



HAL
open science

Bayesian inference of natural selection from spatiotemporal phenotypic data

Olivier David, Gaëlle van Frank, Isabelle Goldringer, Pierre Rivière, Michel
Turbet Delof

► **To cite this version:**

Olivier David, Gaëlle van Frank, Isabelle Goldringer, Pierre Rivière, Michel Turbet Delof. Bayesian inference of natural selection from spatiotemporal phenotypic data. *Theoretical Population Biology*, 2020, 131, pp.100-109. 10.1016/j.tpb.2019.11.007 . hal-03641603

HAL Id: hal-03641603

<https://hal.inrae.fr/hal-03641603>

Submitted on 14 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Bayesian inference of natural selection from spatiotemporal phenotypic data

Olivier David^{a,*}, Gaëlle van Frank^b, Isabelle Goldringer^b, Pierre Rivière^c, Michel Turbet Delof^b

^a*MaIAGE, INRA, Université Paris-Saclay,
78350, Jouy-en-Josas, France*

^b*Génétique Quantitative et Evolution - Le Moulon,
INRA, Université Paris-Saclay, Université Paris-Sud, CNRS, AgroParisTech,
91190, Gif-sur-Yvette, France*

^c*Réseau Semences Paysannes,
47190, Aiguillon, France*

Abstract

Spatiotemporal variations of natural selection may influence the evolution of various features of organisms such as local adaptation or specialisation. This article develops a method for inferring how selection varies between locations and between generations from phenotypic data. It is assumed that generations are non-overlapping and that individuals reproduce by selfing or asexually. A quantitative genetics model taking account of the effects of stabilising natural selection, the environment and mutation on phenotypic means and variances is developed. Explicit results on the evolution of populations are derived and used to develop a Bayesian inference method. The latter is applied to simulated data and to data from a wheat participatory plant breeding programme. It has some ability to infer evolutionary parameters, but estimates may be sensitive to prior distributions, for example when phenotypic time series are short and when environmental effects are large. In such cases, sensitivity to prior distributions may be reported or more data may be collected.

Keywords: Adaptation, Evolution, Quantitative genetics, Statistics, Wheat

*Correspondence: O. David, Unité MaIAGE, Bât. 210, INRA, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France. E-mail: Olivier.David@inrae.fr

This manuscript is a postprint of the article: David, O., van Frank, G., Goldringer, I., Rivière, P. and Turbet Delof, M. (2020) Bayesian inference of natural selection from spatiotemporal phenotypic data. *Theoretical Population Biology*, 131, 100–109. <https://doi.org/10.1016/j.tpb.2019.11.007>

©2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Selection is the process by which favourable types become more frequent in successive generations (Barton et al., 2007, p. 457). Selective pressures may vary in space and in time (Lande and Shannon, 1996; Papaïx et al., 2011; Troth et al., 2018; Garnault et al., 2019) and these variations may influence the evolution of various features of organisms, for example local adaptation (Kawecki and Ebert, 2004; Débarre et al., 2013; David et al., 2017), specialisation (Poisot et al., 2011), dispersal (Friedenberg, 2003) or phenotypic heterogeneity (Sasaki and Ellner, 1995; Ackermann, 2015). Thus, these variations are involved in numerous practical issues, for example they may generate biological diversity in particular when considering divergent selection in subdivided populations (Dobzhansky, 1973; Kocher, 2004; Enjalbert et al., 2011), they may allow some genotypes to better develop in some environments (Rhoné et al., 2008; Gautier et al., 2009; Pariaud et al., 2009; Fabre et al., 2015; Ropars et al., 2016) or they may be involved in the emergence of resistances to pesticides and drugs (Papaïx et al., 2011; REX Consortium, 2013; Garnault et al., 2019). Natural selection has a strong impact on the management and breeding of crop species for instance in crop diversity conservation. In *in situ* conservation approaches, defined in the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA 2009) as "the conservation of ecosystems and natural habitats and the maintenance and recovery of viable populations of species in their natural surroundings and, in the case of domesticated or cultivated plant species, in the surroundings where they have developed their distinctive properties", populations of wild or domesticated species are grown over time in environments submitted to all evolutionary forces including natural selection (Rhoné et al., 2008; Enjalbert et al., 2011). Such populations evolve and may adapt to changing climatic conditions, emerging diseases and agricultural practices depending on the type, strength and direction of selection. Based on this principle, different strategies have been proposed focusing on the conservation and renewing of crop genetic diversity (dynamic management, Enjalbert et al., 2011) or on the improvement of the population (evolutionary plant breeding, Suneson, 1956). More recently, decentralised participatory breeding approaches have developed to produce populations adapted to local contexts and practices with low input requirements in traditional or industrialised agricultural contexts (Sperling et al., 2001; Ceccarelli et al., 2001; Dawson et al., 2011). Characterisation of selective pressures on these populations is important to anticipate their evolution and optimise their management.

Selection can be inferred in various ways, in particular using retrospective methods (Barton et al., 2007, Chap. 19; Lefèvre et al., 2016, p. 665). This approach consists in measuring quantitative traits on individuals sampled in a population at several time points. Phenotypic evolution is then modelled taking account of selection and parameters are estimated by fitting the model to the data. For example such methods have been used to quantify selection for Darwin's finches (Grant and Grant, 1995). Unlike

prospective methods, they do not require measures of fitness components (Lande and Arnold, 1983; Barton et al., 2007, p. 525).

The main retrospective method that has been used is based on the breeder’s equation (Lande, 1976; Grant and Grant, 1995; Gillespie, 1998, Chap. 5). It assumes that genetic variance is constant, but this assumption may not be satisfied in practice (McGuigan, 2006). A few quantitative genetics models allowing genetic variance to evolve have been developed (Cavalli-Sforza and Feldman, 1976; Débarre et al., 2014; Lande and Porcher, 2015). However they are not simple because multiple effects may affect traits besides selection, in particular environmental effects, mutation or drift. As a result they are not easy to fit to data. Few statistical methods have been developed to fit such evolutionary models to phenotypic time series.

This article considers populations evolving in several locations. The objective is to develop a Bayesian retrospective method for inferring how selective pressures vary between locations and between generations. For the sake of simplicity, our study is restricted to organisms with non-overlapping generations that reproduce by selfing or asexually, for example autogamous annual plants. Individuals are characterised by a single trait that evolves under the effect of natural selection and mutation. As the model we want to use is not available in the literature, we develop our own model (Section 2). Model parameters are estimated using a Bayesian framework (Section 3). This method is assessed using simulated data (Section 4) and is applied to data from a wheat participatory plant breeding programme (Section 5).

2. Model

2.1. Notations and assumptions

We consider populations located in several sites and model their phenotypic evolution under the effect of mutation and selection. Selective pressures may vary between locations and between generations. Locations are indexed by $i \in \{1, \dots, I\}$. Populations are assumed to be nested within locations and are indexed by (i, j) , $j \in \{1, \dots, J_i\}$. Populations are assumed to be isolated and large enough that demographic stochasticity (or drift) can be neglected.

The life cycle is divided into a reproduction phase and a growth one (Fig. 1). For the sake of clarity, reproduction is divided into a birth phase and a mutation one, although this distinction is only formal. Generations are non-overlapping, are indexed by $t \in \mathbb{N}$ and start after reproduction. Thus, an offspring of an adult of generation t belongs to generation $t + 1$ after the mutation phase but still belongs to generation t after the birth phase.

In the following, the symbols μ , σ^2 and τ^2 are often used to denote the expectation, variance and inverse of the variance of a random variable respectively, using subscripts reminding the random variable, dropping unnecessary subscripts and avoiding double subscripts. For example the expectation of Z_{ijtk} is denoted by μ_{Zijt} rather than by $\mu_{Z_{ijtk}}$,

if it depends on i , j and t and if it does not depend on k . We let $p(X = x)$ denote the value of the density of the random variable X at point x . We use the common notation $\mathcal{N}(\mu, \sigma^2)$ for the normal distribution with mean μ and variance σ^2 .

