



HAL
open science

Approximate Bayesian computation with functional statistics

Samuel Soubeyrand, Florence Carpentier, François Guiton, Etienne K. Klein

► **To cite this version:**

Samuel Soubeyrand, Florence Carpentier, François Guiton, Etienne K. Klein. Approximate Bayesian computation with functional statistics. *Statistical Applications in Genetics and Molecular Biology*, 2013, 12 (1), pp.17-37. 10.1515/sagmb-2012-0014 . hal-01584412

HAL Id: hal-01584412

<https://agroparistech.hal.science/hal-01584412v1>

Submitted on 29 May 2020

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.

Research Article

Samuel Soubeyrand*, Florence Carpentier, François Guiton and Etienne K. Klein Approximate Bayesian computation with functional statistics

Abstract: Functional statistics are commonly used to characterize spatial patterns in general and spatial genetic structures in population genetics in particular. Such functional statistics also enable the estimation of parameters of spatially explicit (and genetic) models. Recently, Approximate Bayesian Computation (ABC) has been proposed to estimate model parameters from functional statistics. However, applying ABC with functional statistics may be cumbersome because of the high dimension of the set of statistics and the dependences among them. To tackle this difficulty, we propose an ABC procedure which relies on an optimized weighted distance between observed and simulated functional statistics. We applied this procedure to a simple step model, a spatial point process characterized by its pair correlation function and a pollen dispersal model characterized by genetic differentiation as a function of distance. These applications showed how the optimized weighted distance improved estimation accuracy. In the discussion, we consider the application of the proposed ABC procedure to functional statistics characterizing non-spatial processes.

Keywords: dispersal model; marked point process; pairwise genetic distance; parameter estimation; spatial model; TwoGener.

*Corresponding author: Samuel Soubeyrand, INRA, UR546 Biostatistics and Spatial Processes, F-84914 Avignon, France, e-mail: Samuel.Soubeyrand@avignon.inra.fr

Florence Carpentier: AgroParisTech, Dpt. SVS, F-75231 Paris, France

François Guiton and Etienne K. Klein: INRA, UR546 Biostatistics and Spatial Processes, F-84914 Avignon, France

1 Introduction

Statistical analysis of spatial data often rely on functional statistics such as correlograms or variograms in geostatistics (Cressie, 1991; Chilès and Delfiner, 1999), and pair correlation or Ripley's functions for point pattern analysis (Cressie, 1991; Illian et al., 2008). For the particular case of spatial genetic data, functions giving a pairwise genetic distance/index with respect to the pairwise spatial distance between populations or individuals have been widely exploited to characterize spatial genetic structures and to derive effective gene flow. For instance, population geneticists commonly use functions of differentiation between populations, F_{ST} (Rousset, 1997), between individuals (Rousset, 2000) and between pollen pools, Φ_{FT} (Austerlitz and Smouse, 2002), and functions of relatedness between individuals (Hardy, 2003) and between pollen pools (Robledo-Arnuncio et al., 2006). These functions enable the estimation of dispersal parameters and effective population sizes from spatial and genetic data (e.g., Rousset, 2000; Austerlitz et al., 2004).

For these genetic indices, likelihood-based methods are rarely adopted to fit the model due to the complexity of the likelihood (Rousset and Leblois, 2007). Rather, least-square estimates are obtained relying on a simple but robust model which provides expectations of the functional statistics under different parameter values. Such approaches generally omit information about the specific distribution of individuals in the study population. For instance when one estimates pollen dispersal with the TwoGener approach (Austerlitz and Smouse, 2002; Robledo-Arnuncio et al., 2006) the expected functional statistics $\Phi_{FT}(r)$ is computed assuming a Poisson distribution for the pollen sources whereas departures from this hypothesis are frequent and modify

consequently the expected $\Phi_{FT}(r)$ (Robledo-Arnuncio and Austerlitz, 2006). To solve this problem, Carpentier (2010) proposed to apply Approximate Bayesian Computation (ABC) and to use all the pairwise genetic indices Φ_{FT} as summary statistics. However the large number of summary statistics (e.g., 190 summary statistics for only 20 sampling points) prevents the implementation of several efficient extensions of ABC, and Beaumont's local regression in particular (Beaumont et al., 2002). Here, we alternatively propose a method where $r \mapsto \Phi_{FT}(r)$ is viewed as a functional statistics. We propose a new method to optimize the use of functional statistics within ABC in general and apply it to the particular case of pollen dispersal estimation from $\Phi_{FT}(r)$.

