Mathematical Biology

A Bayesian estimation approach for the mortality in a stage-structured demographic model

E. Lanzarone[1] · **S. Pasquali[1]** ·
G. Gilioli[2] · **E. Marchesini[3]**

Abstract Control interventions in sustainable pest management schemes are set according to the phenology and the population abundance of the pests. This information can be obtained using suitable mathematical models that describe the population dynamics based on individual life history responses to environmental conditions and resource availability. These responses are described by development, fecundity and survival rate functions, which can be estimated from laboratory experiments. If experimental data are not available, data on field population dynamics can be used for their estimation. This is the case of the extrinsic mortality term that appears in the mortality rate function due to biotic factors. We propose a Bayesian approach to estimate the probability density functions of the parameters in the extrinsic mortality rate function, starting from data on population abundance. The method investigates the time variability in the mortality parameters by comparing simulated and observed trajectories. The grape berry moth, a pest of great importance in European vineyards, has been considered as a case study. Simulated data have been considered to evaluate the con-

✉ S. Pasquali
  sara.pasquali@mi.imati.cnr.it

  E. Lanzarone
  ettore.lanzarone@cnr.it

  G. Gilioli
  gianni.gilioli@unibs.it

  E. Marchesini
  enrico.marchesini@agrea.it

[1] CNR-IMATI, Via A. Corti 12, 20133 Milan, Italy

[2] Department of Molecular and Translational Medicine, University of Brescia, Viale Europa 11, 25123 Brescia, Italy

[3] AGREA S.r.l. Centro Studi, Via Garibaldi 5/16, 37057 S. Giovanni Lupatoto (VR), Italy

🌀 Springer

vergence of the algorithm, while field data have been used to obtain estimates of the mortality for the grape berry moth.

Keywords Fokker–Plank equations · Bayesian inference · MCMC algorithms · Population dynamics · Pest · *Lobesia botrana*

Mathematics Subject Classification 62F15 · 92D25 · 65C05

1 Introduction

Ecologically based pest management is an important element in the development of sustainable agriculture, to improve system sustainability, farmer health and food safety. Sustainable pest management is becoming a key component in both intensive and marginal farming as confirmed, for instance, by the recent European guidelines (Directive 2009/128/EC) that encourage a sustainable use of pesticides in order to reduce environmental and health impact, and increase product quality.

Since the 1970s, multidisciplinary researches involving biologists, ecologists, entomologists and plant pathologists, mathematicians and meteorologists have developed useful tools to meet some of the multiple objectives of sustainable pest management. Innovative elements of these works have led to the development of mathematical models to support sustainable pest management programs. Many pest populations are stage-structured, and several modelling frameworks for stage-structured population dynamics have been proposed since the last decades of the twentieth century (see, e.g., Metz and Diekmann 1986; McDonald et al. 1989; Huffaker and Gutierrez 1999).

A useful approach to simulate population dynamics is based on Fokker–Planck partial differential equations (Gardiner 1994). Buffoni and Pasquali (2007) followed this approach to model the abundance of an insect in time and physiological age accounting for stochasticity in the development, and they studied existence and stability of equilibrium points for this model. Model formulation includes temperature-driven biodemographic responses that characterize the life history traits of the individuals. These responses are described by development, fecundity and survival rate functions.

In Gilioli et al. (2016), a modelling framework for stage-structured pest population dynamics based on Fokker–Planck partial differential equations has been presented and applied to the grape berry moth (*Lobesia botrana*, Den. & Schiff.), a key pest in European vineyards. In this modelling framework the biodemographic functions are estimated from laboratory experiments. If experimental data are not available, sampling data on field population dynamics can be used for their estimation. This is the case of the extrinsic mortality term that appears in the mortality rate function due to biotic factors. For sake of simplicity, the extrinsic mortality terms were assumed to be constant in time and space, and estimated using a least square estimation method.

In the present paper, we check whether the assumption of a constant extrinsic mortality is realistic or too restrictive, and we propose a Bayesian approach to estimate the stage-specific mortality terms for the same dynamical model presented in Gilioli et al. (2016). Starting from data collected in several years, we consider two cases for the mortality parameters estimation. First, we suppose that the parameters are constant

in time, and we use the data to update year by year the knowledge on the parameters. Second, we assume that the parameters are characteristic of each year and we use the data of each specific year for their estimation.

Time-series population dynamics observations are required to apply the estimation method. They are supposed to be affected by a measurement error, and then considered as random variables centered on the value given by the dynamical model. Also the state variables of the model are stochastic, due to the presence of stochastic parameters in the equations. The overall densities of the observations given the parameters are then used to compute the likelihood function of the unknown parameters that, combined with prior distributions, give the posterior distributions according to the Bayes' Theorem. The population dynamics model considered does not allow knowing posterior distributions in a closed form; thus, they are sampled using a Markov Chain Monte Carlo (MCMC) algorithm.

The Bayesian framework is chosen because of the high flexibility. Bayesian estimation allows to obtain the probability distribution for each parameter, thus providing a complete information about the shape of the density together with all of its quantiles. Another important feature of the Bayesian approaches is that they are efficient also in case of few available data, eventuality often occurring in ecological problems where it is burdensome to collect a large amount of data.

More in general, Bayesian approaches represent a powerful tool to derive knowledge and make predictions in ecology (see Dorazio 2016), and they have been already applied to estimate parameters in population dynamics described by stochastic ordinary differential equations. For example, Gillespie and Golightly (2010) estimated parameters in a stochastic population growth model using a MCMC method; Heydari et al. (2014) proposed a Bayesian parameter estimation for stochastic logistic growth models; Gilioli et al. (2008, 2012) and Martín-Fernández et al. (2014) proposed Bayesian inference methods to estimate parameters in a stochastic predator-prey system. Another Bayesian modelling approach in an age-structured model can be found in Tsehaye et al. (2014), where a multispecies model assesses the trade-offs between predatory demands and prey productivities for a pelagic fish community.

Differently from these cases, in the present paper, the population dynamics is described by a system of partial differential equations (see Gilioli et al. 2016). The appropriateness of Bayesian estimation methods for partial differential equations has been documented in the literature (Prakasa Rao 2000; Xun et al. 2013). For example, Lanzarone et al. (2014) and Martín-Fernández and Lanzarone (2015) recently estimated the thermal conductivity of a polymeric specimen. However, to the best of our knowledge, Bayesian approaches for partial differential equations have been only marginally used in ecology. Hooten and Wikle (2008) presented a hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species. Spence et al. (2016) used a MCMC method to estimate a large number of parameters in a dynamic multispecies size spectrum model for aquatic communities, described by a system of partial differential equations.

Finally, a large number of papers can be found in the frequentist setting for partial differential equations; for example, a recent paper by Roques et al. (2016) provides a maximum likelihood approach to estimate the diffusion rates in a population dynamics model described by a system of partial differential equations.

In this paper, the Bayesian method has been applied to both simulated and field data for the dynamics of the grape berry moth. The application to simulated data is considered to evaluate the performance of the proposed method. The application to field data allows to verify whether the assumption of constant extrinsic mortality terms is realistic, and to make biological considerations on estimated mortality terms of *L. botrana*.

The paper is structured as follows. The adopted population dynamics model and the Bayesian estimation model for the mortality coefficients are presented in Sects. 2 and 3, respectively. Then, the application to *L. botrana*, considering both a simulated dataset and the real field data, is described in Sect. 4. The obtained results are presented in Sect. 5, and some discussions and the conclusion of the work are reported in Sect. 6.