2.2. Phenotype

Adults are characterised by an adaptive trait $Z \in \mathbb{R}$ affecting fitness. This trait is assumed to be the sum of genetic and environmental effects (Gillespie, 1998, p. 105; Barton et al., 2017). Consider an adult individual sampled in population (i, j) at generation t . Its trait Z_{ijt} is assumed to be equal to

$$Z_{ijt} = G_{ijt} + A_{it} + B_{ijt} + C_{ijt}.$$

In this equation, G_{ijt} represents a random genetic effect. A_{it} represents an environmental effect that is common to all the individuals of location i . B_{ijt} is an environmental effect that is common to all the individuals of population (i, j) and that is peculiar to this population. The effects A_{it} and B_{ijt} can be regarded as fixed parameters, for which we will assume a prior distribution in Section 3. C_{ijt} represents a local environmental effect that is peculiar to the individual and that is assumed to be normally distributed:

$$C_{ijt} \sim \mathcal{N}(0, \sigma_{C_{ijt}}^2).$$

Note that the variance $\sigma_{C_{ijt}}^2$ is assumed to depend on the population and the generation. The effects G_{ijt} and C_{ijt} are assumed to be independent, thus, the expectation $\mu_{Z_{ijt}}$ and variance $\sigma_{Z_{ijt}}^2$ of Z_{ijt} satisfy

$$\mu_{Z_{ijt}} = \mu_{G_{ijt}} + A_{it} + B_{ijt}, \quad \sigma_{Z_{ijt}}^2 = \sigma_{G_{ijt}}^2 + \sigma_{C_{ijt}}^2. \quad (1)$$

2.3. Birth

During the birth phase, adults produce offsprings and die. Individuals are assumed to reproduce asexually or by selfing. Fecundity is assumed to have a normal shape (Cavalli-Sforza and Feldman, 1976; Johnson and Barton, 2005; Zhang and Hill, 2005). Thus, an adult with phenotype $z \in \mathbb{R}$ in location i at generation t gives birth to $F_{it}(z)$ offsprings, where $F_{it}(z)$ satisfies

$$\begin{aligned} F_{it}(z) &\propto \exp\left(-\tau_{F_i}^2 (z - \mu_{F_{it}})^2 / 2\right), \\ \mu_{F_{it}} &= \theta_i + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2). \end{aligned}$$

This fecundity function leads to a stabilising selection toward the optimal phenotype $\mu_{F_{it}}$ with a strength quantified by τ_{F_i} . Optimal phenotypes fluctuate randomly around θ_i , which represents the mean optimal phenotype over generations for location i (Sasaki and Ellner, 1995; Scheiner, 2014). Note that F_{it} is a deterministic function of the parental phenotype given ε_{it} and that populations in a given location are assumed to undergo the same selective pressures. Selective pressures are allowed to vary spatially, since τ_{F_i} and $\mu_{F_{it}}$ depend on i , and temporally, since $\mu_{F_{it}}$ depends on t .

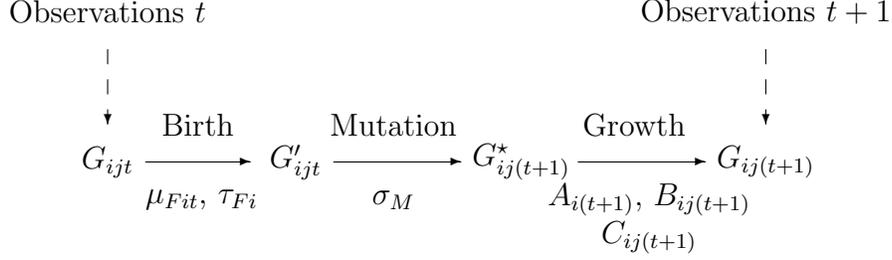


Figure 1: Life cycle diagram, observation times and genetic composition for population j of location i . The life cycle is decomposed into birth, mutation and growth phases. The population is observed when individuals are adults. Individuals are characterised by an adaptive trait that has a genetic component. The random variable G_{ijt} is the genetic effect of an adult sampled in the population at generation t . The random variables G'_{ijt} and $G^*_{ij(t+1)}$ are the genetic effects before and after mutation respectively of an offspring sampled in the population at the end of generation t . Selection favours individuals with trait values close to μ_{Fit} with a strength of τ_{Fi} . Mutation effects with standard deviation σ_M increase genetic variance. Environmental effects $A_{i(t+1)}$, $B_{ij(t+1)}$ and $C_{ij(t+1)}$ occur during growth.

2.4. Mutation and growth

After birth, offspring genetic effects are modified by mutation. Their distribution after mutation is assumed to be normal around their value before mutation with variance σ_M^2 . During growth, individuals undergo environmental effects, express their phenotype and are assumed to survive with the same probability. Thus, the survival probability does not influence the evolution of populations.

2.5. Genetic evolution

The distribution of genetic effects evolves under the effect of selection and mutation (Fig. 1). For $t \in \mathbb{N}$, $i \in \{1, \dots, I\}$ and $j \in \{1, \dots, J_i\}$, consider an adult sampled in population (i, j) at generation t with genetic effect G_{ijt} and trait Z_{ijt} . Likewise consider an offspring sampled independently in the population at the end of generation t after birth and let G'_{ijt} denote its genetic effect. Individuals are assumed to be homozygous or haploid and to transmit their genetic effects to their offsprings but not their environmental effects. Thus, for $g_{ijt} \in \mathbb{R}$, the probability density of genetic effects after birth satisfies

$$\begin{aligned}
 p(G'_{ijt} = g_{ijt}) &= \frac{\int_{\mathbb{R}} F_{it}(z_{ijt}) p(Z_{ijt} = z_{ijt}, G_{ijt} = g_{ijt}) dz_{ijt}}{\int_{\mathbb{R}} \int_{\mathbb{R}} F_{it}(z_{ijt}) p(Z_{ijt} = z_{ijt}, G_{ijt} = g_{ijt}) dz_{ijt} dg_{ijt}}, \\
 &= \frac{w_{it}(g_{ijt}) p(G_{ijt} = g_{ijt})}{\int_{\mathbb{R}} w_{it}(g_{ijt}) p(G_{ijt} = g_{ijt}) dg_{ijt}}, \tag{2}
 \end{aligned}$$

where $w_{it}(g_{ijt}) = \int_{\mathbb{R}} F_{it}(z_{ijt}) p(Z_{ijt} = z_{ijt} | G_{ijt} = g_{ijt}) dz_{ijt}$ and $p(Z_{ijt} = z_{ijt} | G_{ijt} = g_{ijt})$ denotes the density of Z_{ijt} given $\{G_{ijt} = g_{ijt}\}$.

For $g_{ij(t+1)} \in \mathbb{R}$, the density of genetic effects after mutation writes

$$p(G^*_{ij(t+1)} = g_{ij(t+1)}) = \int_{\mathbb{R}} p(G^*_{ij(t+1)} = g_{ij(t+1)} | G'_{ijt} = g_{ijt}) p(G'_{ijt} = g_{ijt}) dg_{ijt},$$

where $G_{ij(t+1)}^*$ is the genetic effect after mutation of the sampled offspring, with distribution $\mathcal{N}(g_{ijt}, \sigma_M^2)$ given $\{G'_{ijt} = g_{ijt}\}$. The distribution of genetic effects does not change during growth.