ABC encompasses a set of inference methods devoted to stochastic implicit models whose distribution theory is intractable but which can be run a large number of times to simulate data sets. Since the proposal of the ABC-rejection procedures (Rubin, 1984; Pritchard et al., 1999), many improvements have been proposed to increase the inference accuracy of ABC; see e.g., the reviews by Beaumont (2010), Csilléry et al. (2010), Marin et al. (2011), Marjoram et al. (2003). In the present study, we are specifically concerned in improving the selection of a distance function between observed and simulated data sets.

In ABC approaches, the distance between observed and simulated data sets is generally the distance between the set of summary statistics computed for the observed and simulated data sets. Traditionally, Euclidean or weighted Euclidean distances are used but the selection of the statistics and the distance function between statistics is crucial. An inappropriate choice of statistics may lead ABC to under-perform; see Blum (2010a) and Leuenberger and Wegmann (2010) who describe the curse of dimensionality in ABC and Haon-Lasportes et al. (2011) who illustrate with a basic example the negative effect of information redundancy, dimension increase and uninformativity in the set of statistics.

Numerous procedures have thus been proposed either to transform an initial set of statistics (possibly of large dimension) or build a distance between observed and simulated data sets in the aim of improving the inference accuracy. For instance, the variances of summary statistics may be equalized (Beaumont et al., 2002), subsets of statistics may be selected based on Kullback-Leibler divergence, relative ε -sufficiency or minimum entropy (Joyce and Marjoram, 2008; Nunes and Balding, 2010; Barnes et al., 2012), the statistics may be transformed into axes of a PLS regression (Wegmann et al., 2009) or into point estimates of the model parameters (Haon-Lasportes et al., 2011; Fearnhead and Prangle, 2012). These procedures reduce mean square errors and/or improve the coverage properties of the posterior intervals (see Blum et al., 2012, which also provides a comparison between dimension reduction methods in ABC).

Here, we propose an alternative procedure adapted to functional statistics. In the transformation mentioned above, all the raw summary statistics are considered equally. However, when one works with functional statistics, there is generally a dependence of the function values along the support of the function. We propose to take into account this dependence by using for the ABC-rejection algorithm a weighted distance between observed and simulated functional statistics. We use a piecewise constant weight function and optimize it by minimizing a loss function computed from pseudo-observed data sets (PODS) drawn in the prior distribution. We also optimize the acceptance rate, which has a large impact on the performance of ABC-rejection algorithm (Beaumont et al., 2002). Our optimization is based either on a Bayesian mean square error or on a partial mean square error after a preliminary pilot ABC run. A similar approach has been proposed by Jung and Marjoram (2011) for non-functional statistics; see Discussion Section 6.

In the following, our approach is presented in Section 2. Then, the method is applied to three examples. First, we considered a simple step model (Section 3) as a simple case where the full Bayesian approach can be compared with our ABC approach. Second, we considered a spatial point process characterized through its pair correlation function (Section 4). This is a typical problem from spatial statistics where ABC may be a promising approach to carry out Bayesian estimation of mechanistic parameters that determine spatial point patterns. Third, we considered a pollen dispersal model including genetic information from molecular markers (Section 5). Here the observation of molecular markers at sampling locations provides a genetic pairwise differentiation index as a function of pairwise distance (Austerlitz and Smouse, 2002). In the three applications, the performance of the optimized weight function is compared with the performance of the constant weight function and the weight function obtained by equalizing the variances of the statistics. For the last example, we also compared our method with the performances of the PLS (Wegmann et al., 2009),

minimum entropy and two-stage (Nunes and Balding, 2010) methods which reduce the dimension of the summary statistics.

2 Method

2.1 Background: The ABC-rejection procedure

Consider observed data $\mathcal{D} \in \mathbb{D}$ which are assumed to be generated under the stochastic model \mathcal{M}_θ parametrized by $\theta \in \Theta$ with prior density π . The data space \mathbb{D} and the parameter space Θ are both included in multidimensional sets of real vectors.

The posterior distribution $p(\theta | \mathcal{D})$ can be estimated using the following ABC-rejection algorithm (Rubin, 1984):

- A1.** Carry out the next two steps, independently for i in $\{1, \dots, I\}$,
1. Generate θ_i from π and simulate \mathcal{D}_i from \mathcal{M}_{θ_i} .
 2. Accept θ_i if $\mathcal{D}_i = \mathcal{D}$, reject it otherwise.

The set of accepted θ_i forms a sample from the posterior distribution

$$p(\theta | \mathcal{D}) = \frac{f(\mathcal{D} | \theta) \pi(\theta)}{\int_{\Theta} f(\mathcal{D} | \alpha) \pi(\alpha) d\alpha},$$

where $f(\mathcal{D} | \theta)$ is the conditional probability distribution function of \mathcal{D} given θ , i.e. the (intractable or unknown) likelihood of the model \mathcal{M}_θ .