2 The population dynamics model

The population dynamics is represented by a system of partial differential equations that describe the abundance of a population in time t and physiological age x. The population is split in generations; for each generation, the model describes a stage-structured population composed by $s - 1$ immature stages and one reproductive stage (stage s). The physiological age x represents the percentage of completion of development in a stage. It ranges in the interval $[0, 1]$, where $x = 0$ when the individual enters in the stage, and $x = 1$ when the individual completes the development in that stage. As mentioned, we refer to the formulation proposed in Gilioli et al. (2016).

Let us denote by $\phi^i(t, x)$ the abundance of individuals in stage $i\,(i = 1, \ldots, s)$ at time t, and by $\phi^i(t, x)dx$ the number of individuals in stage i at time t with physiological age in $(x, x + dx)$. The population dynamics is described by the system:

$$\frac{\partial \phi^i}{\partial t} + \frac{\partial}{\partial x}\left[v^i(t)\phi^i - \sigma^i \frac{\partial \phi^i}{\partial x}\right] + m^i(t)\phi^i = 0 \quad t > t_0, \ x \in (0, 1) \tag{1}$$

$$\left[v^i(t)\phi^i(t, x) - \sigma^i \frac{\partial \phi^i}{\partial x}\right]_{x=0} = F^i(t) \tag{2}$$

$$\left[-\sigma^i \frac{\partial \phi^i}{\partial x}\right]_{x=1} = 0 \tag{3}$$

$$\phi^i(t_0, x) = \hat{\phi}^i(x) \tag{4}$$

where t_0 is the initial time, i.e., the time of the first survey, $\hat{\phi}^i(x)$ is the initial abundance in stage i as function of physiological age, and σ^i are the diffusion coefficients ($i = 1, \ldots, s$), assumed to be time independent.

Moreover, $v^i(t)$ and $m^i(t)$ are the development and the mortality rate functions, respectively. $F^1(t)$ is the egg production flux, given by:

$$F^1(t) = b(t) \int_0^1 f(x)\phi^s(t, x)dx \tag{5}$$

where $b(t)$ summarizes the effects due to both diet and temperature, and $f(x)$ is the fertility profile. The other fluxes from stage $i - 1$ to stage i are given by

$$F^i(t) = v^{(i-1)}(t)\phi^{(i-1)}(t, x), \quad i > 1.$$

Finally, the number of individuals in stage i at time t is obtained by integrating $\phi^i(t, x)$ over x:

$$N^i(t) = \int_0^1 \phi^i(t, x)dx. \tag{6}$$

Further details can be found in Buffoni and Pasquali (2007) and Gilioli et al. (2016). The functions $f(x)$ and $b(t)$ for the egg production flux in Eq. (5) are given by Gutierrez et al. (2012):

$$f(x) = \max\left\{\frac{\delta(\eta x - \xi)}{\theta^{(\eta x - \xi)}}, 0\right\} \tag{7}$$

$$b(t) = b_0\left(P(t)\right) a_0\left(\hat{T}(t)\right) \tag{8}$$

where ξ/η is the physiological age at which oviposition appears (Baumgärtner and Baronio 1988), and $\delta\eta$ and θ^η are the initial fecundity rate and the fecundity decrease with the adult age, respectively (Bieri et al. 1983). Function $b(t)$ depends on temperature and on the phenological stage of the plant $P(t)$ (where $P(t)$ assumes a finite number of values). The function $b_0\left(P(t)\right)$ is a step function that depends on the considered population, and will be reported later in Sect. 4.1 for the case of *L. botrana*. Moreover,

$$a_0(\hat{T}) = 1 - \left[\frac{\hat{T}(t) - \hat{T}_L - \hat{T}_0}{\hat{T}_0}\right]^2 \tag{9}$$

according to Gutierrez et al. (2012), where $\hat{T}(t)$ is the average temperature over a time period $[t - \tau, t]$, which characterizes the temperature regime influencing the status of an adult female, and where τ is a suitable time interval; $\hat{T}_L$ is the minimum temperature of reproduction; and $\hat{T}_0$ is the half-width of the temperature interval of reproduction.

We assume, as in Gilioli et al. (2016), that the development and mortality rate functions depend on time t through the temperature $T(t)$.

The development rate function $v^i(T)$ is given for all i by a Lactin function (Lactin et al. 1995):

$$v^i(T) = \delta^i \max\left\{0, e^{\alpha^i T} - e^{\alpha^i T_m - \frac{T_m - T}{\beta^i}} - \gamma^i\right\} \tag{10}$$

where parameters α^i, β^i, γ^i, δ^i and T_m depend on the population under study.

The mortality rate function $m^i(T)$ is defined for all i as:

$$
m^i(T) = \begin{cases}
-v^i(T)\ln\left(1 - M^i(T)\right) + J_g^i \\
\quad T_{ML}^i \leq T \leq T_{MU}^i \\
\left[-v^i\left(T_{ML}^i\right)\ln\left(1 - M^i(T)\right) + J_g^i\right]\left[\left(T_{ML}^i - T\right)^2 + 1\right] \\
\quad T < T_{ML}^i \\
\left[-v^i\left(T_{MU}^i\right)\ln\left(1 - M^i(T)\right) + J_g^i\right]\left[\left(T - T_{MU}^i\right)^2 + 1\right] \\
\quad T > T_{MU}^i
\end{cases}
\tag{11}
$$

where $M^i(T)$ is the average stage proportional mortality as function of temperature (Briolini et al. 1997):

$$
M^i(T) = \begin{cases}
a^i\left(\dfrac{T - T_{M0}^i}{T_{M0}^i}\right)^2 + \varepsilon & T_{ML}^i \leq T \leq T_{MU}^i \\
0.85 & \text{otherwise}
\end{cases}
\tag{12}
$$

with ε representing the minimum fraction of dying individuals at temperature T_{M0}^i (that we assume equal to 0.1), and

$$
T_{M0}^i = \frac{T_{ML}^i + T_{MU}^i}{2}
$$

$$
a^i = (0.85 - \varepsilon)\left(\frac{T_{MU}^i}{T_{M0}^i} - 1\right)^{-2}.
$$

The interval $\left[T_{ML}^i, T_{MU}^i\right]$ $(i = 1, \ldots, s)$ is enclosed in the temperature interval where the development function is non-zero. The value 0.85 outside the interval $\left[T_{ML}^i, T_{MU}^i\right]$ and where the development is non-zero is arbitrarily chosen to guarantee a little fraction of individuals surviving for low and high temperatures.

The terms J_g^i in (11), which depend on the stage $i (i = 1, \ldots, s)$ and on the generation $g (g = 1, \ldots, G)$, represent the extrinsic mortality terms due to biotic factors (Gilioli et al. 2016).

3 Bayesian estimation method

Estimation of the parameters in the biodemographic functions are based on experimental data, when available. Unfortunately, no literature data are available for the extrinsic mortality terms J_g^i. Hence, in this paper, we propose a Bayesian estimation method for these terms based on population dynamics field data. A previous parameter estimation, based on a least square method, was presented in Gilioli et al. (2016). Here, differently from Gilioli et al. (2016), we do not make the restrictive assumptions that the terms J_g^i are constant for different years, but we allow to have different mortalities

for each year. As mentioned in the introduction, parameters J_g^i are assumed either *constant*, updating year by year their knowledge, or *variable* in time, assuming independent J_g^i for each year. The Bayesian approach gives an estimate of the J_g^i in terms of their probability density functions, which allow to directly evaluate the uncertainty associated to the estimates.