If the initial distribution of genetic effects in the adult population is normal:

$$G_{ij0} \sim \mathcal{N}(\mu_{G_{ij0}}, \sigma_{G_{ij0}}^2),$$

then the distributions of genetic effects among adults in the following generations remain normal (Supplementary material A):

$$\forall t \in \mathbb{N}, G_{ij(t+1)} \sim \mathcal{N}(\mu_{G_{ij(t+1)}}, \sigma_{G_{ij(t+1)}}^2), \quad (3)$$

where $\mu_{G_{ij(t+1)}}$ and $\sigma_{G_{ij(t+1)}}^2$ satisfy the following recursions:

$$\mu_{G_{ij(t+1)}} = \frac{\tau_{G_{ijt}}^2 \mu_{G_{ijt}} + \tau_{S_{ijt}}^2 (\mu_{Fit} - A_{it} - B_{ijt})}{\tau_{G_{ijt}}^2 + \tau_{S_{ijt}}^2}, \quad (4)$$

$$\tau_{G_{ij(t+1)}}^2 = \frac{(\tau_{G_{ijt}}^2 + \tau_{S_{ijt}}^2) \tau_M^2}{\tau_{G_{ijt}}^2 + \tau_{S_{ijt}}^2 + \tau_M^2}, \quad (5)$$

$$\tau_{S_{ijt}}^2 = \frac{\tau_{C_{ijt}}^2 \tau_{Fi}^2}{\tau_{C_{ijt}}^2 + \tau_{Fi}^2},$$

where $\tau_{G_{ijt}}^2 = \sigma_{G_{ijt}}^{-2}$, $\tau_M^2 = \sigma_M^{-2}$, $\tau_{S_{ijt}}$ is the overall selection efficiency and $\tau_{C_{ijt}}^2 = \sigma_{C_{ijt}}^{-2}$. Overall selection efficiency increases with selection strength τ_{Fi} and decreases with environmental variance $\sigma_{C_{ijt}}^2$. Genetic variance $\sigma_{G_{ij(t+1)}}^2$ decreases with selection efficiency $\tau_{S_{ijt}}$ and increases with mutation variance σ_M^2 . In the recursions above, it is assumed that $\tau_{G_{ij0}} > 0$ and $\tau_M > 0$, so that the ratios are well defined. However we allow τ_{Fi} to be equal to 0, which corresponds to the case when evolution is neutral.

3. Bayesian inference

3.1. Observations

Our objective is to estimate model parameters from phenotypic data. For $t \in \{0, \dots, T\}$, $i \in \{1, \dots, I\}$ and $j \in \{1, \dots, J_i\}$, $K_{ijt} > 1$ adults are sampled independently in population (i, j) at generation t and their trait is measured (Fig. 1). Measurement errors are not confounded with local environmental effects since they do not affect trait values and do not interact with selection. However it is difficult to distinguish them from genetic and environmental effects without replicating measurements on some individuals. Thus, measurement errors are here neglected.

For $k \in \{1, \dots, K_{ijt}\}$, the observation Z_{ijtk} for individual k follows a normal distribution:

$$Z_{ijtk} \sim \mathcal{N}(\mu_{Z_{ijt}}, \sigma_{Z_{ijt}}^2),$$

where the expressions for $\mu_{Z_{ijt}}$ and $\sigma_{Z_{ijt}}^2$ result from (1), (4) and (5).

3.2. Prior distribution

A possible joint prior distribution of model parameters is composed of the following independent distributions. Vague normal prior distributions are placed on mean optimal phenotypes and initial genetic means:

$$\theta_i \sim \mathcal{N}(\alpha_\theta, \beta_\theta^2), \quad \mu_{G_{ij}0} \sim \mathcal{N}(\alpha_G, \beta_G^2),$$

where the parameters of these distributions are known constants.

Vague prior distributions on standard deviations are usually specified as half-Student or half-normal distributions (Gelman, 2006). Thus, vague half-normal prior distributions are placed on the standard deviations of mutation effects and of the effects affecting optimal phenotypes:

$$\sigma_M \sim \mathcal{N}^+(0, \gamma_M^2), \quad \sigma_\varepsilon \sim \mathcal{N}^+(0, \gamma_\varepsilon^2),$$

where γ_M and γ_ε are known constants and $\mathcal{N}^+(\mu, \sigma^2)$ is the normal distribution restricted to positive values with parameters μ and σ .

Hierarchical distributions are placed on between-population environmental effects (Robert, 2007, p. 460):

$$\begin{aligned} A_{it} &\sim \mathcal{N}(0, \sigma_A^2), & \sigma_A &\sim \mathcal{N}^+(0, \gamma_A^2), \\ B_{ijt} &\sim \mathcal{N}(0, \sigma_B^2), & \sigma_B &\sim \mathcal{N}^+(0, \gamma_B^2), \end{aligned}$$

where γ_A and γ_B are known constants. Such hierarchical distributions shrink estimates (Robert, 2007, p. 472).

The prior distributions of selection strengths are chosen to allow these parameters to take low values when evolution is neutral and large values when selection is present. Assuming that selection strengths follow vague log-normal distributions does not seem satisfactory in the neutral case since $\ln(\tau_{Fi}) \rightarrow -\infty$ when $\tau_{Fi} \rightarrow 0$. Thus, half-normal prior distributions are placed on selection strengths:

$$\tau_{Fi} \sim \mathcal{N}^+(0, \gamma_F^2),$$

which place a substantial mass on low values. Thus, a large estimate of τ_{Fi} in an analysis will be due to the data rather than the prior distribution. However, our analyses showed that posterior distributions could be sensitive to the value of γ_F used. Thus, rather than considering γ_F as a known constant, it is considered as an unknown parameter that is estimated from the data (Gelman, 2006). It is written as

$$\gamma_F = \omega / \tilde{\sigma}_Z,$$

where $\tilde{\sigma}_Z$ is a prior value for the overall phenotypic standard deviation, which takes account of all possible sources of variation, and ω is an unknown parameter that is assumed to follow a vague half-normal prior distribution:

$$\omega \sim \mathcal{N}^+(0, \gamma_\omega^2),$$

where γ_ω is a known constant.

As only the phenotypic variances are observed, the standard deviations of local environmental effects may be difficult to separate from the genetic standard deviations. As these parameters are numerous, a hierarchical distribution is placed on these parameters to facilitate their estimation:

$$\begin{aligned}\ln(\sigma_{Cijt}) &\sim \mathcal{N}(\ln(\rho_i), \gamma_C^2), \\ \rho_i &\sim \mathcal{N}^+(0, \kappa^2), \quad \gamma_C \sim \mathcal{N}^+(0, \eta^2),\end{aligned}$$

where κ and η are known constants and ρ_i is the median of σ_{Cijt} . This prior distribution shrinks the estimates of σ_{Cijt} toward ρ_i , thereby providing more robust estimates (Robert, 2007, p. 472).

Likewise as only the phenotypic variances are observed, the initial genetic standard deviations may be difficult to estimate. Thus, these parameters are not assumed to follow vague prior distributions but are assumed to be approximately proportional to the medians of the standard deviations of local environmental effects:

$$\sigma_{Gij0} \sim \mathcal{N}^+(\lambda \rho_i, \gamma_G^2),$$

where λ is a known constant and γ_G is a known constant with a moderate value. For example when λ is fixed to one, initial genetic standard deviations and environmental ones are assumed to have similar orders of magnitude. An example of set of prior distributions is given in Supplementary material E.

3.3. Posterior distribution

In the Bayesian framework, the prior distribution is updated into a posterior distribution, that writes (Robert, 2007, p. 22)

$$p(\phi | y) = \frac{p(\phi) p(y | \phi)}{p(y)},$$

where ϕ denotes the vector of model parameters, y the vector of observations, $p(\phi)$ the prior distribution of ϕ , $p(y | \phi)$ the likelihood and $p(y)$ the marginal distribution of y . This posterior distribution cannot be calculated explicitly for our model but it can be estimated using Markov Chain and Monte Carlo (MCMC) methods. Such methods simulate values of ϕ according to a Markov chain that converges in distribution to $p(\phi | y)$ (Robert, 2007, p. 302). MCMC methods are here implemented using the R package `rjags` (Plummer, 2016; R Core Team, 2018). Details on our implementation of MCMC methods are given in Supplementary material B and a script showing how to implement the methods in R is given in Supplementary material C.

