to 25km. We assume that the kernel smoother predicts 0 if no observations are available within the bandwidth.

9

The in-sample forecasts are computed over a regular square grid with 1km×1km cells covering Ω and are compared to several benchmark out-of-sample forecasts computed over the same grid and offering different visions of the *true* sanitary situation between May 2017 and January 2019. Two types of out-of-sample forecasts are considered:

- Local infection proportions: the raw mean of Y_j observed in each grid cell from May 2017 to January 2019. Sampling over the validation period was made in a limited number of grid cells mostly distributed near the shoreline. In particular, areas with very low infection risk based on earlier observations were not sampled (typically high altitude areas). Hence, to better represent the sanitary situation across space in the validation map, we augmented the set of grid cells where local infection proportions could be computed as raw means of Y_j with a set of grid cells where local infection proportions were fixed at zero. This set, displayed in Figure S5 in Online Resource 1, was constructed by identifying grid cells far from past detections of Xf and far from observations during the validation period, that is to say grid cells satisfying two conditions: (i) no infection observed in training data up to 10 km; (ii) no observation in validation data up to 10 km.

- Smoothed infection proportions: the spatial kernel smoother of validation data, i.e. $\{(\mathbf{x}_j, Y_j) : t_j \in [\text{May 2017–January 2019}]\}$, computed at each grid cell center. Here also, the smoother is grounded on an Epanechnikov kernel with bandwidth values ranging from 2.5 to 25km, and it predicts 0 if no observations are available within the bandwidth.

The motivation for considering as benchmarks the smoothed infection proportions with varying bandwidths (in addition to the raw local infection proportions) is that we want to assess the ability of the competing predictors to forecast infection probability at different spatial scales.

Thus, Figure 4 shows the benchmark out-of-sample predictions consisting of the local infection proportions and the smoothed infection proportions obtained with the 15km-kernel bandwidth (top right box), and the five in-sample predictions (as well as posterior standard deviations of u for the best model and the BMA; bottom box). Figures S6 and S7 in Online Resource 1 display the same plots for a kernel bandwidth equal to 5 and 25km, respectively (only the in- and out-of-sample smoothers are different in these figures). Overall, the best model, the BMA, the ensemble and the kernel smoother provide spatially heterogeneous predictions that may be relevant but a quantitative assessment of the prediction quality is needed to finely compare the predictions (see the paragraph below). Furthermore, like for the estimation of the introduction location, BMA yields a clearly more uncertain inference of the infection probability u across space than the best model.

The root-mean-squared error (RMSE) with respect to each benchmark out-of-sample forecast is computed to measure the predictive performance of the BMA, the best model, the ensemble, the climatology and the kernel smoothing, i.e. to measure how close are the forecasts to the *true* sanitary situation between May 2017 and January 2019. This quantity is calculated over the regular square grid with 1km×1km cells already introduced above:

$$\text{RMSE} = \sqrt{\frac{1}{H} \sum_{h=1}^{H} (\hat{\bar{u}}_h - \bar{u}_h)^2},$$

where $\hat{\bar{u}}_h$ is the average (in time and space) prediction of u in grid cell h over May 2017–January 2019 provided by one of the predictors; $\bar{u}_h$ is either the local infection proportion in grid cell h over the validation period or the average (in time and space) of u in grid cell h provided by the spatial kernel smoother with bandwidth $b > 0$ applied to validation data (where b ranges from 2.5 to 25 km); H is the number of grid cells.

Figure 5 shows the RMSE values for the five in-sample forecast approaches and a kernel bandwidth ranging from 2.5 to 25km. Looking only at the comparison with the raw local infection proportions over the validation period (red symbols), the BMA outperforms the other approaches (best model, ensemble, climatology and kernel smoother whatever the bandwidth). The same holds true when one compares the competing in-sample forecasts to the out-of-sample

predictions computed as smoothed infection proportions obtained with different bandwidths (black symbols). The climatology, which predicts the same infection probability everywhere, obviously does not account for the major spatially-structured effect of cold temperatures in winter on Xf reproduction or propagation. For small bandwidths, the high probability areas identified by the kernel smoother applied to training data and to validation data are spatially close but do not exactly coincide. In contrast, the quite smooth predictions based on the BMA, the best model and the ensemble do not predict peaks of infection as observed in the out-of-sample forecast but correctly delineate regions where these peaks can arise. When the bandwidth is large, the out-of-sample forecast tends to a very smooth function that even yields significant positive infection probabilities in regions where Xf reproduction or propagation is hampered (i.e. cold regions in winter). This bias incorporated in what we called a vision of the *true* sanitary situation is also encountered in the in-sample forecast based on kernel smoothing (and to some extent in the climatology that can be considered as a kernel smoothing with very large bandwidth). This bias partly explains the improvement of the prediction ability of the kernel smoothing and the climatology when the bandwidth increases. However, even for large bandwidths, the BMA prediction is better.

7   Discussion

We have presented how to use Bayesian model averaging (BMA) to infer and predict pathogen dynamics from multiple candidate PDE-based models, with an application to *Xylella fastidiosa*. This approach has a large potential of application in epidemiology where compartmental models based on ordinary or partial differential equations are widespread and often used to forecast epidemics as illustrated during the current COVID-19 pandemic (e.g., see Bertozzi et al., 2020; Kissler et al., 2020; Roques et al., 2020, among many other references). BMA is a valuable approach to compensate for the supposed lack of flexibility of such deterministic models. BMA can also be exploited for model selection, using the posterior model probabilities to identify the best model. However, we have tested nine methods proposed in the literature to compute these probabilities and we obtained rather heterogeneous results (even if only 3 models out of 27 were identified as the most probable models by the nine methods). We chose the computation based on un-normalized weights calculated within the importance sampling algorithm. This natural choice appears to be efficient in our application (it led to a BMA-forecast clearly better than the best-model-forecast and the ensemble-forecast), but further work is required to provide robust advises for computing model weights in BMA.

The model-averaging framework that we describe in this article may be used for a broad range of applications not limited to pathogen invasion dynamics. It may especially be applied to predict more general population dynamics (invasive or not) by adapting (i) the competing PDE-based models to the species dynamics and (ii) the observation-process model to the available data. For instance, the model-averaging framework may be adapted to studies in which PDEs were fitted to data concerning the population dynamics of insects (Ovaskainen et al., 2008; Roques et al., 2011), birds (Wikle, 2003), terrestrial mammals (Louvrier et al., 2020) and marine mammals (Williams et al., 2018). The framework is not limited to PDE-based phenomenological models: other relatively concise model formalisms may be used, such as those discussed for plant, animal and pathogen dynamics by Pyšek and Hulme (2005), Schurr et al. (2012), Leitner and Kühn (2018) and Roques and Soubeyrand (2023), even if the additional flexibility eventually offered by alternative models might be an issue depending on the information brought by data. If such a case occurs, the model averaging approach may however allows the analyst to reduce estimation difficulties by simplifying a *comprehensive* model (affected by identifiability problems) into multiple *partial* models (less prone to identifiability problems), which are fitted to data independently and then combined to draw probabilistic predictions. To illustrate this proposal, consider a situation where one aims to infer the dynamics of a species (typically a plant species) with multiple possible dispersal vectors (e.g., rivers, winds, animals and humans). If fitting a *comprehensive* model accounting for all the dispersal vectors leads to identifiability problems,

one may fit *partial* models instead (each of them taking into account a single dispersal vector), and combine these models with BMA to draw predictions, to infer the relative contribution of each dispersal vector, and even to test hypotheses about them (Bartoš et al., 2021). Although appealing, this proposal should be tested with caution: considering a single dispersal vector in a model might lead to biased estimation of the effect of this vector, and might require an informative prior to avoid this bias.

Concerning the emergence and spread of *Xylella fastidiosa*, we obtained consistent results with previous studies. Thus, our estimation of the introduction date around 1956 [1933; 1988] is consistent with the phylogeny-based estimation (1971 [1924; 1994]) for the sequence type 7 of Xf provided by Dupas et al. (2023) and the estimation of Soubeyrand et al. (2018) based on a stochastic temporal model including a hidden host compartment (1985 [1978;1993]). Dupas et al. (2023) mention "a period of massive introduction of exotic plants, particularly in Corsica", starting around the 70's, that supports their estimation, referring to Jeanmonod and Natali (1997) — this statement relies on the credible assumption that the introduction of Xf in an isolated disease-free territory can occur via the introduction of infected plants. Jeanmonod and Natali (1997) indeed estimated a rather constant and high rate of introduction of exotic plant taxa in Corsica from 1970 to 1997: around 50 taxa per decade. In contrast, from 1840 to 1970, this rate was estimated to be around 20 taxa per decade, which is clearly lower, but might be sufficient to favor the introduction of Xf in Corsica. A more in-depth analysis of the introduced plant species, with regard to those which are recognized as hosts of Xf, deserves to be carried out to evaluate the different estimates of the date of introduction. By the way, such an analysis could be exploited to propose a more informative prior of the introduction date in our Bayesian approach. The introduction location, which likely took place in the region of Ajaccio (West coast of Corsica), is consistent with the estimation provided by Abboud et al. (2019), but the application of BMA is expected to better reflect the uncertainty about this location. In addition, BMA provides a posterior distribution for the threshold in winter temperature that includes the threshold identified by Abboud et al. (2019) via a model selection procedure.

Based on the ample literature on model averaging and its benefits, we expected improved predictions and more realistic uncertainty estimation (Hoeting et al., 1999; Wintle et al., 2003). Precisely, in the Xf case study, BMA outperforms the best model in terms of prediction, and credibility intervals provided by BMA are larger than those provided by the best model. This observation could reflect the fact that BMA succeeds in better assessing model uncertainty and avoids overconfidence in the estimations and the predictions. However, for firmly confirming this result, complementary studies should be conducted to assess the calibration of credibility intervals.

As illustrated by the analysis of Xf data, the deterministic candidate models summarized by Equations (1–5) and describing the dynamics of the pathogen capture, overall, the discrepancies between regions of low- and high-probabilities of infection but do not capture the details, e.g., the spatio-temporal disease clusters that can be observed using kernel smoothing with small bandwidth. These details could be implicitly taken into account by coupling the partial differential equation to stochastic terms or observed covariates that would lead to more flexible realizations of the dynamics. Additional model components and covariates could be identified from the series of work grounded on mechanistic models and machine learning investigating the dynamics of *Xylella fastidiosa* and its environmental / biological drivers (Cendoya et al., 2020; Godefroid et al., 2019; Kottelenberg et al., 2021; Martinetti and Soubeyrand, 2019; Raffini et al., 2020; Soubeyrand et al., 2018; White et al., 2020, 2017). An interesting perspective is also to include the models proposed in some of these references directly in the BMA analysis since BMA is expected to take advantage of different model structures.

Acknowledgements

Statements and declarations

The authors declare that the research was conducted in the absence of any financial or non-financial relationships that could be construed as a potential conflict of interest.

Data availability statement

Xylella fastidiosa surveillance data are provided by the ESV Platform and accessible at `https://doi.org/10.15454/RWBIWD` (ESV, 2022). Average of daily minimum temperature in January and February in Corsica are available at `https://doi.org/10.5281/zenodo.6985147` (Soubeyrand and Abboud, 2022).