Each parameter J_g^i is assumed to be a random variable, and the estimate is given in terms of its marginal posterior density, which is derived from the joint posterior distribution of the parameter vector $\mathbf{J} = \left(J_g^i \right), i = 1, \ldots, s; g = 1, \ldots, G$. Such posterior distribution summarizes the information coming from the prior knowledge on the parameters and the process observations (the population dynamics in our case).

The next three subsections describe the elements of the Bayesian estimation, i.e., the likelihood function, the prior density, and the posterior density. The last subsections deal with the parameter estimation in the presence of multi-year observations and with the method adopted to simulate the dynamics, respectively.

3.1 Likelihood function

The likelihood function is obtained from the discretized formulation of the population dynamics model in (1)–(6).

For this purpose, the system of partial differential equations is discretized both in physiological age x and time t for each stage i (Buffoni and Pasquali 2007; Cola et al. 1999). The number of individuals in stage i at time t_h is denoted by $N^i(t_h)$, $h = 1, 2, \ldots, H$, where H is the number of points of the time-discretization.

Each unknown extrinsic mortality term J_g^i in (11) is assumed to be a random variable. In this way, each discretized mortality rate function $\left\{ m^i \left(T(t_h) \right), h = 1, \ldots, H \right\}$ is a random process. As a consequence, given the initial conditions at time t_0, the discretized abundances $\left\{ \phi^i(t_h, x), h = 1, \ldots, H \right\}$ and the discretized numbers of individuals $\left\{ N^i(t_h), h = 1, \ldots, H \right\}$ are also random processes.

We can express the density of each $N^i(t_h)$ as conditioned to the terms J_g^i and the $N^i(t_{h-1})$ values. In particular, following the structure of the system (1)–(6) discretized in time and physiological age, we can write

$$N^i(t_h) \sim \mathcal{L} \left(N^i(t_h) | N^1(t_{h-1}), \ldots, N^s(t_{h-1}), \mathbf{J} \right) \quad \forall i, h \tag{13}$$

where $\mathcal{L}$ denotes the conditioned probability law, which does not follow any known form due to the structure of the system.

Let us suppose, for each i, that observations N_{obs}^i of N^i are taken at some time instants t_{h^*}, where $\{t_{h^*}\}$ is the subset of the discretized instants $\{t_h\}$ in which an observation is available (the discretization step Δt is chosen according to the differential equations, and the frequency of the observations may be less dense than Δt, or observations may not be equally spaced).

We consider that the observations $N_{obs}^i(t_{h^*})$ are subject to a measurement error; then, we assume them as stochastic variables following a Gamma distribution with modal value $N^i(t_{h^*})$:

$$N^i_{obs}(t_{h^*}) \sim \mathcal{G}\left(1 + \beta_{lik} N^i(t_{h^*}), \beta_{lik}\right) \qquad \forall i, h^* \tag{14}$$

where $\mathcal{G}$ denotes the Gamma distribution with shape parameter $1 + \beta_{lik} N^i(t)$ and rate parameter β_{lik}. Parameters β_{lik} have not a biological meaning, while they simply allows tuning the variance of the distributions keeping the modal value fixed at $N^i(t_{h^*})$.

For the sake of simplicity, we consider in the notation the same subset $\{t_{h^*}\}$ for all stages i; however, there is not limitation and the subsets may differ from stage to stage.

Combining (13) and (14), we obtain for each i the conditional law of the observations at each time instant t_{h^*} in which an observation is available. The product of the conditional laws over t_{h^*} determines for each i the likelihood function of the vector of observations given the parameters J^i_g, and their product over i determines the overall likelihood function.

$$f\left(\hat{\mathbf{N}}_{obs}|\mathbf{J}\right) = \prod_{i,h^*} \mathcal{G}\left(1 + \beta_{lik} N^i(t_{h^*}), \beta_{lik}\right)$$
$$\times \mathcal{L}^*\left(N^i(t_{h^*})|N^1(t_{h^*-1}), \ldots, N^s(t_{h^*-1}), \mathbf{J}\right) \tag{15}$$

where $\hat{\mathbf{N}}_{obs} = \left(\hat{N}^1_{obs}, \ldots, \hat{N}^s_{obs}\right)$ denotes the overall set of observations, and $\hat{N}^i_{obs}$ the set of all observations for stage i.

The law $\mathcal{L}^*\left(N^i(t_{h^*})|N^1(t_{h^*-1}), \ldots, N^s(t_{h^*-1}), \mathbf{J}\right)$ represents the marginal density of $N^i(t_{h^*})$, obtained combining all densities $\mathcal{L}$ in (13) between the current observation at t_{h^*} and the previous one at t_{h^*-1}.

Other approaches are possible, e.g., approaches that include the generation of latent observations (see, among others, Durham and Gallant 2002; Elerian et al. 2001; Golightly and Wilkinson 2005; Gilioli et al. 2008, 2012; Lanzarone et al. 2014; Martín-Fernández et al. 2014). We avoid the generation of latent data because, in the proposed form, the distributions of $N^i(t_{h^*})$ already take into account all values of $N^i(t_h)$ in the interval $[t_{h^*-1}, t_{h^*})$.

3.2 Prior density

We denote as $\pi(\mathbf{J})$ the prior distribution for the parameter vector $\mathbf{J}$. Each parameter J^i_g is assumed to be *a priori* independent of the others, so that the joint prior density is the product of the marginal prior densities:

$$\pi(\mathbf{J}) = \prod_{i,g} \pi\left(J^i_g\right)$$

Two types of marginal prior distributions are adopted for each parameter:

1. *Uniform distribution* defined by the minimum value J^{ig}_{min} and the maximum value J^{ig}_{max};

2. *Gamma distribution* defined by the shape $\alpha_{J_g^i}$ and rate $\beta_{J_g^i}$ parameters, respectively.

Both of them refer to non-negative parameters; hence, $J_{min}^{ig} \geq 0$ is imposed for each parameter, while the Gamma density does not allow negative values and no restrictions are required.

The choice about the type of the marginal prior distribution and the associated parameters will be discussed later in the application to the grape berry moth population. For now, we anticipate that the first alternative refers to the case in which the only available information is an admissible range for the parameters, whereas the latter to the case where additional *a priori* knowledge can be exploited.

3.3 Posterior density

The joint posterior density is proportional to the product between the likelihood function and the joint prior distribution, according to the Bayes' theorem:

$$\pi \left(\mathbf{J} | \hat{\mathbf{N}}_{obs} \right) \propto f \left(\hat{\mathbf{N}}_{obs} | \mathbf{J} \right) \pi \left(\mathbf{J} \right) \tag{16}$$

Then, from the joint posterior density $\pi \left(\mathbf{J} | \hat{\mathbf{N}}_{obs} \right)$, each marginal posterior density $\pi \left(J_g^i | \hat{\mathbf{N}}_{obs} \right)$ is obtained.

The posterior estimation cannot be derived in a closed analytical form; hence, a Markov Chain Monte Carlo (MCMC) method is adopted. We use STAN (Stan Development Team 2015) through its R interface, which implements the Hamiltonian Monte Carlo algorithm for obtaining a sequence of random samples from each marginal posterior density. We adopt STAN rather than other more spread tools, e.g. JAGS (Plummer 2003), mainly for two reasons:

– STAN allows to use C++ code with the possibility of including *for* cycles, whereas other tools do not allow cycles and a variable must be declared for each index, with consequent higher memory utilization;
– the Hamiltonian Monte Carlo, also known as hybrid Monte Carlo algorithm, differs from the Metropolis–Hastings algorithm (a widely used MCMC method) because it reduces the correlation between successive samples; this guarantees a quicker convergence with consequent lower computational times (Duane et al. 1987).