4. Simulation study

4.1. Data

The ability of the developed method to infer evolutionary parameters in heterogeneous environments was assessed using simulations. Data were simulated for experimental set-ups similar to that of the wheat application (Section 5). There were T consecutive generations and 10 locations, with 10 populations per location. At each generation, 20 individuals were sampled in each population and their trait was measured. Data were simulated for two quantitative traits using the model of Section 2. For the first (respectively second) trait, parameter values were similar to those of spike weight (respectively plant height) (Section 5). The initial genetic means were equal to

$$\begin{aligned}\mu_{Gij0} &= \mu_{G0} - \delta, \quad j = 1, \dots, 5, \\ \mu_{Gij0} &= \mu_{G0} + \delta, \quad j = 6, \dots, 10,\end{aligned}$$

for $i = 1, \dots, 10$.

Four experimental set-ups, defined by the combinations of two two-level factors, were considered. The first factor quantified the amount of information provided by the set-up, so that the data were either weakly informative or informative. Set-ups providing more information had more generations, smaller between-population environmental variances and a larger δ value. Specifically, parameter values were equal to $T = 4$, $\sigma_A = 0.4$, $\sigma_B = 0.3$, $\gamma_C = 0.2$, $\delta = 0.3$ for the weakly informative set-up and trait 1, to $T = 6$, $\sigma_A = 0.2$, $\sigma_B = 0.2$, $\gamma_C = 0.1$, $\delta = 0.6$ for the informative set-up and trait 1, to $T = 4$, $\sigma_A = 100$, $\sigma_B = 50$, $\gamma_C = 0.2$, $\delta = 100$ for the weakly informative set-up and trait 2, and to $T = 6$, $\sigma_A = 30$, $\sigma_B = 30$, $\gamma_C = 0.1$, $\delta = 200$ for the informative set-up and trait 2. The second factor indicated whether selection was absent or present. When selection was present, $\tau_{Fi} = 1$ (respectively $\tau_{Fi} = 0.003$) for trait 1 (respectively trait 2) and selective pressures varied both spatially and temporally. In particular, mean optimal phenotypes depended on the location: $\theta_i = \mu_{G0} - \Delta$, $i = 1, \dots, 5$, $\theta_i = \mu_{G0} + \Delta$, $i = 6, \dots, 10$. The parameter values used in the simulations are given in Supplementary material D. For each combination of parameter values, four data sets were simulated independently.

4.2. Analysis

The data were analysed using the model of Sections 2 and 3 with $\gamma_\omega = 0.2$. Mutation effects were assumed to be a priori smaller than environmental effects (Lynch, 1988; Houle et al., 1996; Johnson and Barton, 2005), since γ_M had a smaller value than γ_A , γ_B and κ (Supplementary material D). Likewise, γ_ε had a smaller value than γ_A , γ_B and κ to help to separate temporal variations of optimal phenotypes from environmental effects. Parameter λ was fixed to 1. The values of the other parameters of prior distributions are given in Supplementary material D.

The sensitivity of posterior distributions to variations of γ_ω values was investigated. To reduce computing time, the posterior means for $\gamma_\omega = 0.1$ and $\gamma_\omega = 0.3$ were estimated

using importance sampling and the posterior distribution for $\gamma_\omega = 0.2$ as proposal distribution (Carlin and Louis, 2008, p. 182).

The main evolutionary parameters were estimated: selection strengths and mean optimal phenotypes, that characterised spatial variations of selective pressures; σ_ε , that characterised temporal variations of selective pressures; mutation standard deviation. Differences between estimates and true parameter values were quantified using the average absolute error (aae). For example, for selection strengths, experimental set-up s and a given value of γ_ω , aae was equal to

$$aae(s, \gamma_\omega) = \frac{1}{4} \sum_{r=1}^4 \left(\frac{1}{10} \sum_{i=1}^{10} |\hat{\tau}_{Fi}(s, r, \gamma_\omega) - \tau_{Fi}(s)| \right),$$

where $s = 1, \dots, 4$, $\tau_{Fi}(s)$ was selection strength for location i and set-up s and $\hat{\tau}_{Fi}(s, r, \gamma_\omega)$ was the posterior mean of this parameter for set-up s , replication r and parameter value γ_ω .

4.3. Results

The estimates of selection strengths and mutation standard deviation showed lower errors and were less sensitive to variations of prior distributions for informative set-ups than for weakly informative set-ups (Figs. 2 and 3; Figs. D.1 and D.3 in Supplementary material D). For weakly informative set-ups, these estimates both increased with γ_ω , presumably because an increase in selection strengths, that reduced within-population diversities, was counterbalanced by an increase in mutation standard deviation, that increased within-population diversities, in order to give within-population diversities consistent with the data.

Estimation errors were rather large for mean optimal phenotypes (Figs. 2 and 3; Figs. D.2 and D.4 in Supplementary material D) and for σ_ε and trait 2 (Fig. 3). The 95% Highest Posterior Density (HPD) credible intervals of σ_ε were rather wide, for example this interval was equal to [0.000,0.304] (respectively [0.143,99.412]) for trait 1 (respectively trait 2), the informative set-up with selection, replicate 1 and $\gamma_\omega = 0.2$.

5. Application to wheat data

5.1. Data

A participatory plant breeding programme on bread wheat was set up in France by farmers and facilitators from the Réseau Semences Paysannes¹ and researchers from INRA (GQE - le Moulon) to develop heterogeneous populations adapted to farmers' local conditions, practices and objectives (Dawson et al., 2011). Some of these populations were exchanged among farmers. They then evolved several years in the same farms in

¹www.semencespaysannes.org

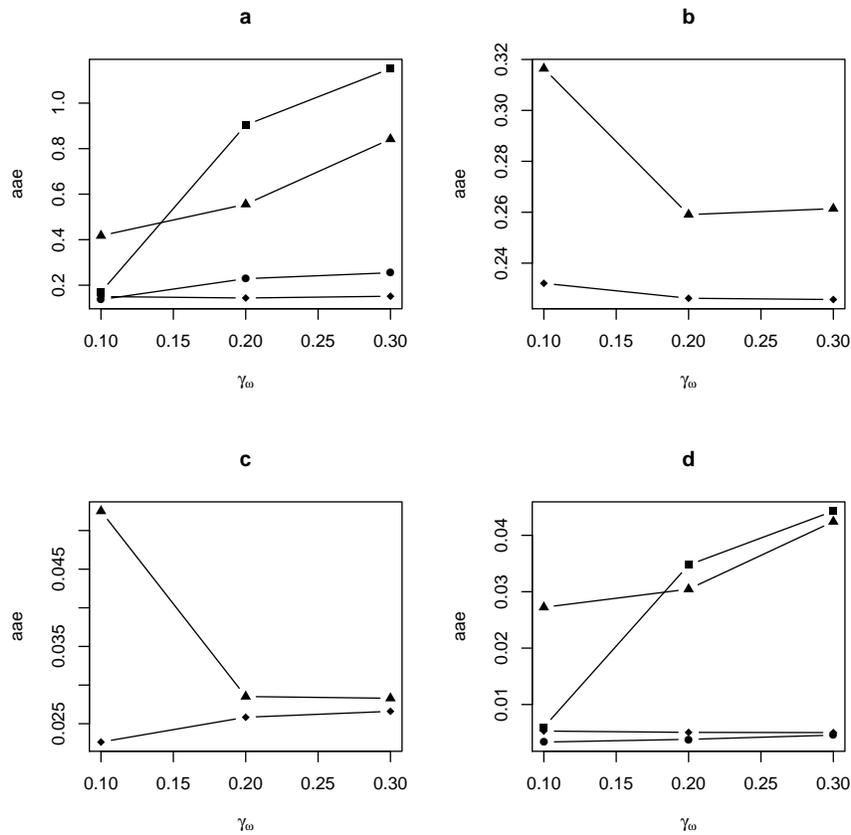


Figure 2: Analysis of simulated data: sensitivity analysis for trait 1, selection strengths (a), mean optimal phenotypes (b), σ_ε (c) and σ_M (d). The quality of estimates is quantified by average absolute error (aae). ■: Weakly informative set-up without selection; ●: Informative set-up without selection; ▲: Weakly informative set-up with selection; ◆: Informative set-up with selection. Estimation errors for mean optimal phenotypes and σ_ε are relevant in the presence of selection only.