Online Resource 1

Section S1: Estimators of the integrated likelihood
Section S2: Specification of prior distributions
Table S1: Specification of models for simulating data
Table S2: Posterior model probabilities for the real case study
Figure S1: Posterior inference for simulated data set $\mathcal{O}^{(1)}$
Figure S2: Posterior inference for simulated data set $\mathcal{O}^{(2)}$
Figure S3: Posterior inference for simulated data set $\mathcal{O}^{(3)}$
Figure S4: Assessment of AMIS stabilization
Figure S5: Areas of infection proportion fixed at zero in the raw validation map
Figure S6: In-sample and out-sample predictions for a 5km-kernel bandwidth
Figure S7: In-sample and out-sample predictions for a 25km-kernel bandwidth

References

Abboud, C., Bonnefon, O., Parent, E., and Soubeyrand, S. (2019). Dating and localizing an invasion from post-introduction data and a coupled reaction–diffusion–absorption model. *Journal of Mathematical Biology*, 79:765–789.

Anas, O., Harrison, U. J., Brannen, P. M., and Sutton, T. B. (2008). The effect of warming winter temperature on the severity of Pierce's disease in the Appalachian mountains and Piedmont of the southeastern United States. *Plant Health Progress.*, 101094:450–459.

Bartoš, F., Gronau, Q. F., Timmers, B., Otte, W. M., Ly, A., and Wagenmakers, E.-J. (2021). Bayesian model-averaged meta-analysis in medicine. *Statistics in Medicine*, 40:6743–6761.

Berliner, L. M. (2003). Physical-statistical modeling in geophysics. *Journal of Geophysical Research: Atmospheres*, 108:8776.

Bertozzi, A., Franco, E., Mohler, G., Short, M., and Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences*, 117:16732–16738.

Botella, C., Bonnet, P., Hui, C., Joly, A., and Richardson, D. M. (2022). Dynamic species distribution modeling reveals the pivotal role of human-mediated long-distance dispersal in plant invasion. *Biology*, 11:1293.

Boys, R. J., Wilkinson, D. J., and Kirkwood, T. B. L. (2008). Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18:125–135.

Bugallo, M. F., Martino, L., and Corander, J. (2015). Adaptive importance sampling in signal processing. *Digital Signal Processing*, 47:36–49.

Burnham, K. P., White, G. C., and Anderson, D. R. (1995). Model selection strategy in the analysis of capture-recapture data. *Biometrics*, 51:888–898.

Cendoya, M., Martínez-Minaya, J., Dalmau, V., Ferrer, A., Saponari, M., Conesa, D., López-Quílez, A., and Vicent, A. (2020). Spatial Bayesian modeling applied to the surveys of *Xylella fastidiosa* in Alicante (Spain) and Apulia (Italy). *Frontiers in plant science*, 11:1204.

Chapman, D. S., White, S. M., Hooftman, D. A., and Bullock, J. M. (2015). Inventory and review of quantitative models for spread of plant pests for use in pest risk assessment for the EU territory. *EFSA Supporting Publications*, 12:EN–795.

Cornuet, J., Marin, J.-M., Mira, A., and Robert, C. P. (2012). Adaptive multiple importance sampling. *Scandinavian Journal of Statistics*, 39:798–812.

Denancé, N., Cesbron, S., Briand, M., Rieux, A., and Jacques, M.-A. (2017a). Is *Xylella fastidiosa* really emerging in France? . In Costa, J. and Koebnik, R., editors, *1st Annual Conference of the EuroXanth - COST Action Integrating Science on Xanthomonadaceae for integrated plant disease management in Europe*, volume 7, Portugal. EuroXanth.

Denancé, N., Legendre, B., Briand, M., Olivier, V., Boisseson, C., Poliakoff, F., and Jacques, M.-A. (2017b). Several subspecies and sequence types are associated with the emergence of *Xylella fastidiosa* in natural settings in France. *Plant Pathology*, 66:1054–1064.

Dupas, E., Durand, K., Rieux, A., Briand, M., Pruvost, O., Cunty, A., Denancé, N., Donnadieu, C., Legendre, B., Lopez-Roques, C., Cesbron, S., Ravigné, V., and Jacques, M.-A. (2023). Suspicions of two bridgehead invasions of *xylella fastidiosa* subsp. *multiplex* in France. *Communications Biology*, 6:103.

ESV, P. (2022). Données de surveillance sur végétaux de *Xylella fastidiosa*. https://doi.org/10.15454/RWBIWD. Recherche Data Gouv.

Evans, L. C. (1998). *Partial differential equations*, volume 19 of *Graduate studies in mathematics*. American Mathematical Society, Rhode Island.

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O. G., and Lemey, P. (2014). The early spread and epidemic ignition of HIV-1 in human populations. *Science*, 346:56–61.

Fletcher, D. (2018). *Model Averaging*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ford, E. B. and Gregory, P. C. (2007). Bayesian model selection and extrasolar planet detection. In *Statistical Challenges in Modern Astronomy IV*, volume 371, page 189. ASP Conference Series.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, 56:501–514.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

Gelman, A., Roberts, G. O., Gilks, W. R., et al. (1996). Efficient Metropolis jumping rules. *Bayesian statistics*, 5:599–608.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88:881–889.

Godefroid, M., Cruaud, A., Streito, J.-C., Rasplus, J.-Y., and Rossi, J.-P. (2019). *Xylella fastidiosa*: climate suitability of European continent. *Scientific Reports*, 9:8844.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model-averaging: a tutorial. *Statistical Science*, 14:382–417.

Huld, T. A., Šúri, M., Dunlop, E. D., and Micale, F. (2006). Estimating average daytime and daily temperature profiles within Europe. *Environmental Modelling & Software*, 21:1650–1661.

Jeanmonod, D. and Natali, A. (1997). Les xénophytes de Corse: un danger pour la flore indigène. *Lagascalia*, 19:783–792.

Jones, D. R. and Baker, R. H. A. (2004). Introductions of non-native plant pathogens into Great Britain, 1970–2004. *Plant Pathology*, 56:891–910.

Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., and Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368:860–868.

Kottelenberg, D., Hemerik, L., Saponari, M., and Van Der Werf, W. (2021). Shape and rate of movement of the invasion front of *Xylella fastidiosa* spp. *pauca* in Puglia. *Scientific reports*, 11:1–14.

Lamon, E. C. and Clyde, M. A. (2000). Accounting for model uncertainty in prediction of chlorophyll a in lake okeechobee. *Journal of Agricultural, Biological, and Environmental Statistics*, 5:297–322.

Lanzarone, E., Pasquali, S., Gilioli, G., and Marchesini, E. (2017). A Bayesian estimation approach for the mortality in a stage-structured demographic model. *Journal of mathematical biology*, 75:759–779.

Leamer, E. (1978). *Specification searches: Ad hoc inference with nonexperimental data*, volume 53. John Wiley & Sons Incorporated.

Leitner, M. and Kühn, I. (2018). Dispersal in plants and animals. In Bunde, A., Caro, J., Kärger, J., and Vogl, G., editors, *Diffusive Spreading in Nature, Technology and Society*, pages 29–47, Cham. Springer International Publishing.

Louvrier, J., Papaix, J., Duchamp, C., and Gimenez, O. (2020). A mechanistic–statistical species distribution model to explain and forecast wolf (*Canis lupus*) colonization in South-Eastern France. *Spatial Statistics*, 36:100428.

Madigan, D., Gavrin, J., and Raftery, A. E. (1995). Enhancing the predictive performance of Bayesian graphical models. *Communications in Statistics-Theory and Methods*, 24:2271–2292.

Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89:1535–1546.

Marin, J.-M., Pudlo, P., and Sedki, M. (2019). Consistency of adaptive importance sampling and recycling schemes. *Bernoulli*, 25:1977 – 1998.

Martinetti, D. and Soubeyrand, S. (2019). Identifying lookouts for epidemio-surveillance: application to the emergence of Xylella fastidiosa in France. *Phytopathology*, 109:265–276.

Mason, S. J. (2004). On using "climatology" as a reference strategy in the Brier and ranked probability skill scores. *Monthly Weather Review*, 132:1891–1895.

McElreath, R. (2018). Overfitting, regularization, and information criteria. In *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society B*, 56:3–26.

Oehler, V. G., Yeung, K. Y., Choi, Y. E., Bumgarner, R. E., Raftery, A. E., and Radich, J. P. (2009). The derivation of diagnostic markers of chronic myeloid leukemia progression from microarray data. *Blood*, 114:3292–3298.

Okubo, A. and Levin, S. (2002). *Diffusion and ecological problems: Modern Perspectives*. Springer Science & Business Media, New York.

Ovaskainen, O., Rekola, H., Meyke, E., and Arjas, E. (2008). Bayesian methods for analyzing movements in heterogeneous landscapes from mark-recapture data. *Ecology*, 89:542–554.

Parkinson, D. and Liddle, A. R. (2013). Bayesian model averaging in astrophysics: a review. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 6:3–14.

Peterson, R. O., Vucetich, J. A., Page, R. E., Chouinard, A., et al. (2003). Temporal and spatial

aspects of predator–prey dynamics. *Alces*, 39:215–232.

Protter, MH and Weinberger, HF (1967). *Maximum Principles in Differential Equations*. Prentice-Hall, Englewood Cliffs, New Jersey.

Pyšek, P. and Hulme, P. E. (2005). Spatio-temporal dynamics of plant invasions: linking pattern to process. *Ecoscience*, 12:302–315.

Raffini, F., Bertorelle, G., Biello, R., D'Urso, G., Russo, D., and Bosso, L. (2020). From nucleotides to satellite imagery: approaches to identify and manage the invasive pathogen *Xylella fastidiosa* and its insect vectors in Europe. *Sustainability*, 12:4508.

Raftery, A. E. (1996). Approximate bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika*, 83:251–266.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133:1155–1174.

Roques, L. (2013). *Modèles de réaction-diffusion pour l'écologie spatiale*. Editions Quae, Paris.

Roques, L., Bonnefon, O., Baudrot, V., Soubeyrand, S., and Berestycki, H. (2020). A parsimonious approach for spatial transmission and heterogeneity in the COVID-19 propagation. *Royal Society Open Science*, 7:201382.

Roques, L. and Soubeyrand, S. (2023). Les invasions biologiques à la lumière des modèles. In Lannou, C., Rasplus, J.-Y., Soubeyrand, S., Gautier, M., and Rossi, J.-P., editors, *Crises sanitaires en agriculture: Les espèces invasives sous surveillance*, pages 167–185, Versailles. Editions Quae.

Roques, L., Soubeyrand, S., and Rousselet, J. (2011). A statistical-reaction–diffusion approach for analyzing expansion processes. *Journal of Theoretical Biology*, 274:43–51.

Roques, L., Walker, E., Franck, P., Soubeyrand, S., and Klein, E. (2016). Using genetic data to estimate diffusion rates in heterogeneous landscapes. *Journal of Mathematical Biology*, 73:397–422.

Rubin, D. B. and Schenker, N. (1986). Efficiently simulating the coverage properties of interval estimates. *Journal of the Royal Statistical Society C*, 35:159–167.