3.4 Multi-year approach

Data on grape berry moth population abundance are available during the growing season, i.e., from Spring to Autum. In the presence of data acquired in several years, our approach can be iteratively applied year by year, once the new data of a year are acquired. In this case, our estimation method can be applied in two different ways, i.e., following a *constant approach* or a *variable approach*. The first is used when the mortality parameters can be assumed to be constant over

the years, while the latter when the mortality parameters may evolve from year to year. Every year, the initial condition for system (1)–(4) is given by the number of adults collected per week in that year, until the first larvae of the first generation are observed.

- *Constant approach*

  If the mortality parameters are constant, we can increase their knowledge year by year by adding the new observations. The marginal posterior densities of $\mathbf{J}$ obtained in a year are used as the prior independent densities of J_g^i for the following year. If the mortality parameters are actually constant, their marginal posterior densities should concentrate year by year around the *true* parameter values with stricter variances. The algorithm for the *constant approach* can be summarized as follows:

  1. First year

     At the first year, the only available information is an admissible range for all mortality parameters. Then, a Uniform prior distribution on this interval has been chosen for each parameter, and parameters are assumed to be independent. The joint posterior distribution of $\mathbf{J}$ is then obtained applying the estimation method described in Sects. 3.1–3.3. Such joint density is finally marginalized into the marginal densities of each mortality parameter J_g^i.

  2. Each following year

     We start from the marginal posterior densities of the previous year. Each one is fitted either with a Uniform or a Gamma density depending on its shape: a Gamma density truncated in the admissible range is used if the density is strict around a peak value, while a Uniform density on the admissible range is chosen in case a peak is not present. These marginal and fitted densities are then used as independent priors for the current year. Once again, the joint posterior distribution of $\mathbf{J}$ is obtained applying the Bayesian estimation method. Finally, such joint density is marginalized into the marginal densities of each mortality parameter J_g^i.

- *Variable approach*

  The parameter estimation is independently performed year by year, starting from the same Uniform prior distributions at each year. We only apply step 1 of the algorithm presented for the *constant approach*. In this way, we obtain different posterior distributions for each considered year.

3.5 Generation of simulated dynamics

The population dynamics are simulated considering all samples of parameter vector $\mathbf{J}$ generated by the MCMC algorithm. For each sample of the vector, we draw the trajectory of all populations. Then, the median trend and the 2.5–97.5% confidence bands among the sampled trajectories are extracted point by point from the trajectories, as done, e.g., in Lamonica et al. (2016a, b).

As for the *constant approach*, we use the posterior density of the last year to generate the dynamics of all years. As for the *variable approach*, the dynamics of each year are generated with the posterior density of the same year.

4 Application to the grape berry moth population

The proposed Bayesian estimation approach has been applied to the population dynamics of the grape berry moth *L. botrana*, i.e., the most important pest of grape (*Vitis vinifera*) in the Mediterranean basin (CABI 2014). The population is composed by 4 biological stages: the first 3 stages ($i = 1, 2, 3$) are immature (i.e., eggs, larvae and pupae), while the fourth stage ($i = 4$) represents the reproductive adults.

Data on population abundance (number of individuals per sampling unit) were collected, for the late cultivar Garganega, in a vineyard in Colognola ai Colli (Italy) for the years 2008, 2009, and 2011, as presented in Gilioli et al. (2016). At each year, the number of eggs, larvae, pupae and adults have been weekly recorded, from April up to grape harvest in September, on a sample of 100 grapes. The number of available observations is between 25 and 31 for each year. Moreover, the temperatures have been hourly recorded by a meteorological station placed nearby the vineyard. Collected data show four generations of *L. botrana* ($G = 4$), i.e., the maximum number of generations in Southern Europe (Marchesini and Dalla Montà 2004; Pavan et al. 2010).

To validate the approach, we first apply the *constant approach* to a simulated dataset generated with constant parameters (see Sect. 4.2). The goal is to check whether the marginal posterior distributions concentrate year by year around the *true* values of the parameters (those used to generate the simulated data). Then, we apply both the approaches to the real field data, for which we do not know whether the parameters are constant or vary from year to year.

The discretization step Δt is 2 h and each t_{h*} is the day and time slot in which the observation N_{obs}^i is taken. The parameters β_{lik} are taken equal to 1 for all the stages, i.e., we assume Gamma densities (14) with equal mean and variance.

Estimations are run on a Server equipped with X86-64 AMD Opteron processor 6328 and 26 GB of installed RAM, considering 1000 iterations after a warm up of 100 iterations. This order of magnitude for number of iterations is admissible for the Hamiltonian Monte Carlo implemented in STAN (Stan Development Team 2015), because of the reduced correlation between successive samples (Duane et al. 1987). In addition, standard convergence diagnostics have been checked using the CODA package (Plummer et al. 2006), and trace plots, autocorrelations and bivariate scatter plots have been checked for all parameters. All of these analyses indicated in all cases the convergence of the chain with the used number of iterations.

4.1 Parameters of biodemographic functions

The parameters of the biodemographic functions adopted in this work are the same estimated in Gilioli et al. (2016) by means of a least square method. Stage-dependent parameters are reported in Table 1.

Moreover, T_m in (10) is equal to 36 °C, while the values of the parameters appearing in (7) are $\delta = 1.4175$, $\eta = 380$, $\xi = 16$, and $\theta = 1.025$, respectively. The temperatures in (9) are $\hat{T}_L = 17$ °C and $\hat{T}_0 = 7.5$ °C. The diffusion coefficients σ^i are taken equal to 0.0001 for all i. The time interval τ is 10 days.

Table 1 Parameters of the development (10), and mortality (11) rate functions of *L. botrana* (Gilioli et al. 2016), for the different stages: eggs ($i = 1$), larvae ($i = 2$), pupae ($i = 3$), adults ($i = 4$)

Stage	α^i	β^i	γ^i	δ^i	T^i_{ML}	T^i_{MU}
$i = 1$	0.01	0.8051	1.0904	1	8.8	34.79
$i = 2$	0.003	0.662	1.0281	1	9.1	34.27
$i = 3$	0.0076	1.7099	1.0929	1.1	11.73	32.67
$i = 4$	0.0076	1.7099	1.0929	1.1	11.73	32.67

Finally, the discrete function $b_0(P(t))$ takes the following values depending on the plant stages (Gutierrez et al. 2012)

$$b_0(P(t)) = \begin{cases} 0.31 & P(t) = \text{inflorescence stage} \\ 0.48 & P(t) = \text{green berries} \\ 1 & P(t) = \text{maturing fruits} \end{cases}$$

We anticipate here that, based on the results obtained in Gilioli et al. (2016), we assume that parameters J^i_g are constrained in the interval [0, 0.25]. The left bound is due to the fact that the parameters must be non-negative, according to their definition, while the right bound exceeds the values of natural mortality reported in literature (Marchesini 2007). Hence, values J^{ig}_{min} and J^{ig}_{max} of the Uniform prior densities are assumed equal to 0 and 0.25 for each i and g, respectively.