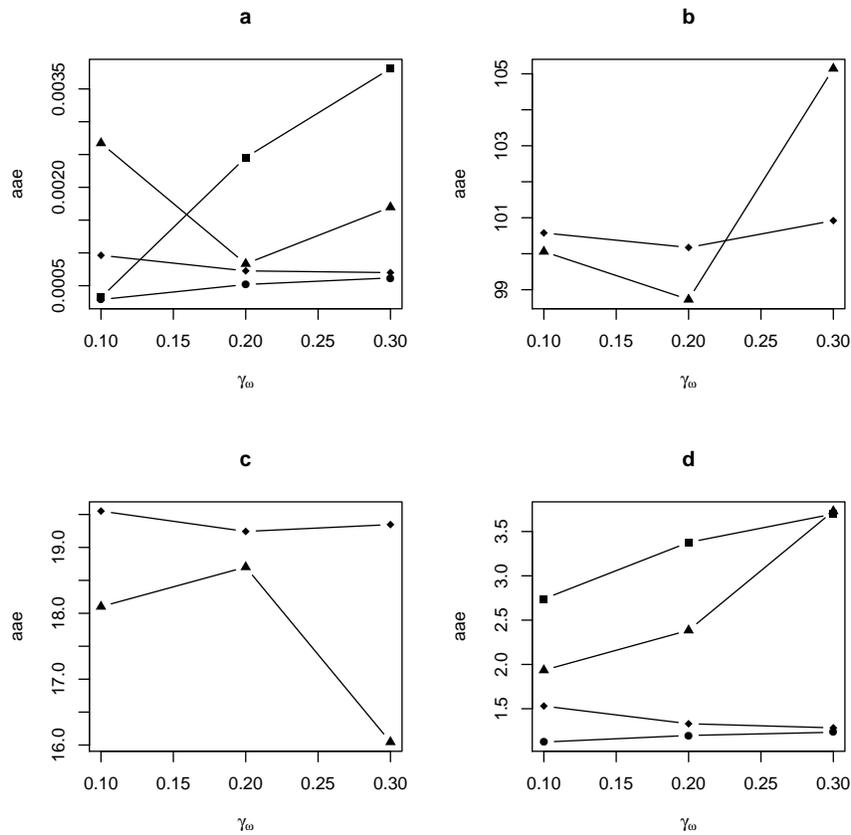


Figure 3: Analysis of simulated data: sensitivity analysis for trait 2, selection strengths (a), mean optimal phenotypes (b), σ_ε (c) and σ_M (d). The quality of estimates is quantified by average absolute error (aae). ■: Weakly informative set-up without selection; ●: Informative set-up without selection; ▲: Weakly informative set-up with selection; ◆: Informative set-up with selection. Estimation errors for mean optimal phenotypes and σ_ε are relevant in the presence of selection only.

Variable	Number of				
	populations	farms	years	trials	observations
Spike weight	106	12	3.5	48	14172
Plant height	86	7	4.3	37	12352

Table 1: Wheat data. The number of years corresponds to the mean number of consecutive years a population is sown in a farm, while a trial is defined as a farm x year combination.

field trials and sometimes underwent farmers’ mass selection. Whenever populations were cultivated on-farm, some characters such as plant height, spike length and spike weight were measured on several individuals per population (10 to 50).

The data used for this analysis consisted of a subset of the data gathered since the beginning of the project in 2006 (Tab. 1). Two variables were analysed: spike weight and plant height. The populations selected (i) evolved for at least three years in a given farm, (ii) were measured for at least three consecutive years, (iii) had at least 10 observations per year and (iv) were not subjected to mass selection. We studied the selective pressures undergone by these populations grown in different environments and the influence of natural selection on their evolution. In this application, locations were farms, trials were farm x year combinations and on average populations were sown and measured for 3.5 and 4.3 consecutive years for spike weight and plant height respectively. Measurement errors, that were neglected in our model, were likely to be smaller for spike weight than for plant height, since the former was measured in the laboratory while the latter was measured in the field. More information on the data collected in each farm is given in Supplementary material E.

5.2. Analysis

As wheat is an autogamous annual plant, the data were analysed using the model of Section 2 and the prior distributions given in Supplementary material E. As in the simulation study, γ_M and γ_ε had a smaller value than γ_A , γ_B and κ , and λ was fixed to 1. Parameter γ_ω took two possible values, $\gamma_\omega = 0.1$ or $\gamma_\omega = 0.4$, in order to study the influence of this parameter on posterior distributions.

5.3. Results

Selection strengths and mean optimal phenotypes. The posterior distributions of τ_{Fi} and θ_i were sensitive to changes in γ_ω values for both variables (Figs. 4 and 5). For spike weight, the HPD credible intervals of selection strengths contained values close to 0 for all the locations when $\gamma_\omega = 0.1$ (Fig. 4, a), so that the presence of selection was not significant. On the contrary when $\gamma_\omega = 0.4$, the presence of selection was significant in 4 locations (Fig. 4, b). The significant estimate of selection strength in location 5 could be a false positive since there were only 6 populations observed for 3 years in this location (Supplementary material E). The posterior medians of mean optimal phenotypes were

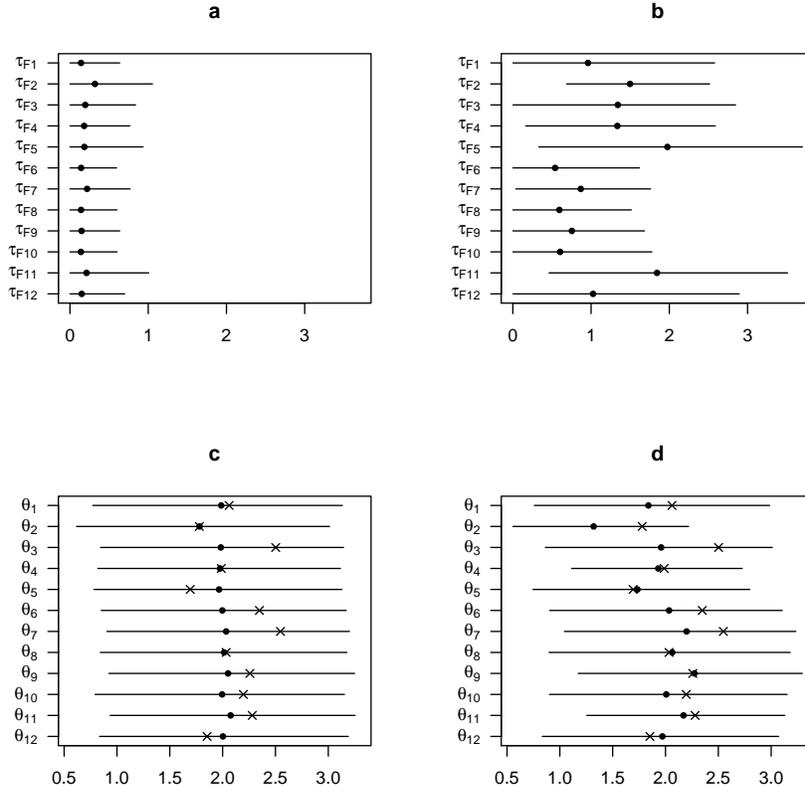


Figure 4: Analysis of wheat data: selection strengths and mean optimal phenotypes for spike weight. Estimates (\bullet) and 95% HPD credible intervals (solid lines) of selection strengths in 12 locations for $\gamma_\omega = 0.1$ (a) and $\gamma_\omega = 0.4$ (b). Sample location means (\times), estimates (\bullet) and 95% HPD credible intervals (solid lines) of mean optimal phenotypes in the same locations for $\gamma_\omega = 0.1$ (c) and $\gamma_\omega = 0.4$ (d). Results are in g for mean optimal phenotypes and in g^{-1} for selection strengths.