Schurr, F. M., Pagel, J., Cabral, J. S., Groeneveld, J., Bykova, O., O'Hara, R. B., Hartig, F., Kissling, W. D., Linder, H. P., Midgley, G. F., Schröder, B., Singer, A., and Zimmermann, N. E. (2012). How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography*, 39:2146–2162.

Shigesada, N., Kawasaki, K., and Takeda, Y. (1995). Modeling stratified diffusion in biological invasions. *The American Naturalist*, 146:229–251.

Sidman, A. H., Mak, M., and Lebo, M. J. (2008). Forecasting non-incumbent presidential elections: Lessons learned from the 2000 election. *International Journal of Forecasting*, 24:237 – 258.

Skellam, J. G. (1951). Random dispersal in theoretical populations. *Biometrika*, 38:196–218.

Soubeyrand, S. and Abboud, C. (2022). Data set: Average daily minimum temperature in January and February in Corsica. https://doi.org/10.5281/zenodo.6985148. Zenodo.

Soubeyrand, S., de Jerphanion, P., Martin, O., Saussac, M., Manceau, C., Hendrikx, P., and Lannou, C. (2018). What dynamics underly temporal observations? Application to the emergence of *Xylella fastidiosa* in France: probably not a recent story. *New Phytologist*, 219:824–836.

Soubeyrand, S. and Roques, L. (2014). Parameter estimation for reaction-diffusion models of biological invasions. *Population Ecology*, 56:427–434.

Turchin, P. (1998). *Quantitative Analysis of Movement: measuring and modeling population redistribution in plants and animals*. Sinauer, Sunderland, Massachusetts.

Viallefont, V., Raftery, A. E., and Richardson, S. (2001). Variable selection and bayesian model averaging in case-control studies. *Statistics in Medicine*, 20:3215–3230.

White, S. M., Bullock, J. M., Hooftman, D. A. P., and Chapman, D. S. (2017). Modelling the

spread and control of *Xylella fastidiosa* in the early stages of invasion in Apulia, Italy. *Biological Invasions*, 19:1825–1837.

White, S. M., Navas-Cortés, J. A., Bullock, J. M., Boscia, D., and Chapman, D. S. (2020). Estimating the epidemiology of emerging *Xylella fastidiosa* outbreaks in olives. *Plant Pathology*, 69:1403–1413.

Wikle, C. K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84:1382–1394.

Williams, P. J., Hooten, M. B., Womble, J. N., Esslinger, G. G., and Bower, M. R. (2018). Monitoring dynamic spatio-temporal ecological processes optimally. *Ecology*, 99:524–535.

Wintle, B. A., McCarthy, M. A., Volinsky, C. T., and Kavanagh, R. P. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, 17:1579–1590.

Yeung, K. Y., Bumgarner, R. E., and Raftery, A. E. (2005). Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics*, 21:2394–2402.

Table 1: Posterior model probabilities obtained for each simulated data set using different methods to assess the integrated likelihood, namely, those based on BIC, DIC_1 and DIC_2, UWE, HME_1, HME_2, $WAIC_1$, $WAIC_2$ and IC.

Method	Model	Diffusion	Posterior model probability	
			$\mathcal{M}_0(5.5)$	$\mathcal{M}_1(5.5)$
BIC	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	1.00	0.00
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	1.00	0.00
		$D_2 = 0.1D_1$	1.00	0.00
DIC_1	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	1.00	0.00
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	<0.01	>0.99
		$D_2 = 0.1D_1$	1.00	0.00
DIC_2	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	1.00	0.00
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	0.57	0.43
		$D_2 = 0.1D_1$	0.30	0.70
UWE	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	0.58	0.42
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	0.23	0.77
		$D_2 = 0.1D_1$	0.15	0.85
HME_1	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	1.00	0.00
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	0.80	0.20
		$D_2 = 0.1D_1$	0.23	0.77
HME_2	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	1.00	0.00
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	0.12	0.88
		$D_2 = 0.1D_1$	0.08	0.92
$WAIC_1$	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	1.00	0.00
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	0.40	0.60
		$D_2 = 0.1D_1$	0.41	0.59
$WAIC_2$	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	1.00	0.00
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	0.40	0.60
		$D_2 = 0.1D_1$	0.40	0.60
IC	$\mathcal{M}_0(5.5)$	$D_2 = D_1$	1.00	0.00
	$\mathcal{M}_1(5.5)$	$D_2 = 0.9D_1$	0.00	1.00
		$D_2 = 0.1D_1$	1.00	0.00
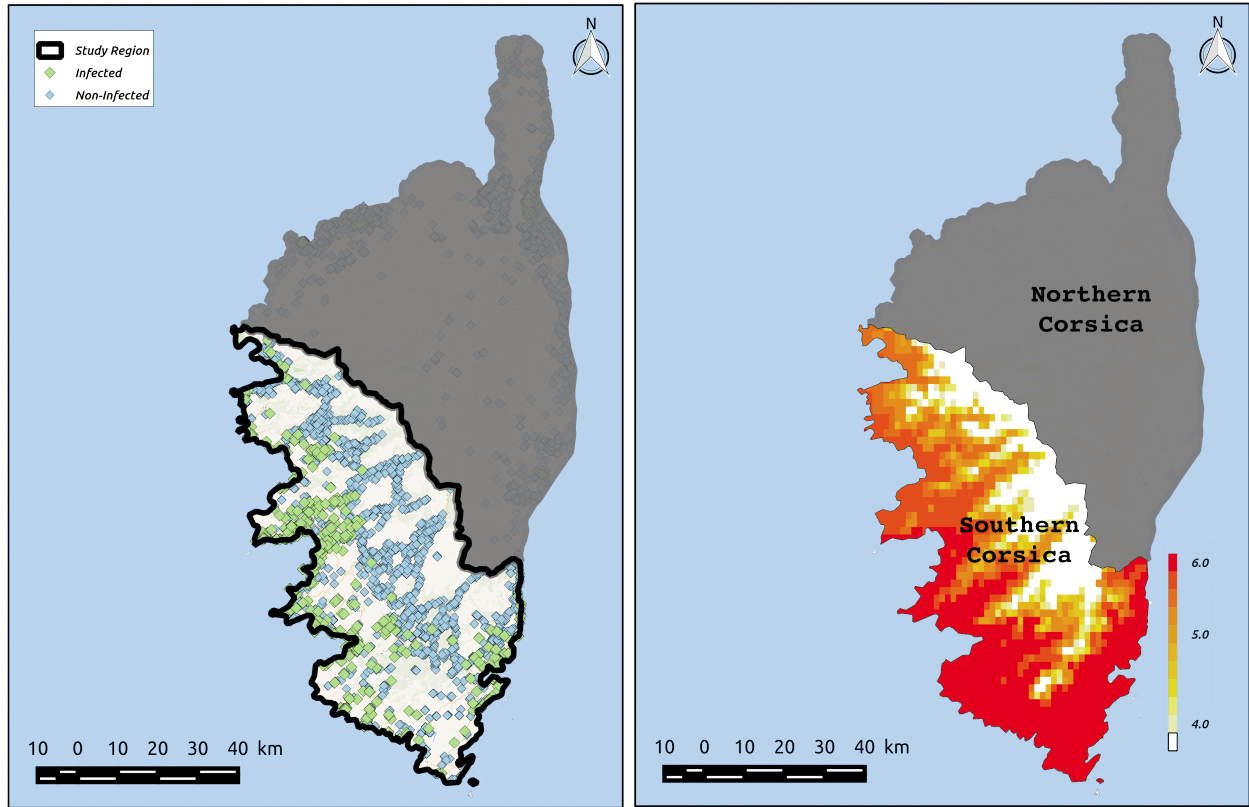
Figure 1: Locations of plants (left), sampled from July 2015 to January 2019, that have been detected as positive (green dots) or negative (blue dots) to Xf in South Corsica, France, and map of the average of the daily minimum temperature (right) in Celsius degrees over January and February and a 1×1km-pixel grid.
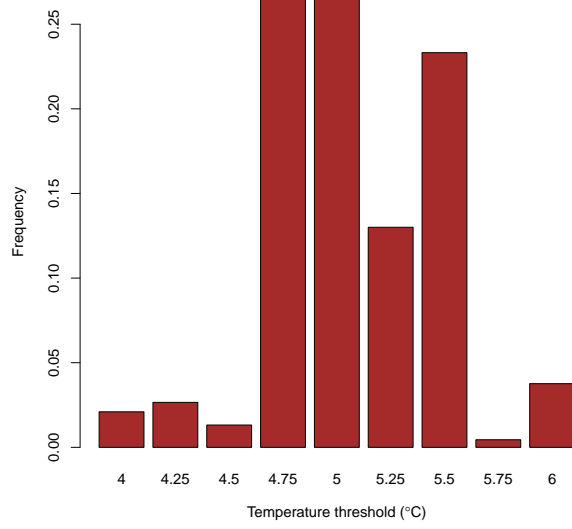
Figure 2: BMA marginal posterior distribution of the temperature threshold $\tilde{T}$.
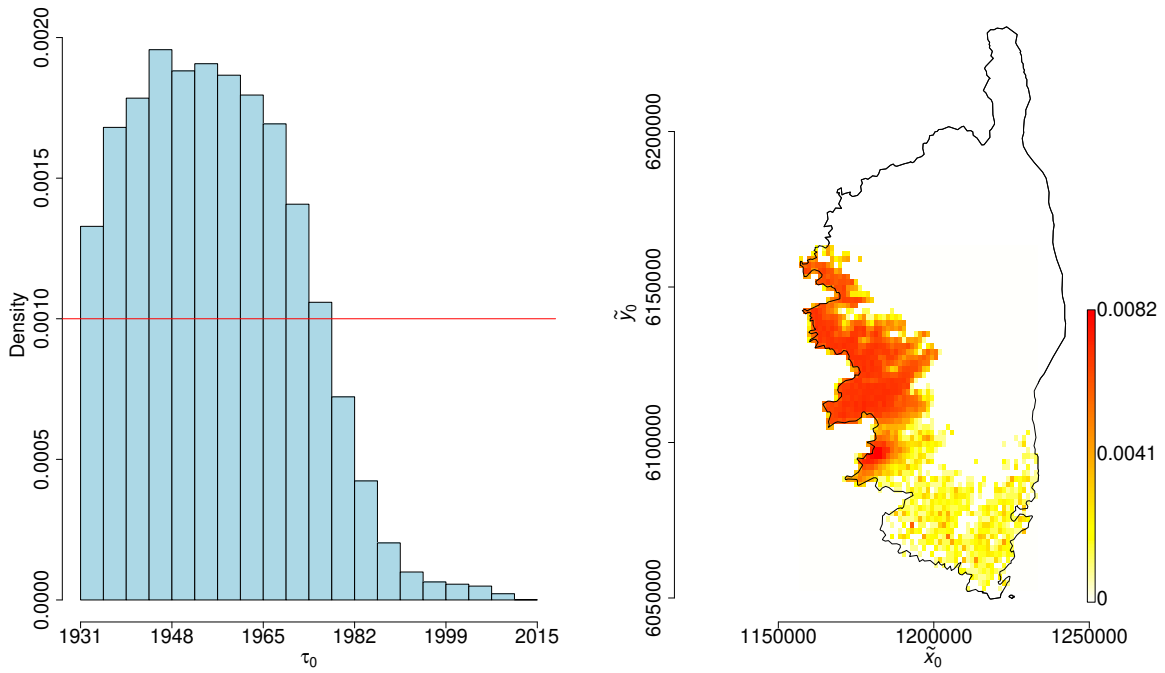
Figure 3: BMA Posterior distributions of the introduction time τ_0 (histogram) and the introduction point $\tilde{\mathbf{x}}_0$ (color palette). The prior for τ_0 was uniform over $[1931, 2015]$ (red line). The prior for $\tilde{\mathbf{x}}_0$ was uniform over $\Omega_{\tilde{T}1}$ for model $\mathcal{M}_i(\tilde{T})$, $0 \leq i \leq I$.
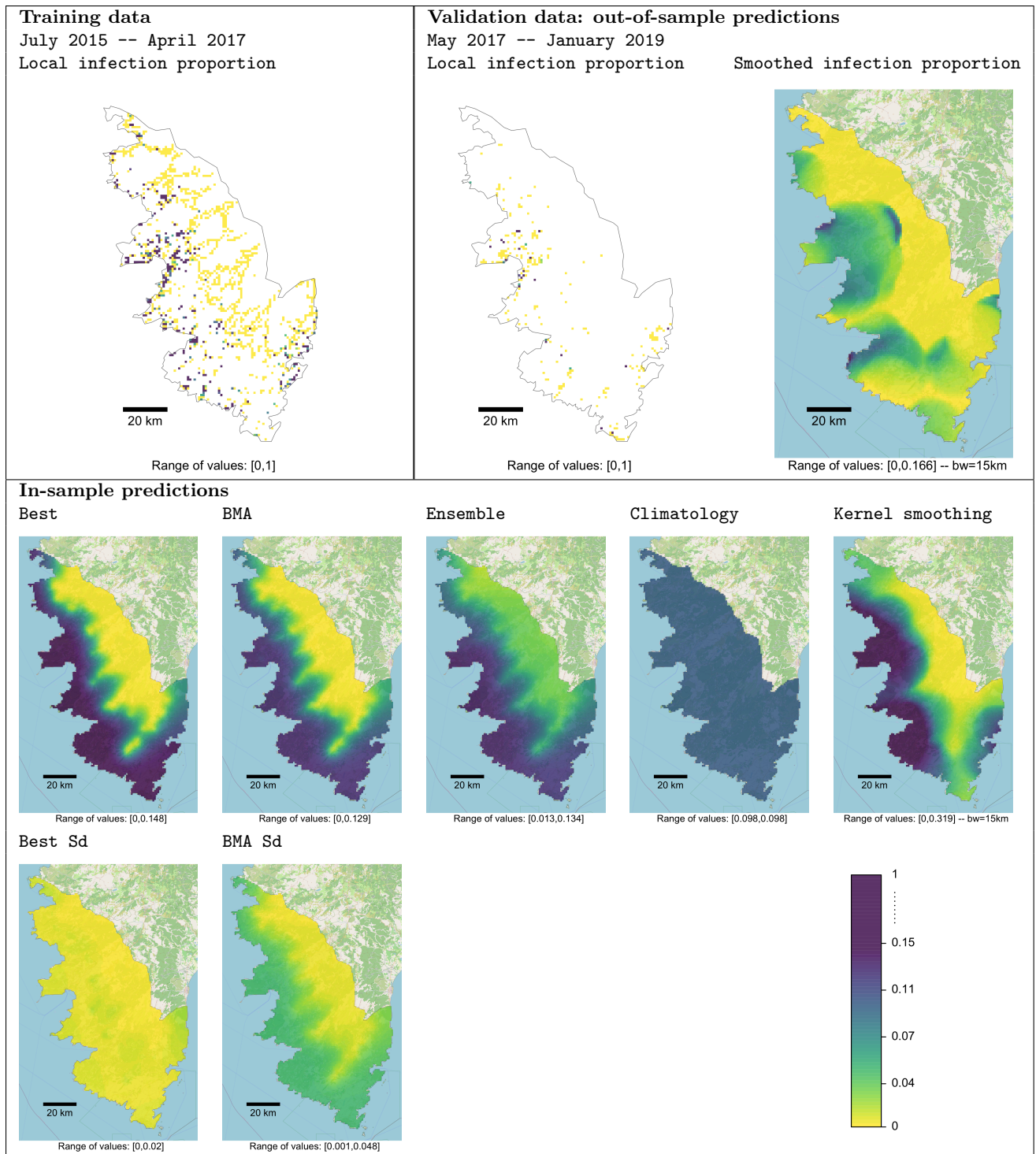
Figure 4: Training data (top left box), out-of-sample predictions computed either as a local infection proportion or as a kernel smoothing with a 15km-bandwidth (top right box), and in-sample predictions (i.e. posterior means provided by the best model and the BMA, ensemble, climatology and kernel smoothing with a bandwidth of 15 km; bottom box) obtained from training data. Map of standard deviations are also provided in the bottom box for the best model and BMA. The color palette (bottom right) applies to all plots. The range of values taken in each map is provided below the map.
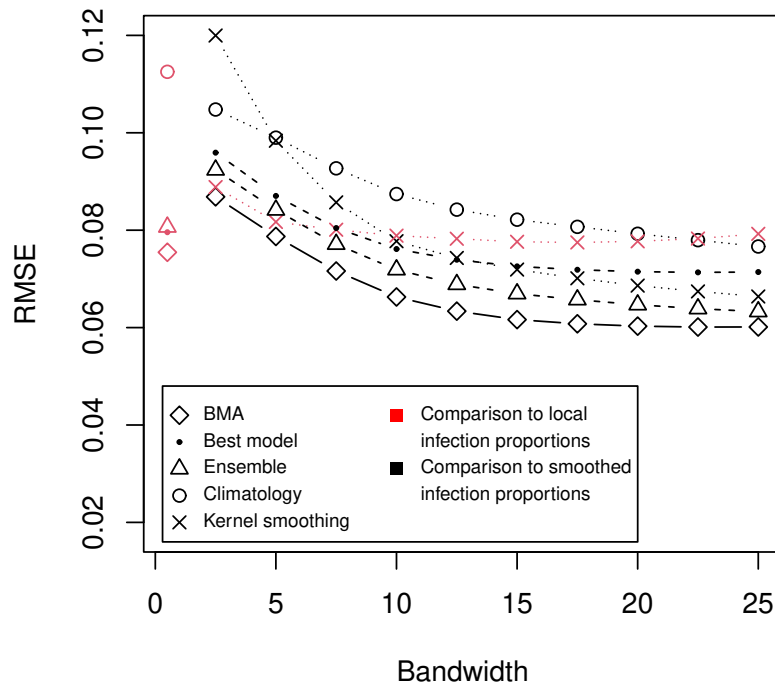
Figure 5: RMSE values showing the performance of the in-sample predictors (BMA; best model; ensemble; climatology; kernel smoothing with varying bandwidth) with respect to each of the out-of-sample predictions (local infection proportions in red; smoothed infection proportions with varying bandwidth in black). Each in-sample kernel smoother obtained with a given bandwidth value is compared both to the local infection proportions (red crosses) and to the smoothed infection proportion using the same bandwidth (black crosses), but not to the smoothed infection proportions using other bandwidth values.