4.2 Generation of the simulated dataset

Simulated data are generated for the same 3 years starting from the same initial conditions of the field data, i.e., the number of adults collected per week until the first larvae of the first generation are observed. The same temperature hourly recorded have been considered, and the mortality parameters J^i_g reported in Table 2 have been adopted, which are an approximation of those estimated in Gilioli et al. (2016). Finally, as for the other parameters of the biodemographic functions, those described in Sect. 4.1 have been uesed.

We run the discretized version of the demographic model in (1)–(6) to obtain weekly data for eggs, larvae, pupae and adults. We generate the values at each discretized time instant up to the third adults' generation, and we take as simulated data the values at the beginning of the day in which a field observation has been collected. Depending on the year, we obtain from 18 to 20 data for each simulated dynamics.

Table 2 Values of parameters J^i_g used to generate the simulated data

J^i_g	$i = 1$ (eggs)	$i = 2$ (larvae)	$i = 3$ (pupae)	$i = 4$ (adults)
$g = 1$	–	–	–	0
$g = 2$	0	0.05	0	0
$g = 3$	0.2	0	0.05	0.2

5 Results

5.1 Simulated data

As mentioned, in the case of simulated data, we applied the *constant approach* for the three years 2008, 2009, and 2011 in sequence.

Figure 1 reports the histograms of the posterior distributions after the third year 2011 (i.e., the marginal densities of $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs\,2011}\right)$ obtained using the marginal densities of $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs\,2009}\right)$ as prior distributions), together with the fitted Gamma distributions. Results show that all histograms are well fitted by a truncated Gamma density, both in terms of modal value and variability.

Figure 2 shows the box-plots of the marginal posterior distributions for the 3 years (i.e., of $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs\,2008}\right)$ obtained with the uniform prior distributions, of $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs\,2009}\right)$ obtained with the marginal densities of $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs\,2008}\right)$ as prior distri-
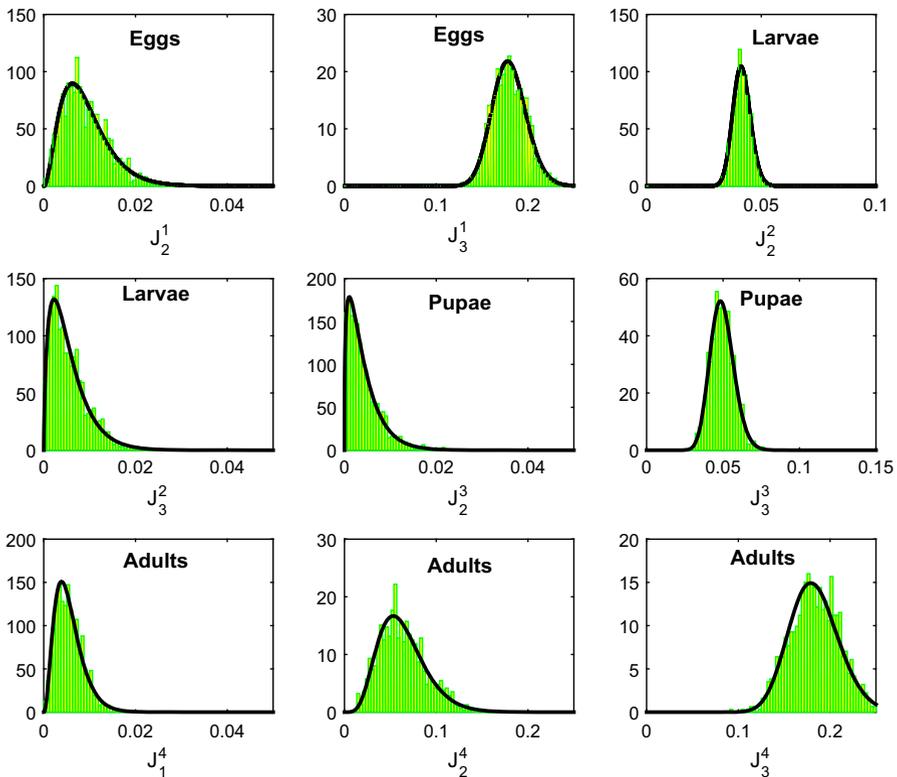
Fig. 1 Histograms of the J_g^i marginal posterior distributions after the third year 2011, for the case of simulated data: generations $g = 2, 3$ for stages $i = 1, 2, 3$ (i.e., eggs, larvae and pupae, respectively), and generations $g = 1, 2, 3$ for $i = 4$ (i.e., adults). *Black lines* represent the fit with a truncated Gamma density on [0,0.25]
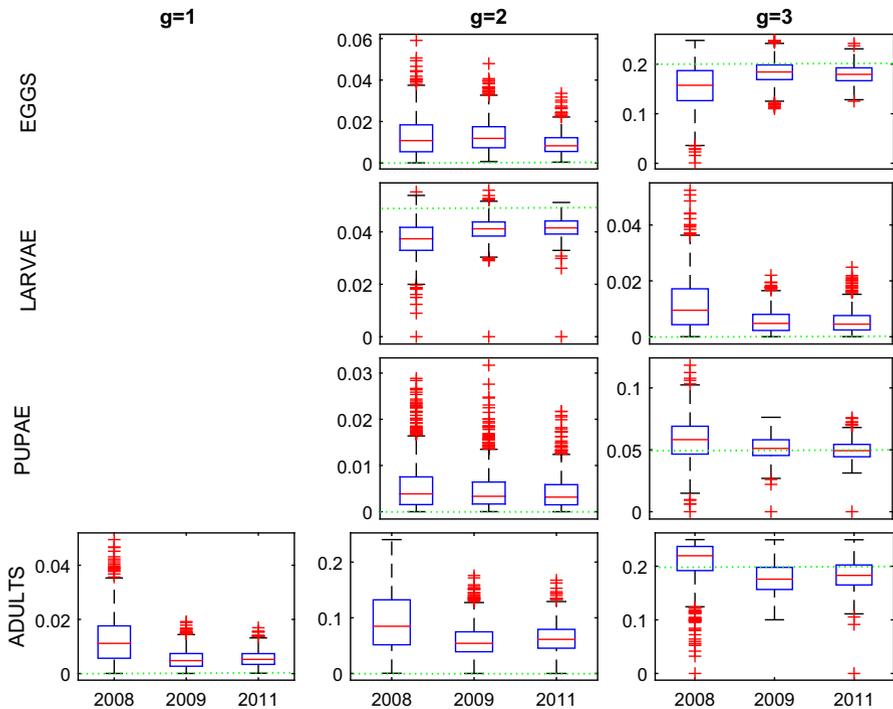
Fig. 2 *Box-plots* of the J_g^i marginal posterior distributions for all years 2008, 2009, and 2011, for the case of simulated data. *Dashed lines* represent the *true* values of the parameters (Table 2)

butions, and of $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs2011}\right)$ obtained with the marginal densities of $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs2009}\right)$ as prior distributions), compared with the *true* values used for generating the dataset (Table 2). It can be observed that the *true* values are always included in the 2.5–97.5% confidence interval of the posterior densities, but they are often far from the median of the distribution. However, this is not a drawback because the simulated dynamics reproduce the dataset very well. In particular, all simulated observations are included in the 2.5–97.5% confidence interval of the dynamics and, above all, most of the observations are very close to the median (Fig. 3).

The box-plots show a shrinkage of the 25–75% confidence intervals from 2008 to 2011. The variance of the curves decreases in time and the greater shrinkage of the distributions is observed from the first to the second year.

Finally, we remark that strict 2.5–97.5% confidence bands are obtained for all dynamics (Fig. 3), confirming the converge of the approach.