close to 2 when $\gamma_\omega = 0.1$, while they ranged from 1.320 to 2.199 and tended to be smaller than sample means when $\gamma_\omega = 0.4$ (Fig. 4, c and d). For plant height, the presence of selection was not significant for both values of γ_ω , but estimates of selection strengths and mean optimal phenotypes varied with this parameter (Fig. 5).

Environmental and mutation standard deviations. When $\gamma_\omega = 0.1$, the credible intervals of σ_A , σ_B and ρ_i did not contain values close to 0 (Fig. 6). The estimates of σ_A were larger than those of σ_B . The estimates of ρ_i appeared more variable between locations for spike weight than for plant height. The posterior median of σ_M was equal to 0.033 (respectively 0.132) when $\gamma_\omega = 0.1$ (respectively $\gamma_\omega = 0.4$) for spike weight. It was equal to 21.316 (respectively 22.705) when $\gamma_\omega = 0.1$ (respectively $\gamma_\omega = 0.4$) for plant height. The credible intervals of σ_ε were rather wide for both values of γ_ω . The 95%

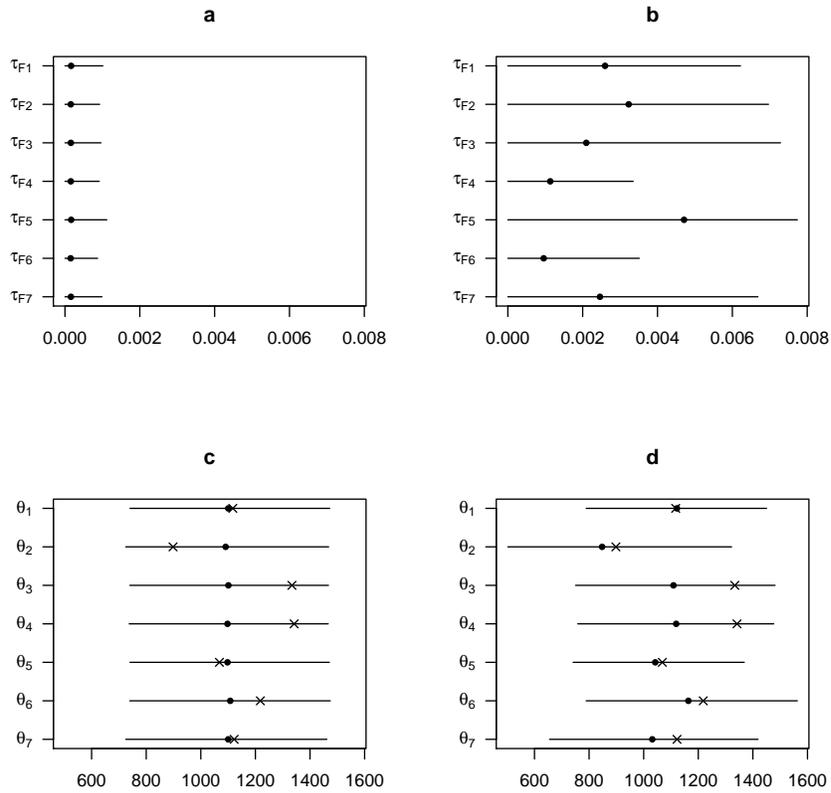


Figure 5: Analysis of wheat data: selection strengths and mean optimal phenotypes for plant height. Estimates (\bullet) and 95% HPD credible intervals (solid lines) of selection strengths in 7 locations for $\gamma_\omega = 0.1$ (a) and $\gamma_\omega = 0.4$ (b). Sample location means (\times), estimates (\bullet) and 95% HPD credible intervals (solid lines) of mean optimal phenotypes in the same locations for $\gamma_\omega = 0.1$ (c) and $\gamma_\omega = 0.4$ (d). Results are in mm for mean optimal phenotypes and in mm^{-1} for selection strengths.

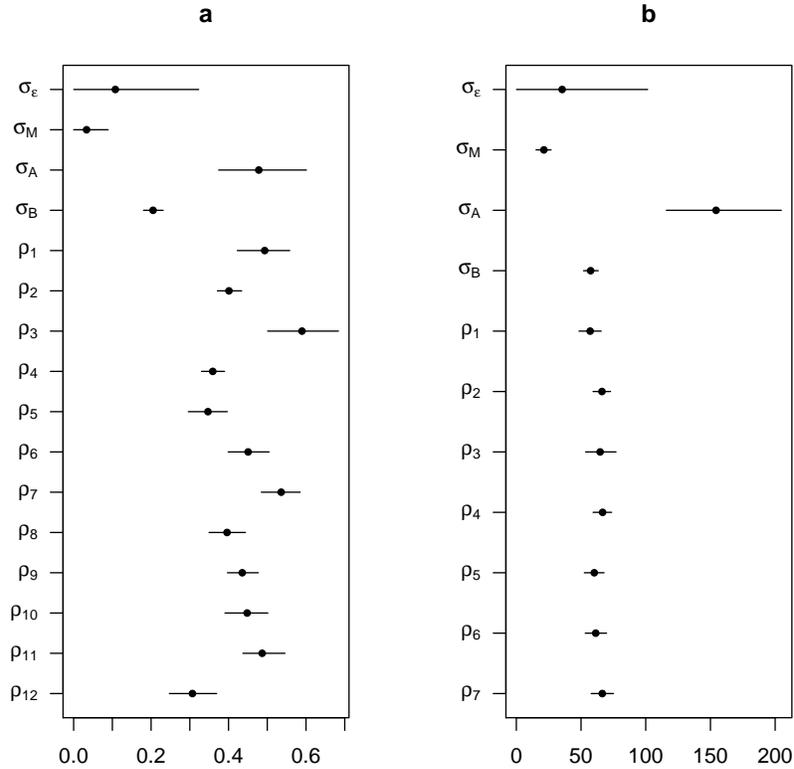


Figure 6: Analysis of wheat data: environmental and mutation standard deviations. Estimates (\bullet) and 95% HPD credible intervals (solid lines) for spike weight (in g, a) and plant height (in mm, b). In these analyses, $\gamma_\omega = 0.1$.

credible interval of γ_C was equal to $[0.341, 0.434]$ (respectively $[0.334, 0.441]$) for spike weight (respectively plant height) and $\gamma_\omega = 0.1$, and thus did not contain values close to 0. Thus, the standard deviations of local environmental effects varied between years and populations, in agreement with observed phenotypic standard deviations (Fig. 7, b and d).

Population means and standard deviations. Estimates of population phenotypic means and standard deviations were consistent with observed values (Fig. 7). Genetic means and standard deviations tended to evolve toward equilibrium values (Cavalli-Sforza and Feldman, 1976; Johnson and Barton, 2005; Zhang and Hill, 2005).