Forecasting pathogen dynamics with Bayesian model averaging: Application to *Xylella fastidiosa*

—

Online Resource 1

March 5, 2023

Candy Abboud[1,2,*], Eric Parent[3], Olivier Bonnefon[2], Samuel Soubeyrand[2,**]

[1] Current affiliation: Engineering and Technology, American University of the Middle East, Kuwait

[2] INRAE, BioSP, 84914 Avignon, France

[3] AgroParisTech, INRAE, UMR 518 Math. Info. Appli., Paris, France

[*] Corresponding author: candy.abboud@aum.edu.kw

[**] Secondary corresponding author: samuel.soubeyrand@inrae.fr

S1 Estimators of the integrated likelihood

Estimator grounded on un-normalized weights. An interesting benefit of importance sampling schemes is the relatively direct assessment of the integrated likelihood, computed as the average of un-normalized weights (see Bugallo et al., 2015). This unbiased estimator, say UWE which stands for un-normalized weight estimator, satisfies for model $\mathcal{M}_i(\tilde{T})$:

$$[\widehat{Y|\mathcal{M}_i(\tilde{T})}] = \frac{1}{ML} \sum_{m=1}^{M} \sum_{l=1}^{L} \tilde{w}_m^l, \tag{1}$$

where the un-normalized weights $\tilde{w}_m^l$ are computed at the last AMIS iteration using Equations (5) and (6) of Abboud et al. (2019).

Estimators grounded on information criteria. Information criteria are generally used for model comparison but can also be exploited to assess the integrated likelihood in BMA as presented in McElreath (2018). Thus, the WAIC (Watanabe–Akaike information criterion; Watanabe, 2010) has been used by McElreath (2018) to approach the integrated likelihood $[Y|\mathcal{M}_i(\tilde{T})]$ with the so-called

Akaike weight:

$$[Y|\widehat{\mathcal{M}_i(\tilde{T})}] = \frac{\exp(-\frac{1}{2}\mathrm{dWAIC}_{i\tilde{T}})}{\displaystyle\sum_{0 \le i \le I, \tilde{T} \in \mathcal{T}} \exp(-\frac{1}{2}d\mathrm{WAIC}_{i\tilde{T}})}, \tag{2}$$

where $\mathrm{dWAIC}_{i\tilde{T}} = \mathrm{WAIC}_{i\tilde{T}} - \min_{i,\tilde{T}}\{\mathrm{WAIC}_{i\tilde{T}}\}$. If prior probabilities of models are the same, then Akaike weights and posterior model probabilities coincide.

McElreath (2018) also suggested to replace WAIC by other information criteria in (2) and we will test this proposal by using BIC, DIC and IC. Below, we provide the expressions of all the criteria taken into consideration and show how they are computed in practice.