5.2 Field data

For the field data we firstly applied the *constant approach* over the three years 2008, 2009, and 2011, computing the posterior distributions as in the simulated data case. Differently from the simulated data case, here we do not observe a shrinkage in time
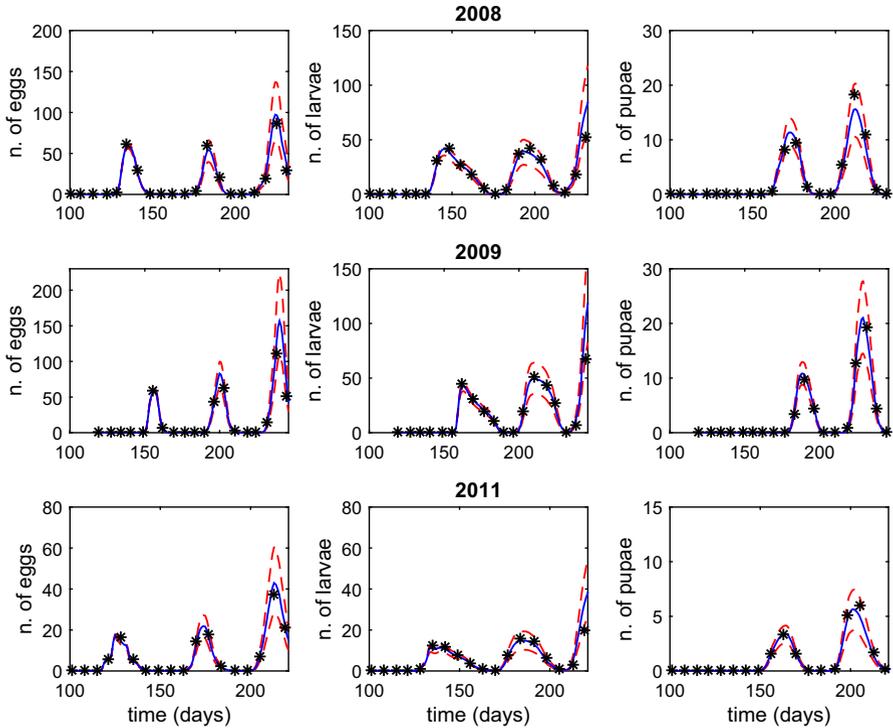
Fig. 3 Simulated dynamics of *L. botrana* for all years 2008 (*first row*), 2009 (*second row*), and 2011 (*third row*), for the case of simulated data: median (*continuous line*) and 2.5–97.5% confidence bands (*dashed lines*). *Asterisks* represent the simulated data

for all the mortality parameters (Fig. 4), especially for the second generation ($g = 2$) and for pupae ($i = 3$). Moreover, we do not observe any trend over the years, and most of the 2.5–97.5% confidence intervals overlap among years.

Then, we applied the *variable approach*, where no constraints link the parameters among the years; the box-plots of the marginal posterior distributions of $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs\,2008}\right)$, $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs\,2009}\right)$ and $\pi\left(\mathbf{J}|\hat{\mathbf{N}}_{obs\,2011}\right)$, all obtained with the uniform prior distributions, are reported in Fig. 5. From the comparison with Fig. 4, it can be seen that a large variability among years exists in both cases for the pupae, while the estimates of extrinsic mortality rates are quite constant for the eggs of the last two generations and the larvae of the third generation. For all stages and generations, the variability is greater in the *variable* (Fig. 5) than in the *constant* approach (Fig. 4), especially for adults, because of the Uniform priors used in all the years. Also in this case, we do not observe any trend over the years, and the 2.5–97.5% confidence intervals overlap among years.

Looking at the box-plots of the posterior distributions, even if a certain overlapping is observed, it is not possible to establish whether parameters are constant or variable in time. Wilcoxon hypothesis test, performed for each extrinsic mortality term to investigate the differences from year to year, reveals significant differences for both
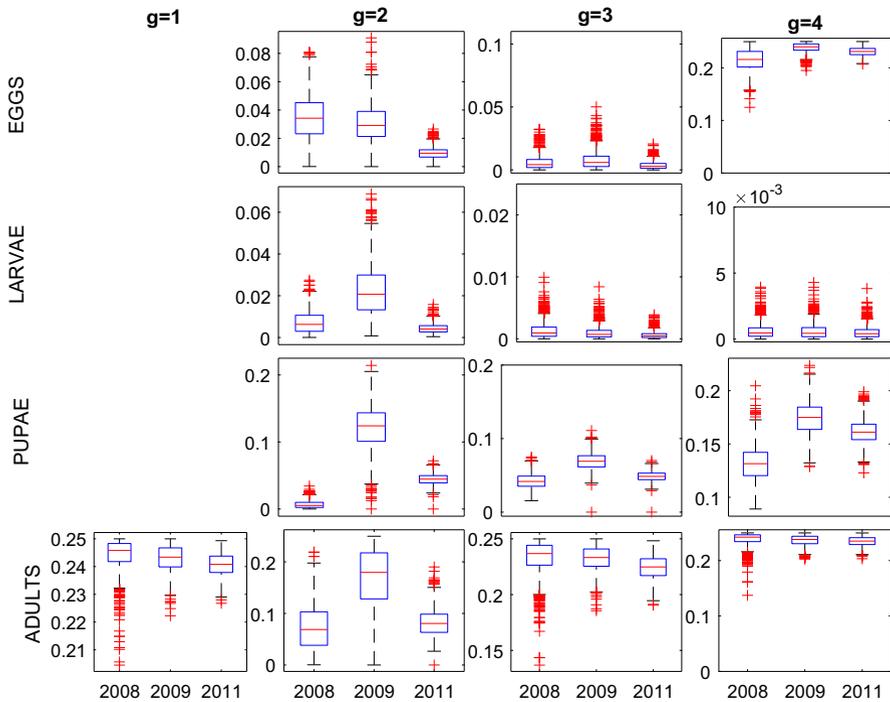
Fig. 4 *Box-plots of the* J_g^i *marginal posterior distributions for all years 2008, 2009, and 2011, for the case of field data, under the* constant approach

the case of constant and variable approach. However, these significant differences are not associated with any trend and cannot be explained by any environmental covariate; therefore this variability can be considered of poor biological relevance (EFSA 2011). To the end of decision support in sustainable pest control, it is important that the simulated trajectories satisfactory interpret the phenology of the pest (i.e., the temporal dynamics of the major events describing the change in the biological stages). A very similar phenology can be observed by comparing the simulated dynamics (median and 2.5–97.5% confidence band) of the *constant* and the *variable approaches*, which is also in agreement with the field data. It follows that both approaches are effective for the purpose of decision support.

Thus, we keep the assumption of constant parameters, as it gives a lower variability in the posterior densities of the extrinsic mortalities, and it has also positive implications from the applicative point of view, because it avoids the need of new estimates for every year. Once the parameter estimation is performed for a certain number of years, we can guess that the obtained posterior distributions are also valid for the subsequent years.