6. Discussion

Selection. This article develops a quantitative genetics model involving selection. Here fecundity is assumed to have a normal shape so that selection favours an interme-

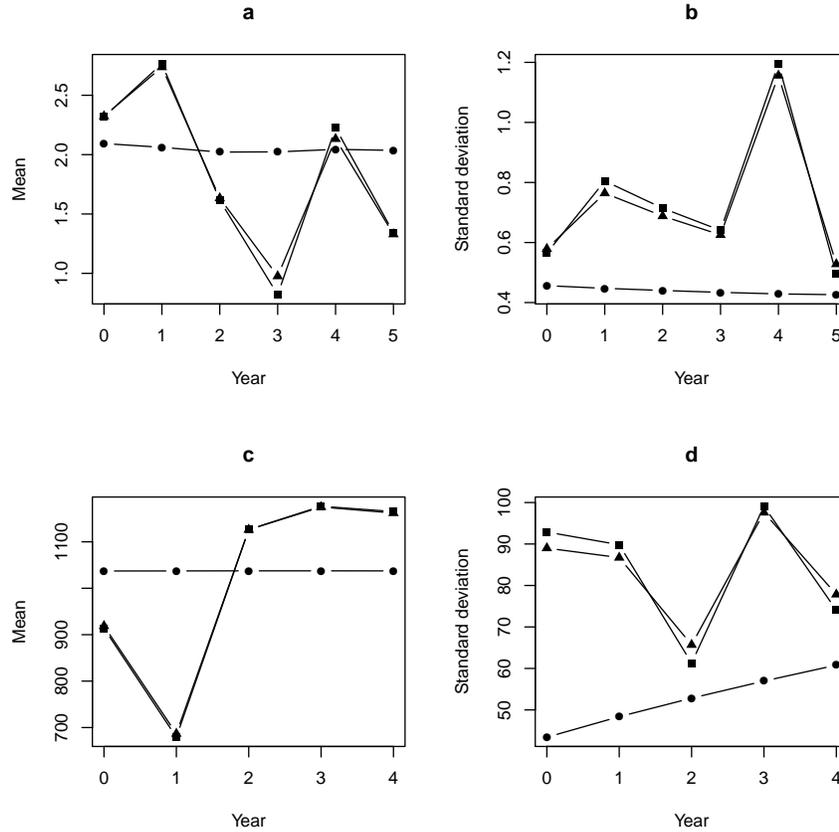


Figure 7: Analysis of wheat data: population means and standard deviations. Sample phenotypic means (■), estimated phenotypic means (▲) and estimated genetic means (●) for a population of location 2 and spike weight (a). Estimated phenotypic (respectively genetic) means are the posterior medians of μ_{Zijt} (respectively μ_{Gijt}). Sample phenotypic standard deviations (■), estimated phenotypic standard deviations (▲) and estimated genetic standard deviations (●) for the same population and spike weight (b). Estimated phenotypic (respectively genetic) standard deviations are the posterior medians of σ_{Zijt} (respectively σ_{Gijt}). Sample phenotypic means (■), estimated phenotypic means (▲) and estimated genetic means (●) for a population of location 5 and plant height (c). Sample phenotypic standard deviations (■), estimated phenotypic standard deviations (▲) and estimated genetic standard deviations (●) for the same population and plant height (d). In these analyses, $\gamma_\omega = 0.1$. Results are in g (respectively mm) for spike weight (respectively plant height).

diate trait value. This type of selection appears common in nature (Kingsolver et al., 2001) and leads to explicit formulae for the distribution of observations, which simplifies the implementation of statistical methods. Fecundity may also be modelled using a quadratic function (Débarre et al., 2014). This model takes account not only of stabilising selection but also of directional selection, which favours extreme trait values in one direction, and of diversifying selection, which favours extreme trait values. However it assumes that selection is weak and that genetic effects have a normal distribution at each generation. On the contrary, the model with a normal fecundity function does not assume that selection is weak and assumes that genetic effects have a normal distribution at the first generation only. This is why we here restricted ourselves to a normal fecundity function.

Selection on a trait may be real or apparent (Johnson and Barton, 2005; Zhang and Hill, 2005; Barton et al., 2007, p. 526). Selection is real when fitness differences are caused by the trait of interest and apparent when the trait of interest is neutral but evolves under the effect of the selection acting on correlated traits. Our model is univariate and is not able to distinguish between real or apparent selections. The targets of selection could be better studied by measuring many traits and by developing multivariate models allowing for the correlations between traits (Lande and Arnold, 1983; Troth et al., 2018). However this approach can hardly measure all the traits that are under real selection.

In principle, selection may affect both fecundity and survival. Here it is assumed to affect fecundity only to adapt the model to the available data. As our data are collected on adults, they provide more information on the selective pressures occurring during reproduction than during growth. In addition it is difficult to separate the selective pressures occurring during these phases using a single observation time per generation.

In the simulation study, estimates tend to be less sensitive to prior distributions when the data are more informative. Posterior distributions combine a prior information with the information brought by the data. Thus, they may be sensitive to changes in prior distributions when the data provide little information. This is the case in the wheat application, so that it is difficult to infer how selection varies between farms and between years. This is presumably because phenotypic time series are short and perturbed by large environmental effects. It is recommended to carry out a sensitivity analysis to prior distributions. If the results are sensitive to prior distributions, one may report this sensitivity and conclude that the available data do not allow to estimate parameters precisely. Alternatively, additional information may be introduced in the analysis by collecting more data (Carlin and Louis, 2008, p. 181).

For the sake of simplicity, the significance of estimated selection strengths are here assessed by examining if HPD credible intervals contain low values. Testing if these parameters vanish formally is difficult because zero is the boundary of the parameter space (Müller et al., 2013; Baey et al., 2019). More work is needed to provide rigorous tests of selection strengths.

Environmental effects. In our model, environmental effects are decomposed into between- and within-population effects, with standard deviations σ_A , σ_B and σ_{Cijt} . Both effects appear present in the wheat application. The standard deviation of local effects σ_{Cijt} is allowed to vary between generations to take account of environmental stochasticity (Caswell, 2001, Chap. 14). In the wheat application, this stochasticity could be due to variations in climatic conditions, soil fertility or farming practices over years.

Model assumptions. Our model assumes that populations are isolated and that individuals reproduce by selfing or asexually and are homozygous or haploid. These assumptions are partly justified in our application since wheat plants mainly reproduce by selfing, but not fully however since new genotypes may appear in populations because plants are not fully homozygous, because plants reproduce by a mixture of self fertilisation and outcrossing and because of seed dispersal. Thus, our estimates of mutation effects could capture various diversifying effects besides mutation. The selfing assumption could be relaxed by using the model developed by Cavalli-Sforza and Feldman (1976), which is based on outcrossing.

This article considers large populations and neglects demographic stochasticity, thus the developed method should be applied to small populations with caution. Populations are not very large in the wheat application, so that demographic stochasticity should be present. However its influence on the estimation of selection parameters should be limited since locations hold several populations. Evaluating the relative importances of evolutionary forces remains an open issue in evolutionary biology (Wright, 1931; Lande, 1976; Barton et al., 2007, Chap. 19). Our model allows to investigate the relative importances of selection and mutation but not those of selection and demographic stochasticity.

Acknowledgements

This work was supported by INRA (project AgriBio 4). We are grateful to the INRA MIGALE bioinformatics platform (<http://migale.jouy.inra.fr>) for providing computational resources. We thank Charles Ladmiraal and two anonymous reviewers for helpful comments.

References

- Ackermann, M., 2015. A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology* 13, 497–508.
- Baey, C., Cournède, P.H., Kuhn, E., 2019. Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. *Computational Statistics & Data Analysis* 135, 107 – 122. doi:10.1016/j.csda.2019.01.014.