The WAIC defined in the deviance scale satisfies for model $\mathcal{M}_i(\tilde{T})$:

$$\mathrm{WAIC}_{i\tilde{T}} = -2(\mathrm{lppd}_{i\tilde{T}} - \mathrm{p}_{\mathrm{WAIC}_{i\tilde{T}}}),$$

where $\mathrm{lppd}_{i\tilde{T}}$ is the log pointwise predictive density and $p_{\mathrm{WAIC}_{i\tilde{T}}}$ is an estimate of the effective number of free parameters in the model. The term $\mathrm{lppd}_{i\tilde{T}}$ is the logarithm of the predictive density integrated over the posterior distribution of the model parameters summed over all observations: $\mathrm{lppd}_{i\tilde{T}} = \sum_{j=1}^{J} \log \int [Y_j|\Theta, \mathcal{M}_i(\tilde{T})] \times [\Theta|Y, \mathcal{M}_i(\tilde{T})]d\Theta$. In practice, $\mathrm{lppd}_{i\tilde{T}}$ is replaced in the WAIC formula by its empirical analogue (see Gelman et al., 2014) using the weighted posterior sample: $\sum_{j=1}^{J} \log \left(\sum_{m=1}^{M} \sum_{l=1}^{L} [Y_j|\Theta_m^l, \mathcal{M}_i(\tilde{T})]w_m^l \right)$. Following the initial proposal by Watanabe (2010), the term $p_{\mathrm{WAIC}_{i\tilde{T}}}$ can be expressed as:

$$\mathrm{p}_{\mathrm{WAIC}_{i\tilde{T}}} = 2 \sum_{j=1}^{J} \left(\log \left(\mathbb{E}\left[Y_j|\Theta, \mathcal{M}_i(\tilde{T}) \right] \right) - \mathbb{E}\left(\log \left[Y_j|\Theta, \mathcal{M}_i(\tilde{T}) \right] \right) \right),$$

but we will also consider the following expression advocated by Gelman et al. (2014):

$$\mathrm{p}_{\mathrm{WAIC}_{i\tilde{T}}} = \sum_{j=1}^{J} \mathbb{V}\left(\log[Y_j|\Theta, \mathcal{M}_i(\tilde{T})] \right).$$

Thereafter, WAIC_1 and WAIC_2 denote the versions of WAIC based on the former and the latter expressions of $\mathrm{p}_{\mathrm{WAIC}_{i\tilde{T}}}$, respectively. In practice, expectations and variances arising in the two versions of $\mathrm{p}_{\mathrm{WAIC}_{i\tilde{T}}}$ are replaced by their empirical values using the weighted posterior sample $\{(\Delta_m^l, w_m^l) : 1 \le m \le M, 1 \le l \le L\}$ provided by the AMIS algorithm applied to model $\mathcal{M}_i(\tilde{T})$.

The BIC (Bayesian information criterion; Schwarz et al., 1978) satisfies for model $\mathcal{M}_i(\tilde{T})$: $\mathrm{BIC}_{i\tilde{T}} = -2\log[Y|\hat{\Theta}, \mathcal{M}_i(\tilde{T})] + \xi_{i\tilde{T}} \log J$, where J is the sample size, $\xi_{i\tilde{T}}$ is the number of model parameters, and $\hat{\Theta}$ is the maximum likelihood estimate of the parameter vector Θ of model $\mathcal{M}_i(\tilde{T})$ in the support $\mathcal{S}_{i\tilde{T}}$ of Θ: $\hat{\Theta} = \underset{\Theta \in \mathcal{S}_{i\tilde{T}}}{\mathrm{argmax}}[Y|\Theta, \mathcal{M}_i(\tilde{T})]$.

The DIC (deviance information criterion; Gelman et al., 2014) satisfies for model $\mathcal{M}_i(\tilde{T})$: $\mathrm{DIC}_{i\tilde{T}} =$

2

$\bar{\mathcal{D}} + p_{\text{eff}}$, where $\bar{\mathcal{D}}$ is the posterior mean of the deviance $\mathcal{D}(\Theta) = -2\log[Y|\Theta, \mathcal{M}_i(\tilde{T})] + C$ (C is a constant that cancels out when one compares different models) and p_{eff} is the effective number of parameters of the model. The difference in the two versions of the DIC considered here lies in the calculation of p_{eff}. In the first version proposed by Spiegelhalter et al. (2002), $p_{\text{eff}} = p_{\mathcal{D}} = \bar{\mathcal{D}} - \mathcal{D}(\bar{\Theta})$, where $\bar{\Theta}$ is the posterior mean of Θ under the model $\mathcal{M}_i(\tilde{T})$: $\bar{\Theta} = \mathbb{E}[\Theta|Y, \mathcal{M}_i(\tilde{T})]$. In the second version proposed by Gelman et al. (2003), $p_{\text{eff}} = \frac{1}{2}\mathbb{V}(\mathcal{D}(\Theta)|Y, \mathcal{M}_i(\tilde{T}))$, where $\mathbb{V}(\mathcal{D}(\Theta)|Y, \mathcal{M}_i(\tilde{T}))$ is the posterior variance of $\mathcal{D}(\Theta)$. Let DIC_1 and DIC_2 denote the versions of DIC based on the former and the latter expressions of p_{eff}, respectively.

The IC (information criterion) of Ando (2011), which is supposed to solve over-fitting issues in comparison with DIC, satisfies: $\text{IC}_{i\tilde{T}} = \bar{\mathcal{D}} + 2p_{\mathcal{D}} := 3\bar{\mathcal{D}} - 2\mathcal{D}(\bar{\Theta})$.

In practice, the different terms appearing in BIC, DIC and IC, namely $\hat{\Theta}$, $\bar{\Theta}$, $\bar{\mathcal{D}}$ and $\mathbb{V}(\mathcal{D}(\Theta)|Y, \mathcal{M}_i(\tilde{T}))$, are replaced by their empirical values using the weighted posterior sample $\{(\Delta_m^l, w_m^l) : 1 \le m \le M, 1 \le l \le L\}$ provided by AMIS applied to model $\mathcal{M}_i(\tilde{T})$.

Harmonic mean estimators. For approaching the integrated likelihood, Newton and Raftery (1994) noticed, from Bayes theorem,

$$[Y|\mathcal{M}_i(\tilde{T})]^{-1} = \int [\Theta|Y, \mathcal{M}_i(\tilde{T})][Y|\Theta, \mathcal{M}_i(\tilde{T})]^{-1}d\Theta = \mathbb{E}([Y|\Theta, \mathcal{M}_i(\tilde{T})]^{-1}|Y),$$

and proposed the sample harmonic mean of the likelihood as an estimator of $[Y|\mathcal{M}_i(\tilde{T})]$. Thus, using the weighted posterior sample $\{(\Delta_m^l, w_m^l) : 1 \le m \le M, 1 \le l \le L\}$, the harmonic mean estimator, say HME_1, satisfies:

$$[\widehat{Y|\mathcal{M}_i(\tilde{T})}] = \left(\sum_{m=1}^{M} \sum_{l=1}^{L} \frac{1}{[Y|\Theta_m^l, \mathcal{M}_i(\tilde{T})]} w_m^l \right)^{-1}.$$

This estimator is easily computed and is consistent if the sample size tends to infinity but is rather unstable in practice and in theory (the variance of the weights is not finite).

Gelfand and Dey (1994) proposed the following generalized version, say HME_2, of HME_1:

$$[\widehat{Y|\mathcal{M}_i(\tilde{T})}] = \left(\sum_{m=1}^{M} \sum_{l=1}^{L} \frac{f(\Theta_m^l)}{[Y|\Theta_m^l, \mathcal{M}_i(\tilde{T})] \times [\Theta_m^l|\mathcal{M}_i(\tilde{T})]} w_m^l \right)^{-1}, \quad (3)$$

where $f(.)$ is an importance probability density function. HME_2 is an unbiased and consistent estimator of the integrated likelihood if $\int f(\Theta)[Y|\Theta, \mathcal{M}_i(\tilde{T})]^{-1}[\Theta|\mathcal{M}_i(\tilde{T})]^{-1}d\Theta < \infty$. We choose for f an approximation of the posterior distribution of Θ, namely the multivariate normal distribution with mean vector and covariance matrix estimated from the weighted posterior sample $\{(\Theta_m^l, w_m^l) : 1 \le m \le M, 1 \le l \le L\}$. This is supposed to give a stable version of the harmonic mean estimator (Raftery et al., 2006).

S2 Specification of prior distributions

The prior distributions $[\Theta_{i\tilde{T}}|\mathcal{M}_i(\tilde{T})]$ combine vague uniform and Dirac distributions. Dirac distributions are considered for r_0 and p_0, which are related to the initial conditions. These parameters are kept fixed for identifiability issues.

$$
\begin{aligned}
[\Theta_{0\tilde{T}}|\mathcal{M}_0(\tilde{T})] =& \frac{1}{(10^8 - 50) \times 100 \times 1 \times 100 \times 1000 \times |\Omega_{\tilde{T}1}|} \\
& \times \mathbb{1}(D \in [50; 10^8], b \in [0; 100], K \in]0; 1], \alpha \in [0; 100], \tau_0 \in [-1000; 0], \tilde{\mathbf{x}}_0 \in \Omega_{\tilde{T}1}) \\
& \times \mathrm{Dirac}_{5000}(r_0) \times \mathrm{Dirac}_{0.1}(p_0). \\
[\Theta_{1\tilde{T}}|\mathcal{M}_1(\tilde{T})] =& \frac{1}{(10^8 - 100) \times (10^4 - 10) \times 100 \times 1 \times 100 \times 1000 \times |\Omega_{\tilde{T}1}|} \\
& \times \mathbb{1}(D_{1\tilde{T}1} \in [100; 10^8], D_{1\tilde{T}2} \in [10; 10^4], b \in [0; 100], K \in]0; 1], \alpha \in [0; 100], \\
& \quad \tau_0 \in [-1000; 0], \tilde{\mathbf{x}}_0 \in \Omega_{\tilde{T}1}) \\
& \times \mathrm{Dirac}_{5000}(r_0) \times \mathrm{Dirac}_{0.1}(p_0). \\
[\Theta_{2\tilde{T}}|\mathcal{M}_2(\tilde{T})] =& \frac{1}{(10^8 - 50) \times (10^5 - 50) \times 100 \times 1 \times 1000 \times |\Omega_{\tilde{T}1}|} \\
& \times \mathbb{1}(D_{2\tilde{T}1} \in [50; 10^8], D_{2\tilde{T}2} \in [50; 10^5], b \in [0; 100], K \in]0; 1], \tau_0 \in [-1000; 0], \tilde{\mathbf{x}}_0 \in \Omega_{\tilde{T}1}) \\
& \times \mathrm{Dirac}_{5000}(r_0) \times \mathrm{Dirac}_{0.1}(p_0).
\end{aligned}
$$

The prior for τ_0 ranging from -1000 to 0 corresponds to an introduction a priori happening between 1931 and 2015, which includes available estimates of the introduction date of the detected strains of *Xylella fastidiosa* in Corsica, France: namely 1985 or 2001 for Soubeyrand et al. (2018) depending on the model used, 1959 for Abboud et al. (2019) and 1971 or 1987 for Dupas et al. (2022) depending on the considered strains.

References

Abboud, C., Bonnefon, O., Parent, E., and Soubeyrand, S. (2019). Dating and localizing an invasion from post-introduction data and a coupled reaction–diffusion–absorption model. *Journal of Mathematical Biology*, 79:765–789.

Ando, T. (2011). Predictive Bayesian model selection. *American Journal of Mathematical and Management Sciences*, 31:13–38.

Bugallo, M. F., Martino, L., and Corander, J. (2015). Adaptive importance sampling in signal processing. *Digital Signal Processing*, 47:36–49.

Dupas, E., Durand, K., Rieux, A., Briand, M., Pruvost, O., Cunty, A., Denancé, N., Donnadieu,

C., Legendre, B., Lopez-Roques, C., Cesbron, S., Ravigné, V., and Jacques, M.-A. (2022). Two bridgehead invasions of Xylella fastidiosa subsp. multiplex in France. *bioRxiv*, 480026:1–68.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, 56:501–514.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian data analysis.* Texts in statistical science series. Chapman & Hall/CRC, New York, 2 edition.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24:997–1016.

McElreath, R. (2018). Overfitting, regularization, and information criteria. In *Statistical rethinking: A Bayesian course with examples in R and Stan.* Chapman and Hall/CRC.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society B*, 56:3–26.