The simulated dynamics (median and 2.5–97.5% confidence band) for the chosen *constant approach* are reported in Fig. 6 for all years 2008, 2009, and 2011. Differently from the case of simulated data, a certain amount of observations fall outside of the 2.5–97.5% confidence band. However, this does not indicate bad estimation
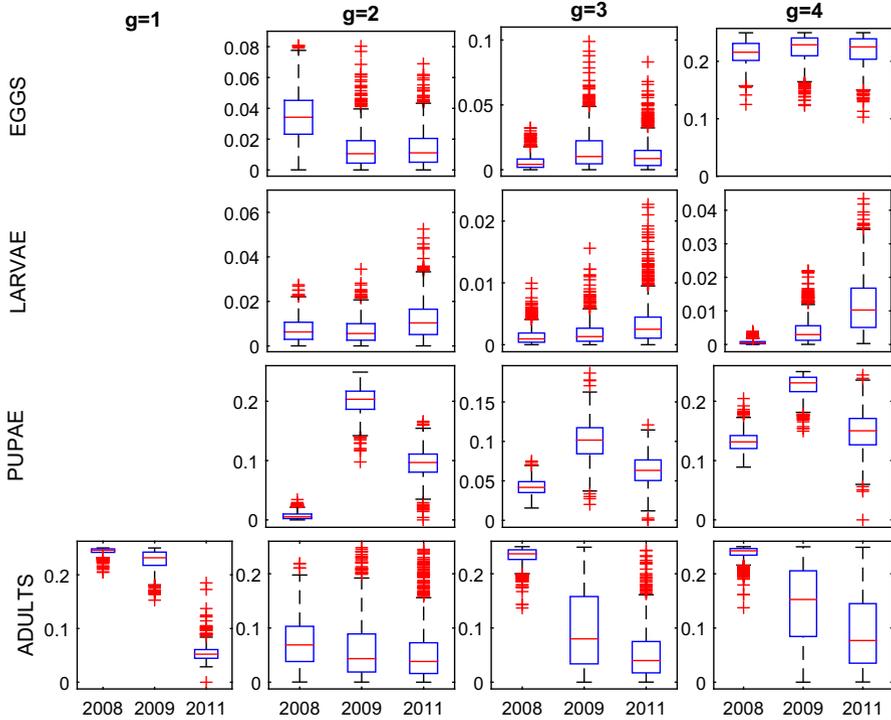
Fig. 5 *Box-plots* of the J_g^i marginal posterior distributions for all years 2008, 2009, and 2011, for the case of field data, under the *variable approach*

performance of the proposed approach, because the field observations are affected by high sampling errors. Thus, the estimated parameters try to compensate the error between the observations and the simulated trajectories of all the stages obtained from the demographic model. As a matter of fact, the results with the simulated data (which are not affected by measurement errors) are highly satisfactory.

Furthermore, as mentioned, in the perspective of a threshold-based decision making for control intervention the model correctly reveals when the abundance crosses the control threshold. For example, using an action threshold for the second larval generation of 15 larvae per 100 berry bunch, the model provides exact indications in all 3 years (CABI 2014).

Comparing the results here obtained with those in Gilioli et al. (2016), we observe that also the simulated trajectories obtained with the least square estimation present overestimation and underestimation of population abundance. Moreover, the phenology of the grape berry moth is well reproduced in both cases, as well as the crossing of the control threshold for larvae. In this paper, differently from Gilioli et al. (2016), we test whether the mortality parameters are constant or vary among years, finding no evidence that they differ from year to year. Moreover, thanks to the Bayesian setting, we have here obtained the entire probability density functions of the estimates, and the confidence bands for the trajectories have been directly generated from the MCMC
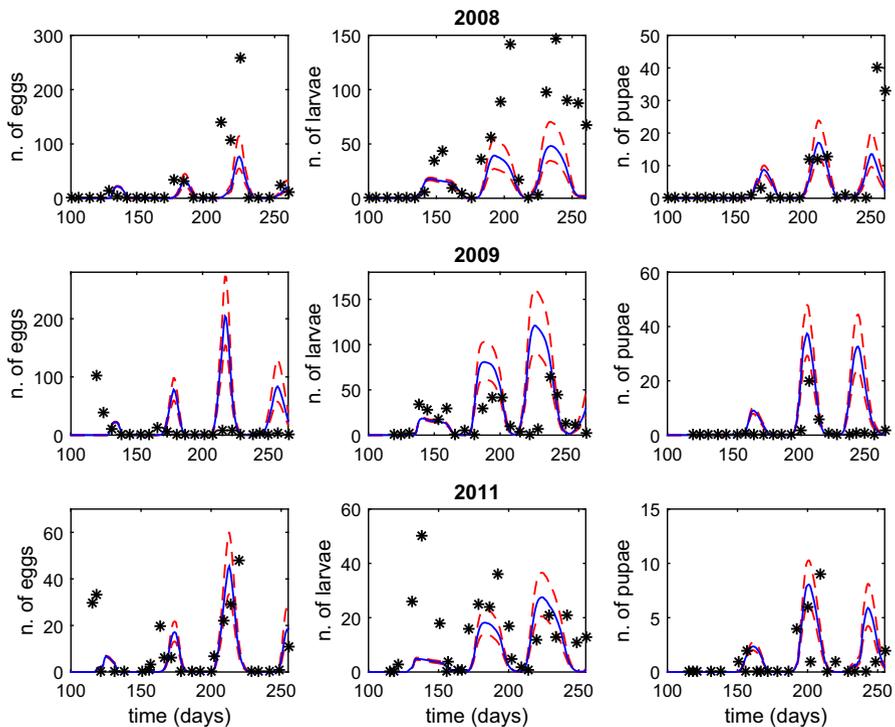
Fig. 6 Simulated dynamics of *L. botrana* for all years 2008 (*first row*), 2009 (*second row*), and 2011 (*third row*), for the case of field data, using the *constant approach*: median (*continuous line*) and 2.5–97.5% confidence bands (*dashed lines*). *Asterisks* represent the field data collected in Colognola ai Colli

parameter estimation. Data collected for a new year can be used without redoing the estimation for the past observations. In fact, under the *constant approach*, every year we consider only the observations of that year, while information coming from previous observations are summarized in the prior distribution. With the least square method, estimation has to be performed every year considering all the observations.

6 Discussion and conclusions

In this paper we consider a stage-structured demographic model describing the dynamics of one species, and we propose a Bayesian approach to estimate a mortality term for each stage due to biotic factors, namely the extrinsic mortality term. The estimation method is sufficiently general to be used for all species whose dynamics can be described by the model presented in Sect. 2, and it can be applied also when abundance data are available only for some stages.

Some advantages derive from the adoption of the Bayesian framework. If some knowledge about the extrinsic mortality parameters is available, it can be summarized into a prior distribution that, suitably updated with the observations on the dynamics, gives a posterior distribution for each unknown parameter. The posterior distribution

also allows to evaluate the uncertainty associated to each parameter. This uncertainty, in the proposed constant multi-year approach, can be also reduced by using observations from different years. In such approach, we update the knowledge of the mortality parameters year by year, by deriving the prior density at a given year from the posterior of the previous one and exploiting only the current year observations in the likelihood function.

To illustrate the use of the proposed method, an application to a pest of vineyards, the grape berry moth, has been considered. We started from Uniform prior distributions, because the only available information on the parameters to estimate was the admissible range. In a first analysis with simulated data, the convergence of the estimation method is observed while assuming the parameters to be constant (*constant approach*). In a second analysis with field data, we inferred the behaviour of the parameters and we found that the *constant* and the *variable approaches* give similar results in terms of phenology of the grape berry moth. Thus, we opted for the assumption of constant parameters, which simplifies the estimation process and allows us to fully exploit the above mentioned advantages of the Bayesian approach. In this way, the parameter distributions obtained for the last year can be updated using the new collected observations and, after a certain number of years, we obtain parameter posterior distributions that can be considered valid for all subsequent years and are characterized by narrowed confidence intervals. This should improve management decision process after some years.