- Barton, N.H., Briggs, D.E.G., Eisen, J.A., Goldstein, D.B., Patel, N.H., 2007. *Evolution*. Cold Spring Harbor Laboratory Press, NY.
- Barton, N.H., Etheridge, A.M., Véber, A., 2017. The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* 118, 50 – 73. doi:10.1016/j.tpb.2017.06.001.
- Carlin, B.P., Louis, T.A., 2008. *Bayesian methods for data analysis*. third ed., CRC Press.
- Caswell, H., 2001. *Matrix Population Models*. Sinauer Associates Inc. Publishers, Sunderland (USA).
- Cavalli-Sforza, L.L., Feldman, M.W., 1976. Evolution of continuous variation: Direct approach through joint distribution of genotypes and phenotypes. *Proceedings of the National Academy of Sciences* 73, 1689–1692.
- Ceccarelli, S., Grando, S., Bailey, E., Amri, A., El-Felah, M., Nassif, F., Rezgui, S., Yahyaoui, A., 2001. Farmer participation in barley breeding in Syria, Morocco and Tunisia. *Euphytica* 122, 521–536.
- David, O., Lannou, C., Monod, H., Papaïx, J., Traore, D., 2017. Adaptive diversification in heterogeneous environments. *Theoretical Population Biology* 114, 1–9. doi:10.1016/j.tpb.2016.11.003.
- Dawson, J.C., Rivière, P., Berthelot, J.F., Mercier, F., de Kochko, P., Galic, N., Pin, S., Serpolay, E., Thomas, M., Giuliano, S., Goldringer, I., 2011. Collaborative plant breeding for organic agricultural systems in developed countries. *Sustainability* 3, 1206–1223. doi:10.3390/su3081206.
- Débarre, F., Nuismer, S.L., Doebeli, M., 2014. Multidimensional (co)evolutionary stability. *Am Nat* 184, 158–171. doi:10.1086/677137.
- Débarre, F., Ronce, O., Gandon, S., 2013. Quantifying the effects of migration and mutation on adaptation and demography in spatially heterogeneous environments. *J Evol Biol* 26, 1185–1202. doi:10.1111/jeb.12132.
- Dobzhansky, T., 1973. Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher* 35, 125–129.
- Enjalbert, J., Dawson, J.C., Paillard, S., Rhoné, B., Rousselle, Y., Thomas, M., Goldringer, I., 2011. Dynamic management of crop diversity: From an experimental approach to on-farm conservation. *C R Biol* 334, 458–468. doi:10.1016/j.crvi.2011.03.005.

- Fabre, F., Rousseau, E., Mailleret, L., Moury, B., 2015. Epidemiological and evolutionary management of plant resistance: optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. *Evolutionary Applications* 8, 919–932.
- Friedenberg, N.A., 2003. Experimental evolution of dispersal in spatiotemporally variable microcosms. *Ecology Letters* 6, 953–959.
- Garnault, M., Duplaix, C., Leroux, P., Couleaud, G., Carpentier, F., David, O., Walker, A.S., 2019. Spatiotemporal dynamics of fungicide resistance in the wheat pathogen *Zymoseptoria Tritici* in France. *Pest Management Science* 75, 1794–1807. doi:10.1002/ps.5360.
- Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K., Foulley, J.L., 2009. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10, 550.
- Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis* 1, 515–534.
- Gillespie, J.H., 1998. Population genetics: a concise guide. JHU Press.
- Grant, P.R., Grant, B.R., 1995. Predicting microevolutionary responses to directional selection on heritable variation. *Evolution* 49, 241–251.
- Houle, D., Morikawa, B., Lynch, M., 1996. Comparing mutational variabilities. *Genetics* 143, 1467–1483.
- Johnson, T., Barton, N., 2005. Theoretical models of selection and mutation on quantitative traits. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 360, 1411–1425. doi:10.1098/rstb.2005.1667.
- Kawecki, T.J., Ebert, D., 2004. Conceptual issues in local adaptation. *Ecology Letters* 7, 1225–1241. doi:10.1111/j.1461-0248.2004.00684.x.
- Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E., Hoang, A., Gibert, P., Beerli, P., 2001. The strength of phenotypic selection in natural populations. *The American Naturalist* 157, 245–261.
- Kocher, T.D., 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics* 5, 288–298.
- Lande, R., 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30, 314–334.
- Lande, R., Arnold, S.J., 1983. The measurement of selection on correlated characters. *Evolution* 37, 1210–1226.

- Lande, R., Porcher, E., 2015. Maintenance of quantitative genetic variance under partial self-fertilization, with implications for evolution of selfing. *Genetics* 200, 891–906.
- Lande, R., Shannon, S., 1996. The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* 50, 434–437.
- Lefèvre, T., Raymond, M., Thomas, F., 2016. *Biologie évolutive*. De Boeck Supérieur.
- Lynch, M., 1988. The rate of polygenic mutation. *Genetical Research* 51, 137–148. doi:10.1017/S0016672300024150.
- McGuigan, K., 2006. Studying phenotypic evolution using multivariate quantitative genetics. *Molecular ecology* 15, 883–896.
- Müller, S., Scaely, J.L., Welsh, A.H., 2013. Model selection in linear mixed models. *Statistical Science* 28, 135–167.
- Papaïx, J., Monod, H., Goyeau, H., du Cheyron, P., Lannou, C., 2011. Influence of cultivated landscape composition on variety resistance: an assessment based on wheat leaf rust epidemics. *New Phytologist* 191, 1095–1107.
- Pariaud, B., Ravigné, V., Halkett, F., Goyeau, H., Carlier, J., Lannou, C., 2009. Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathology* 58, 409–424.
- Plummer, M., 2016. *rjags: Bayesian Graphical Models using MCMC*. URL: <http://CRAN.R-project.org/package=rjags>. R package version 4-6.
- Poisot, T., Bever, J.D., Nemri, A., Thrall, P.H., Hochberg, M.E., 2011. A conceptual framework for the evolution of ecological specialisation. *Ecology Letters* 14, 841–851.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- REX Consortium, 2013. Heterogeneity of selection and the evolution of resistance. *Trends in Ecology & Evolution* 28, 110–118. doi:http://dx.doi.org/10.1016/j.tree.2012.09.001.
- Rhoné, B., Remoue, C., Galic, N., Goldringer, I., Bonnin, I., 2008. Insight into the genetic bases of climatic adaptation in experimentally evolving wheat populations. *Molecular Ecology* 17, 930–943.
- Robert, C., 2007. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.

- Ropars, J., de La Vega, R.R., López-Villavicencio, M., Gouzy, J., Dupont, J., Swennen, D., Dumas, E., Giraud, T., Branca, A., 2016. Diversity and mechanisms of genomic adaptation in penicillium, in: DeVries, R.G., Andersen, M.R. (Eds.), *Aspergillus and Penicillium in the Post-Genomic Era*, pp. 27–42.
- Sasaki, A., Ellner, S., 1995. The evolutionarily stable phenotype distribution in a random environment. *Evolution* 49, 337–350.
- Scheiner, S.M., 2014. Bet-hedging as a complex interaction among developmental instability, environmental heterogeneity, dispersal, and life-history strategy. *Ecology and evolution* 4, 505–515.
- Sperling, L., Ashby, J.A., Smith, M.E., Weltzien, E., McGuire, S., 2001. A framework for analyzing participatory plant breeding approaches and results. *Euphytica* 122, 439–450.
- Suneson, C.A., 1956. An evolutionary plant breeding method. *Agronomy Journal* 48, 188–191.
- Troth, A., Puzey, J.R., Kim, R.S., Willis, J.H., Kelly, J.K., 2018. Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science* 361, 475–478.
- Wright, S., 1931. Evolution in mendelian populations. *Genetics* 16, 97.
- Zhang, X.S., Hill, W.G., 2005. Genetic variability under mutation selection balance. *Trends in Ecology & Evolution* 20, 468–470.