Raftery, A. E., Newton, M. A., Satagopan, J. M., and Krivitsky, P. N. (2006). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian Statistics 8*, page 1–45.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6:461–464.

Soubeyrand, S., de Jerphanion, P., Martin, O., Saussac, M., Manceau, C., Hendrikx, P., and Lannou, C. (2018). What dynamics underly temporal observations? Application to the emergence of *Xylella fastidiosa* in France: probably not a recent story. *New Phytologist*, 219:824–836.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64:583–639.

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.

Table S1: Specification of models $\mathcal{M}_0(5.5)$ and $\mathcal{M}_1(5.5)$ from which data sets were generated for the simulation study.

Dataset	Model	Diffusion	Parameter values
$\mathcal{O}_1$	$\mathcal{M}_0(5.5)$	$D_1 = D_2 = D$	$D = 3.2\mathrm{e}{+}05$, $b = 0.049$, $K = 0.15$,
			$\alpha = 0.26$, $\tilde{\mathbf{x}}_0 = (1.176023\mathrm{e}{+}06, 6.108375\mathrm{e}{+}06)$, $\tau_0 = -335$
$\mathcal{O}_2$	$\mathcal{M}_1(5.5)$	$D_1 = 0.9D_2$	$D_1 = 2.0\mathrm{e}{+}05$, $D_2 = 1.8\mathrm{e}{+}05$, $b = 0.025$, $K = 0.14$,
			$\alpha = 0.50$, $\tilde{\mathbf{x}}_0 = (1.178962\mathrm{e}{+}06, 6.113653\mathrm{e}{+}06)$, $\tau_0 = -650$
$\mathcal{O}_3$	$\mathcal{M}_1(5.5)$	$D_1 = 0.1D_2$	$D_1 = 2.0\mathrm{e}{+}05$, $D_2 = 2.0\mathrm{e}{+}04$, $b = 0.025$, $K = 0.14$,
			$\alpha = 0.50$, $\tilde{\mathbf{x}}_0 = (1.178962\mathrm{e}{+}06, 6.113653\mathrm{e}{+}06)$, $\tau_0 = -650$

Table S2: Posterior model probabilities computed from BIC, DIC$_1$, DIC$_2$, IC, WAIC$_1$, WAIC$_2$, HME$_1$, HME$_2$ and UWE. The highest posterior model probability value obtained with each method is highlighted in yellow.

|					Posterior model probability				
$\mathcal{M}_i(\tilde{T})$		BIC	DIC$_1$	DIC$_2$	IC	WAIC$_1$	WAIC$_2$	HME$_1$	HME$_2$	UWE
$\tilde{T} = 4$ ℃										
	$\mathcal{M}_0$	1.083×10^{-5}	1.760×10^{-30}	8.622×10^{-7}	1.445×10^{-58}	2.362×10^{-7}	2.387×10^{-7}	5.748×10^{-6}	2.441×10^{-5}	8.677×10^{-3}
	$\mathcal{M}_1$	1.079×10^{-16}	1.254×10^{-38}	1.595×10^{-14}	2.002×10^{-67}	1.621×10^{-14}	1.633×10^{-14}	1.130×10^{-14}	3.086×10^{-16}	4.962×10^{-3}
	$\mathcal{M}_2$	4.712×10^{-65}	3.268×10^{-89}	6.187×10^{-68}	3.075×10^{-116}	4.839×10^{-67}	4.914×10^{-67}	7.509×10^{-68}	7.197×10^{-66}	6.877×10^{-3}
$\tilde{T} = 4.25$ ℃										
	$\mathcal{M}_0$	9.232×10^{-6}	2.779×10^{-29}	2.042×10^{-6}	8.857×10^{-57}	8.006×10^{-7}	7.710×10^{-7}	2.935×10^{-5}	1.459×10^{-4}	1.972×10^{-2}
	$\mathcal{M}_1$	2.248×10^{-13}	1.559×10^{-34}	1.212×10^{-11}	7.220×10^{-62}	2.625×10^{-12}	2.573×10^{-12}	3.256×10^{-12}	4.595×10^{-14}	1.471×10^{-3}
	$\mathcal{M}_2$	3.443×10^{-64}	2.739×10^{-87}	2.054×10^{-66}	2.181×10^{-113}	3.939×10^{-66}	3.920×10^{-66}	7.323×10^{-66}	1.580×10^{-66}	4.406×10^{-3}
$\tilde{T} = 4.5$ ℃										
	$\mathcal{M}_0$	4.855×10^{-4}	2.622×10^{-29}	7.222×10^{-6}	1.601×10^{-57}	5.107×10^{-6}	4.540×10^{-6}	6.383×10^{-5}	1.662×10^{-6}	2.283×10^{-3}
	$\mathcal{M}_1$	3.110×10^{-13}	4.242×10^{-36}	2.492×10^{-11}	4.826×10^{-65}	4.158×10^{-12}	4.530×10^{-12}	3.900×10^{-11}	5.254×10^{-13}	2.555×10^{-3}
	$\mathcal{M}_2$	2.930×10^{-65}	1.515×10^{-90}	2.405×10^{-66}	1.325×10^{-119}	1.878×10^{-66}	1.904×10^{-66}	4.550×10^{-65}	1.784×10^{-66}	8.078×10^{-3}
$\tilde{T} = 4.75$ ℃										
	$\mathcal{M}_0$	1.013×10^{-2}	2.224×10^{-22}	4.685×10^{-4}	1.292×10^{-45}	3.864×10^{-4}	3.444×10^{-4}	2.123×10^{-2}	9.355×10^{-1}	2.289×10^{-1}
	$\mathcal{M}_1$	9.615×10^{-8}	8.132×10^{-28}	2.933×10^{-8}	3.439×10^{-52}	8.272×10^{-9}	7.186×10^{-9}	6.346×10^{-7}	2.328×10^{-4}	2.566×10^{-2}