Moreover, from the biological point of view, the assumption of constant parameters seems to support the hypothesis of a well-established community of natural enemies which provide a significant contribution to natural pest control with relatively low inter-annual variability.

Finally we remark that, with the estimation method here proposed, data collected in a year can be used only at the end of the year. Other real-time approaches are also possible, based, e.g., on particle filters, alternatively known as sequential Monte Carlo methods. They allow to use an observation as soon as it becomes available, and to follow the temporal variation of the estimated parameters (Martín-Fernández et al. 2014; Martín-Fernández and Lanzarone 2015).

References

Baumgärtner J, Baronio P (1988) Modello fenologico di volo di *Lobesia botrana* Den. et Schiff. (Lep. Tortricidae) relativo alla situazione ambientale della Emilia-Romagna. Bollettino dell'Istituto di Entomologia. 'Guido Grandi' Dell'Università di Bologna 43:157–170

Bieri M, Baumgärtner J, Bianchi G, Delucchi V, Von Arx R (1983) Development and fecundity of pea aphid (Acyrthosiphon pisum Harris) as affected by constant temperatures and by pea varieties. Mitt Schweiz Entomol Ges 56:163–171

Briolini G, Di Cola G, Gilioli G (1997) Stochastic model for population development of *L. botrana* (Den. et Schiff.). IOBC/WPRS Bull 21:79–81

Buffoni G, Pasquali S (2007) Structured population dynamics: continuous size and discontinuous stage structures. J Math Biol 54(4):555–595

CABI Invasive Species Compendium (2014) *Lobesia botrana* (grape berry moth). www.cabi.org/isc/datasheet/42794. Accessed 14 July 2016

Di Cola G, Gilioli G, Baumgärtner J (1999) Mathematical models for age-structured population dynamics. In: Huffaker CB, Gutierrez AP (eds) Ecological entomology. Wiley, New York, pp 503–534

Dorazio RM (2016) Bayesian data analysis in population ecology: motivations, methods, and benefits. Popul Ecol 58:3144

Duane S, Kennedy AD, Pendleton BJ, Roweth D (1987) Hybrid monte carlo. Phys Lett B 195(2):216–222

Durham GB, Gallant AR (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. J Bus Econ Stat 20:297–338

EFSA Scientific Commettee (2011) Statistical significance and biological relevance. EFSA J 9(9):2372

Elerian O, Chib S, Shephard N (2001) Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69:959–993

Gardiner CW (1994) Handbook of stochastic methods. Springer, Berlin

Gilioli G, Pasquali S, Marchesini E (2016) A modelling framework for pest population dynamics and management: an application to the grape berry moth. Ecol Model 320:348–357

Gilioli G, Pasquali S, Ruggeri F (2008) Bayesian inference for functional response in a stochastic predator–prey system. Bull Math Biol 70:358–381

Gilioli G, Pasquali S, Ruggeri F (2012) Nonlinear functional response parameter estimation in a stochastic predator–prey model. Math Biosci Eng 9(1):75–96

Gillespie CS, Golightly A (2010) Bayesian inference for generalized stochastic population growth models with applications to aphids. J R Stat Soc 59(2):341–357

Golightly A, Wilkinson D (2005) Bayesian inference for stochastic kinetic models using a diffusion approximation. Biometrics 61:781–788

Golightly A, Wilkinson D (2008) Bayesian inference for nonlinear multivariate diffusion models observed with error. Comput Stat Data Anal 52:1674–1693

Gutierrez AP, Ponti L, Cooper ML, Gilioli G, Baumgärtner J (2012) Prospective analysis of the invasive potential of the European grapevine moth *L. Botrana* (De. & Schiff.) in California. Agric For Entomol 14:225–238

Heydari J, Lawless C, Lydall DA, Wilkinson DJ (2014) Fast Bayesian parameter estimation for stochastic logistic growth models. Biosystems 122:55–72

Hooten MB, Wikle CK (2008) A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian collared-dove. Environ Ecol Stat 15:59–70

Huffaker CB, Gutierrez AP (eds) (1999) Ecological entomology. Wiley, New York

Lactin DJ, Holliday NJ, Johnson DL, Craigen R (1995) Improved rate model of temperature-dependent development by arthropods. Environ Entomol 24:68–75

Lamonica D, Herbach U, Orias F, Clément B, Charles S, Lopes C (2016a) Mechanisitc modelling of daphnid–algae dynamics within a laboratory microcosm. Ecol Model 320:213–230

Lamonica D, Clément B, Charles S, Lopes C (2016b) Modelling algae–duckweed interaction under chemical pressure within a laboratory microcosm. Ecotoxicol Environ Saf 128:252–265

Lanzarone E, Pasquali S, Mussi V, Ruggeri F (2014) Bayesian estimation of thermal conductivity and temperature profile in a homogeneous mass. Numer Heat Transf Part B Fundam 66(5):397–421

Marchesini E (2007) I limitatori naturali della tignoletta nell'agroecosistema vigneto. In: Anfora G et al (eds) Le tignole della vite. Agricoltura Integrata, Istituto Agrario di San Michele all'Adige, pp 40–55

Marchesini E, Dalla Montà L (2004) Nel Veneto quattro generazioni di tignoletta della vite. Inf Agrar 60(4):75–78

Martín-Fernández L, Gilioli G, Lanzarone E, Míguez J, Pasquali S, Ruggeri F, Ruiz DP (2014) A Rao–Blackwellized particle filter for joint parameter estimation and biomass tracking in a stochastic predator–prey system. Math Biosci Eng 11(3):573–597

Martín-Fernández L, Lanzarone E (2015) A particle-filtering approach for real-time estimation of thermal conductivity and temperature tracking in homogeneous masses. Numer Heat Transf Part B Fundam 67(6):507–530

McDonald L, Manly B, Lockwood J, Logan JA (eds) (1989) Estimation and analysis of insect populations. Springer, Berlin

Metz JAJ, Diekmann O (eds) (1986) The dynamics of physiologically structured populations. Springer, Berlin

Pavan F, Floreani C, Barro P, Zandigiacomo P, Dalla Montà L (2010) Influence of generation and photoperiod on larval development of *L. Botrana* (Lepidoptera: Tortricidae). Environ Entomol 39(5):1652–1658

Plummer M (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Hornik K, Leisch F, Zeileis A (eds) Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003). Technische Universität Wien, Vienna, pp 20–22

Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. R News 6:711

Prakasa Rao BLS (2000) Bayes estimation for some stochastic partial differential equations. J Stat Plan Inference 91(2):511–524

Roques L, Walker E, Franck P, Soubeyrand S, Klein EK (2016) Using genetic data to estimate diffusion rates in heterogeneous landscape. J Math Biol. doi:10.1007/s00285-015-0954-4

Stan Development Team (2015) Stan modeling language users guide and reference manual (version 2.9.0) http://mc-stan.org

Spence MA, Blackwell PG, Blanchard JL (2016) Parameter uncertainty of a dynamic multispecies size spectrum model. Can J Fish Aquat Sci 73(4):589–597

Tsehaye I, Jones ML, Bence JR, Brenden TO, Madenjian CP, Warner DM (2014) A multispecies statistical age-structured model to assess predator–prey balance: application to an intensively managed Lake Michigan pelagic fish community. Can J Fish Aquat Sci 71(4):627–644

Xun X, Cao J, Mallick B, Maity A, Carroll RJ (2013) Parameter estimation of partial differential equation models. J Am Stat Assoc 108(503):1009–1020