	$\mathcal{M}_2$	1.033×10^{-64}	1.770×10^{-90}	2.616×10^{-66}	1.752×10^{-119}	2.250×10^{-66}	2.265×10^{-66}	4.228×10^{-65}	7.768×10^{-66}	9.172×10^{-3}
$\tilde{T} = 5$ ℃										
	$\mathcal{M}_0$	5.570×10^{-4}	1.435×10^{-22}	8.449×10^{-5}	2.739×10^{-45}	5.263×10^{-5}	4.739×10^{-5}	1.710×10^{-3}	2.444×10^{-2}	1.515×10^{-1}
	$\mathcal{M}_1$	2.295×10^{-9}	1.004×10^{-24}	1.322×10^{-7}	1.312×10^{-46}	3.448×10^{-8}	3.575×10^{-8}	1.558×10^{-5}	1.130×10^{-4}	8.080×10^{-2}
	$\mathcal{M}_2$	1.458×10^{-60}	4.828×10^{-85}	9.892×10^{-63}	2.749×10^{-112}	1.269×10^{-62}	1.254×10^{-62}	1.340×10^{-61}	2.580×10^{-61}	2.888×10^{-2}
$\tilde{T} = 5.25$ ℃										
	$\mathcal{M}_0$	2.511×10^{-15}	4.706×10^{-43}	8.435×10^{-17}	9.856×10^{-75}	6.452×10^{-16}	6.278×10^{-16}	6.742×10^{-15}	5.238×10^{-13}	8.780×10^{-3}
	$\mathcal{M}_1$	5.081×10^{-10}	9.268×10^{-27}	7.848×10^{-8}	2.036×10^{-50}	4.015×10^{-8}	3.769×10^{-8}	3.368×10^{-7}	7.808×10^{-6}	1.151×10^{-1}
	$\mathcal{M}_2$	1.160×10^{-65}	4.931×10^{-91}	4.431×10^{-67}	2.834×10^{-120}	4.676×10^{-67}	4.598×10^{-67}	2.839×10^{-66}	1.708×10^{-66}	1.118×10^{-2}
$\tilde{T} = 5.5$ ℃										
	$\mathcal{M}_0$	2.890×10^{-3}	2.454×10^{-23}	3.552×10^{-3}	3.623×10^{-48}	1.490×10^{-3}	1.363×10^{-3}	2.271×10^{-1}	3.925×10^{-2}	1.249×10^{-1}
	$\mathcal{M}_1$	4.226×10^{-7}	1.531×10^{-24}	8.348×10^{-6}	5.305×10^{-48}	3.386×10^{-6}	3.047×10^{-6}	2.084×10^{-4}	2.427×10^{-4}	8.566×10^{-2}
	$\mathcal{M}_2$	1.932×10^{-64}	2.882×10^{-90}	9.144×10^{-67}	5.855×10^{-119}	1.302×10^{-66}	1.261×10^{-66}	8.925×10^{-66}	1.189×10^{-65}	1.514×10^{-2}
$\tilde{T} = 5.75$ ℃										
	$\mathcal{M}_0$	1.181×10^{-3}	1.000×10^{0}	6.670×10^{-9}	1.000×10^{0}	2.297×10^{-5}	2.408×10^{-5}	2.246×10^{-6}	3.055×10^{-10}	1.206×10^{-3}
	$\mathcal{M}_1$	9.847×10^{-1}	4.901×10^{-49}	9.959×10^{-1}	2.808×10^{-78}	9.980×10^{-1}	9.982×10^{-1}	7.497×10^{-1}	7.811×10^{-10}	1.413×10^{-3}
	$\mathcal{M}_2$	7.952×10^{-79}	1.269×10^{-103}	7.912×10^{-79}	2.658×10^{-132}	2.873×10^{-79}	2.367×10^{-79}	3.484×10^{-78}	2.800×10^{-73}	1.665×10^{-2}
$\tilde{T} = 6$ ℃										
	$\mathcal{M}_0$	3.195×10^{-75}	4.870×10^{-101}	2.194×10^{-75}	1.289×10^{-130}	4.840×10^{-76}	4.608×10^{-76}	1.386×10^{-76}	2.468×10^{-82}	2.067×10^{-3}
	$\mathcal{M}_1$	1.918×10^{-71}	3.273×10^{-95}	6.284×10^{-70}	1.334×10^{-124}	9.743×10^{-71}	8.740×10^{-71}	5.464×10^{-71}	1.738×10^{-77}	1.288×10^{-3}
	$\mathcal{M}_2$	1.204×10^{-64}	1.442×10^{-90}	9.837×10^{-67}	1.666×10^{-119}	1.638×10^{-66}	1.643×10^{-66}	6.166×10^{-66}	2.084×10^{-65}	3.267×10^{-2}
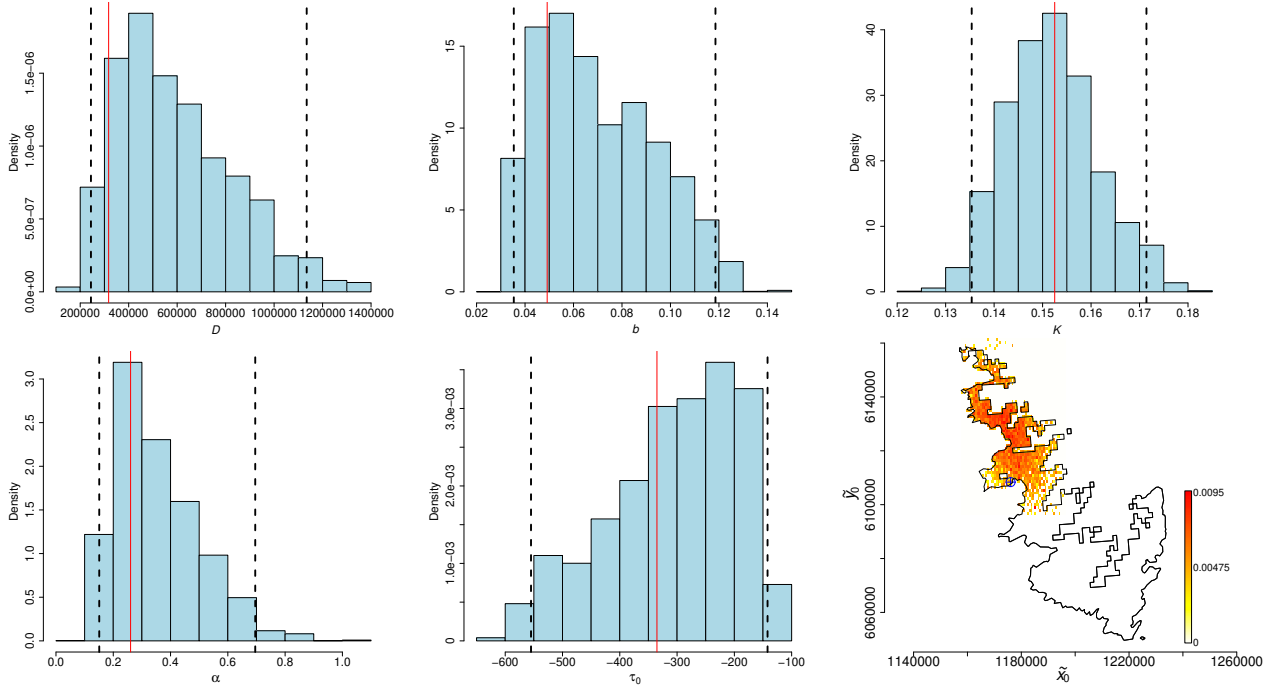
Figure S1: Posterior map of the introduction location and histograms of marginal posterior distributions for all the other parameters of model $\mathcal{M}_0(5.5)$ when it is fitted to data $\mathcal{O}_1$ in the simulation study. Red line: true value. Dashed black lines 0.025 and 0.975 posterior quantiles. Blue cross on the map: true introduction location. Color palette: posterior probability of the introduction location.
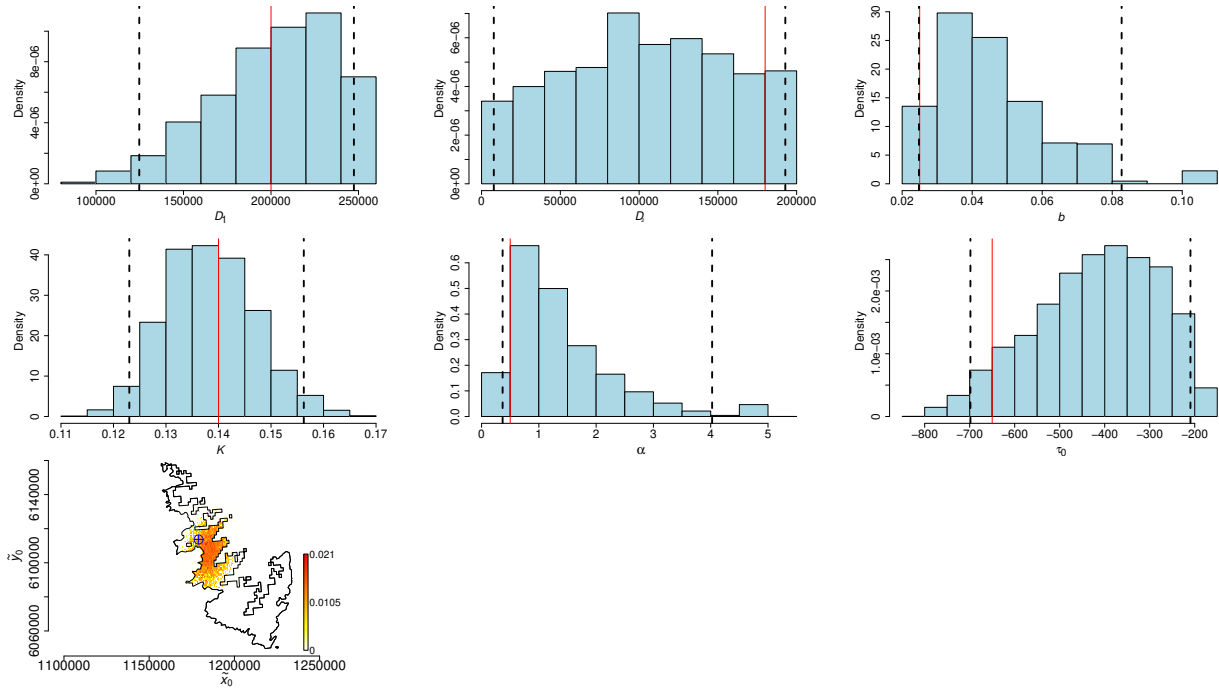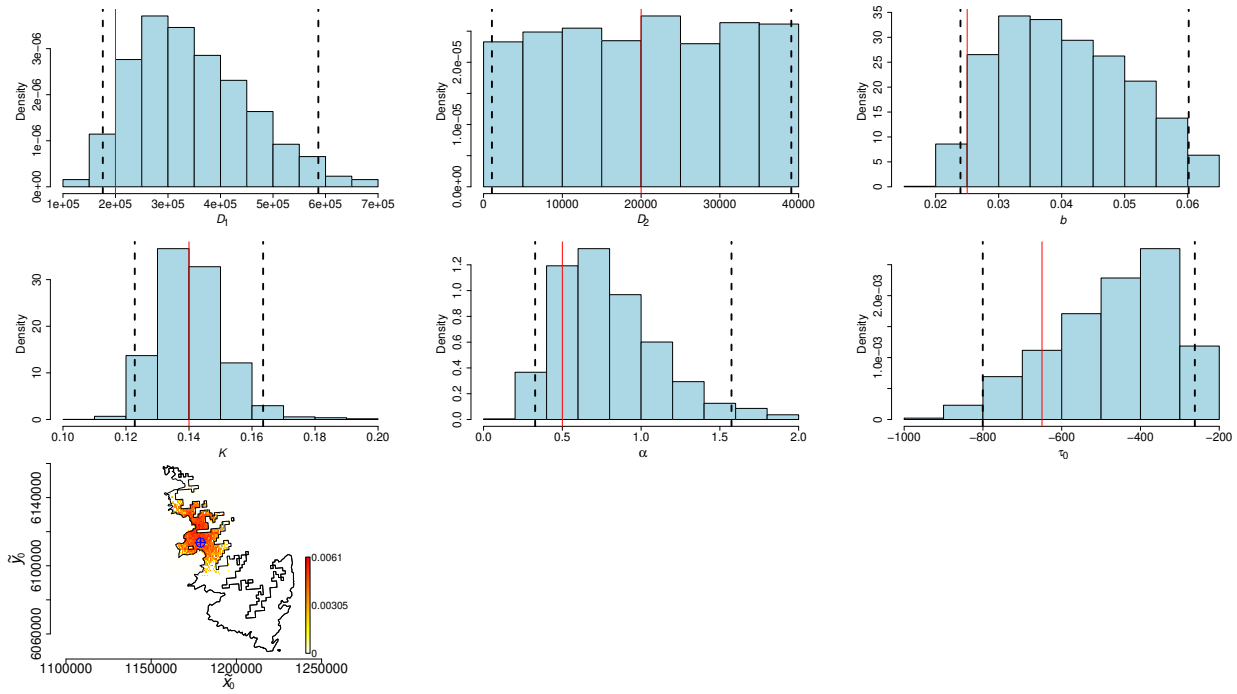
Figure S2: Posterior map of the introduction location and histograms of marginal posterior distributions for all the other parameters of model $\mathcal{M}_1(5.5)$ when it is fitted to data $\mathcal{O}_2$ in the simulation study. Red line: true value. Dashed black lines 0.025 and 0.975 posterior quantiles. Blue cross on the map: true introduction location. Color palette: posterior probability of the introduction location.

9

Figure S3: Posterior map of the introduction location and histograms of marginal posterior distributions for all the other parameters of model $\mathcal{M}_1(5.5)$ when it is fitted to data $\mathcal{O}_3$ in the simulation study. Red line: true value. Dashed black lines 0.025 and 0.975 posterior quantiles. Blue cross on the map: true introduction location. Color palette: posterior probability of the introduction location.
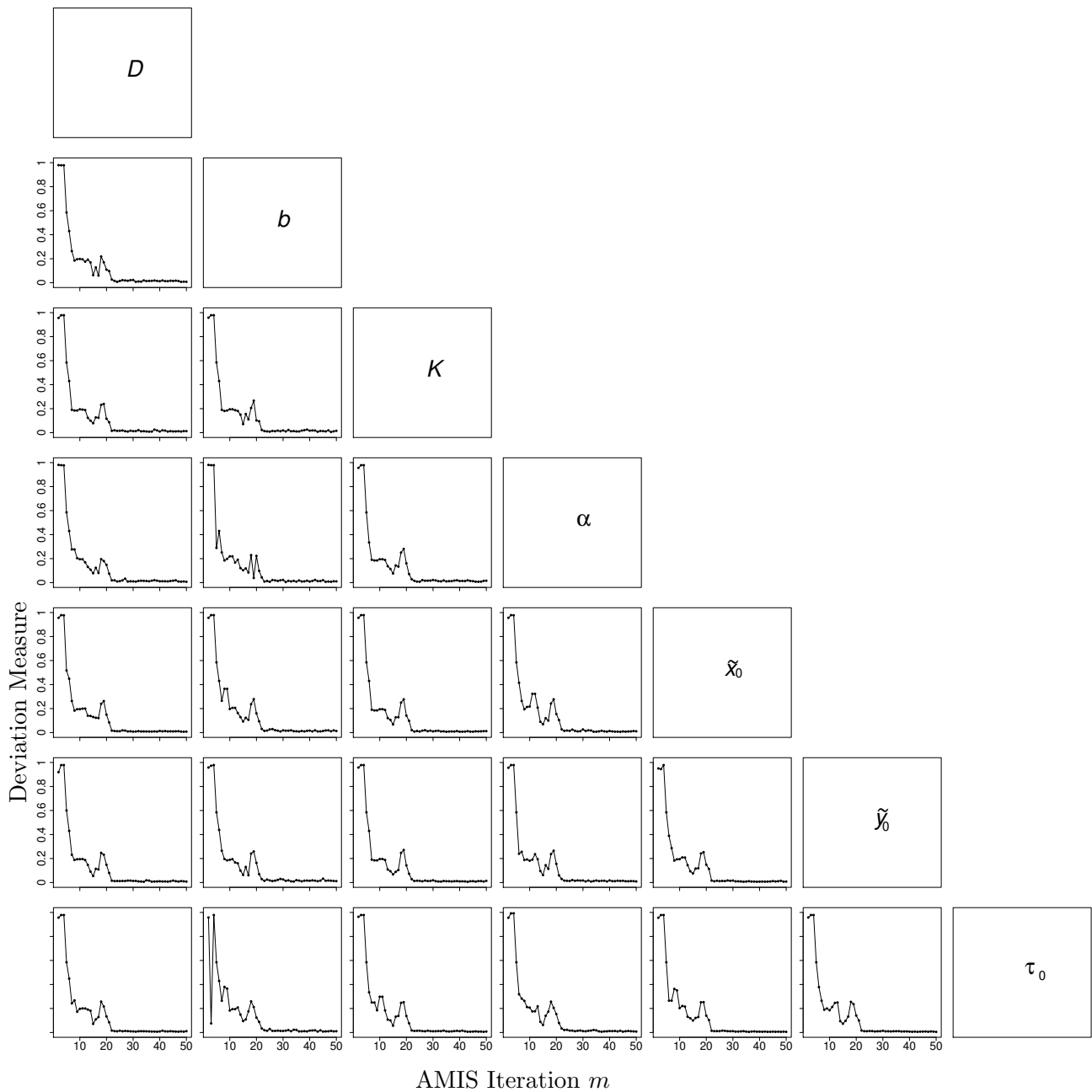
Figure S4: Variation in the deviation measure between the assessments of the posterior distribution at iteration $m-1$ and $m > 1$ of the AMIS algorithm for model $\mathcal{M}_0(4.75)$. The deviation measure is computed as suggested by Abboud et al. (2019) to assess the stabilization of all the 2D posterior distributions of parameters D, b, K, α, $\tilde{x}_0$, $\tilde{y}_0$ and τ_0 of model $\mathcal{M}_0(4.75)$.
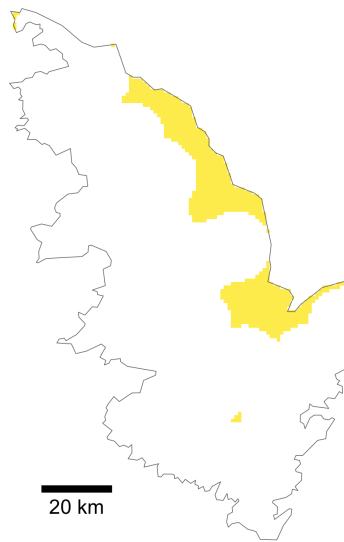
Figure S5: Set of grid cells (in yellow) in which the infection proportion is fixed at zero in the validation map (i.e., the out-of-sample prediction) providing local infection proportions. These grid cells are those satisfying two conditions: (i) no infection observed in training data up to 10 km; (ii) no observation in validation data up to 10 km.
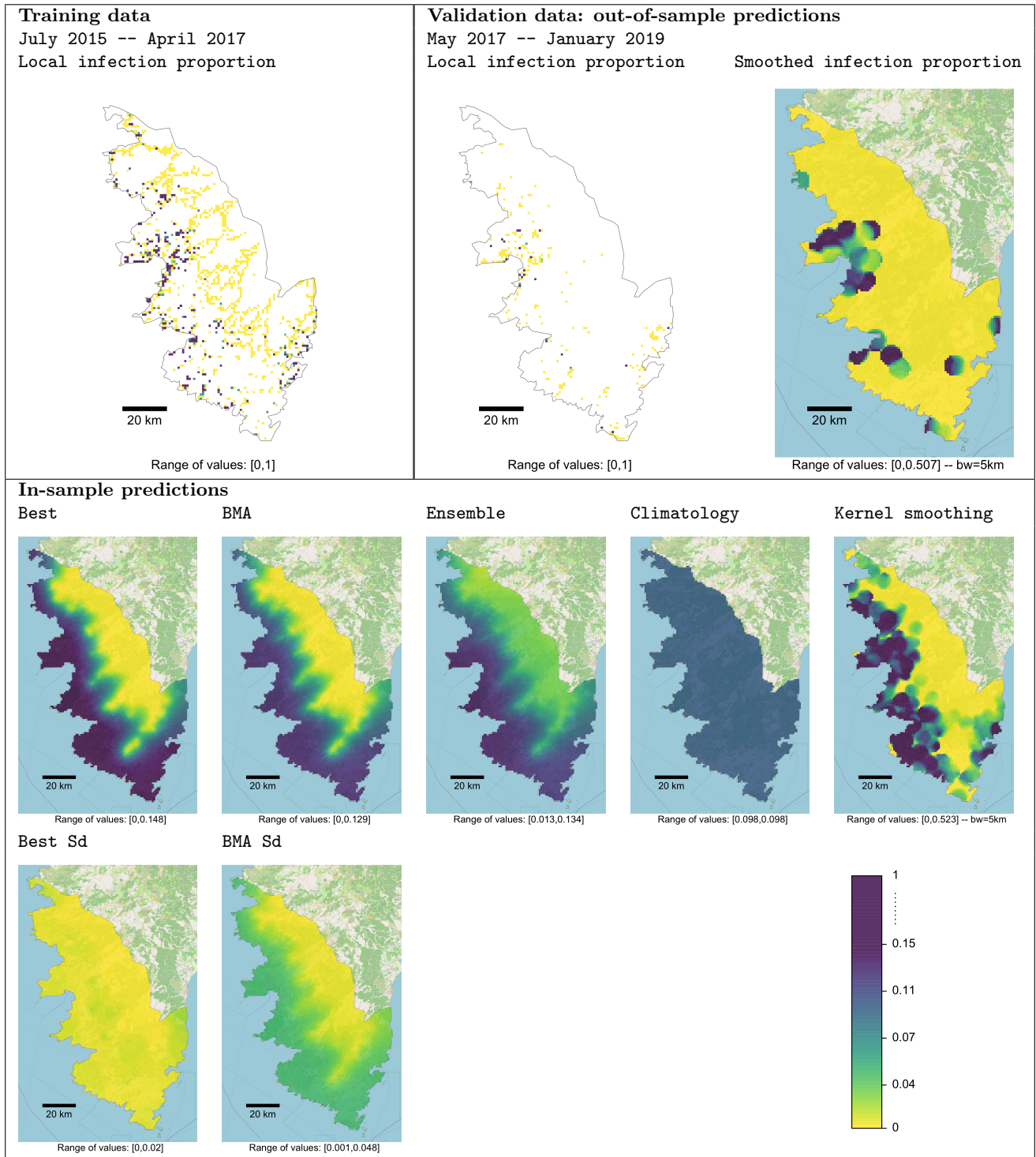
Figure S6: Training data (top left box), out-of-sample predictions computed either as a local infection proportion or as a kernel smoothing with a 5km-bandwidth (top right box), and in-sample predictions (i.e. posterior means provided by the best model and the BMA, ensemble, climatology and kernel smoothing with a bandwidth of 5 km; bottom box) obtained from training data. Map of standard deviations are also provided in the bottom box for the best model and BMA. The color palette (bottom right) applies to all plots. The range of values taken in each map is provided below the map.
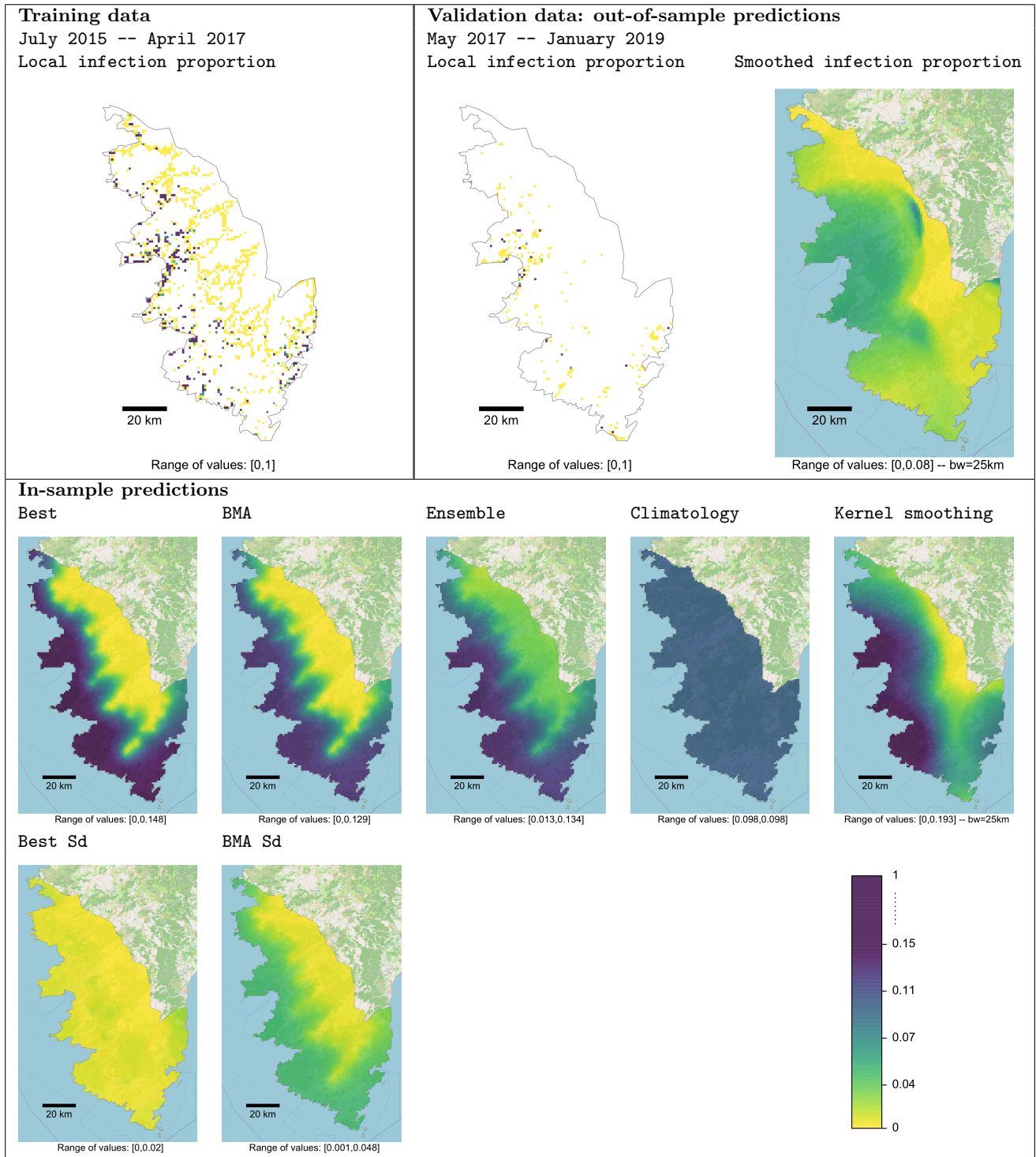
13

Figure S7: Training data (top left box), out-of-sample predictions computed either as a local infection proportion or as a kernel smoothing with a 25km-bandwidth (top right box), and in-sample predictions (i.e. posterior means provided by the best model and the BMA, ensemble, climatology and kernel smoothing with a bandwidth of 25 km; bottom box) obtained from training data. Map of standard deviations are also provided in the bottom box for the best model and BMA. The color palette (bottom right) applies to all plots. The range of values taken in each map is provided below the map.

14