Algorithm **A1** is rarely usable because the probability of generating \mathcal{D}_i equal to \mathcal{D} is very low when the dimensionality of the data space \mathbb{D} is large and this probability is even zero for continuous data. To circumvent this difficulty, two ideas have been applied: the introduction of a tolerance threshold and the replacement of the raw data by summary statistics (Pritchard et al., 1999). This leads to the following ABC-rejection algorithm which will be extensively used in this article:

- A2.** Carry out the next three steps, independently for i in $\{1, \dots, I\}$,
1. Generate θ_i from π and simulate \mathcal{D}_i from \mathcal{M}_{θ_i} .
 2. Compute the statistics $S_i = s(\mathcal{D}_i)$, where s is a function from \mathbb{D} to the space \mathbb{S} of statistics.
 3. Accept θ_i if $d(S_i, S) \leq \varepsilon(\tau)$, where d is a distance over \mathbb{S} and $\varepsilon(\tau) \in \mathbb{R}_+$ is a tolerance threshold for the distance between the observed statistics $S = s(\mathcal{D})$ and the simulated ones. The threshold $\varepsilon(\tau)$ depends on the proportion τ of accepted θ_i among the I simulated parameters; $\varepsilon(\tau)$ is the empirical quantile of order τ . Thereafter, τ is called the acceptance rate.

The set of accepted parameters, say $\Theta_{\tau, I} = \{\theta_i : d(S_i, S) \leq \varepsilon(\tau), i=1, \dots, I\}$, forms a sample from the posterior distribution

$$p_{\varepsilon(\tau)}(\theta | S) = \frac{\left(\int_{B(S, \varepsilon(\tau))} \tilde{f}(z | \theta) dz \right) \pi(\theta)}{\int_{\Theta} \left(\int_{B(S, \varepsilon(\tau))} \tilde{f}(z | \alpha) dz \right) \pi(\alpha) d\alpha}, \quad (1)$$

where $\tilde{f}(S | \theta)$ is the conditional probability distribution function of S given θ and $B(S, \varepsilon(\tau))$ is the ball with center S and radius $\varepsilon(\tau)$ in the space \mathbb{S} with distance d .

When $\varepsilon(\tau)$ tends to zero, $p_{\varepsilon(\tau)}(\theta | S)$ may be a good approximation of the posterior distribution conditional on the statistics (Appendix A), i.e.

$$p(\theta|S) = \frac{\tilde{f}(S|\theta)\pi(\theta)}{\int_{\Theta} \tilde{f}(S|\alpha)\pi(\alpha)d\alpha}, \quad (2)$$

and the sample $\Theta_{\tau,l}$ of accepted parameters is approximately distributed under this posterior distribution. If, in addition, the statistics are sufficient, then $\tilde{f}(S|\theta) = f(\mathcal{D}|\theta)$ and $\Theta_{\tau,l}$ is approximately a sample from the classical posterior distribution $p(\theta|\mathcal{D})$ conditional on the data.

2.2 Selecting a weight function for functional statistics

Here, S is a functional statistics. The space \mathbb{S} of statistics is included in the space of real-valued and square-integrable functions defined over \mathbb{R} :

$$\mathbb{S} \subset \left\{ g: \mathbb{R} \rightarrow \mathbb{R}, \int_{\mathbb{R}} g^2 < \infty \right\}.$$

Besides, we assume that the distance $d: \mathbb{S}^2 \rightarrow \mathbb{R}^+$, used in algorithm **A2** to compare observed and simulated statistics, is parametrized by a non-negative weight function $w: \mathbb{R} \rightarrow \mathbb{R}_+$ and satisfies:

$$d(S_i, S; w) = \int_{\mathbb{R}} w(r) \{S_i(r) - S(r)\}^2 dr. \quad (3)$$

The weight function is expected to modulate the squared difference between $S_i(r)$ and $S(r)$ with respect to the information about the parameters brought by the statistics at r . Other weighted distances could be used and compared, e.g. the integral or the maximum of weighted absolute values $|S_i(r) - S(r)|$. It has to be noted that if the distance is changed, then the conditions satisfied by the functional statistics S may be different: here, the distance given in Eq. (3) is properly defined because functional statistics are square-integrable functions.

We used three weight functions. The first one is the constant function:

$$w_{cst}(r) = 1.$$

The second weight function is the inverse variance function, which corresponds to a classical scaling procedure in ABC consisting in equalizing the variances of the statistics (Beaumont et al., 2002):

$$w_{var}(r) = \begin{cases} \text{Var}(S_i(r))^{-1} & \text{if } \text{Var}(S_i(r)) > 0 \\ 0 & \text{otherwise;} \end{cases}$$

The third weight function is the optimized function w_{opt} obtained by minimizing a mean square error (MSE) of a point estimate of θ (Rohatgi, 2003, chap. 4). The MSE that we used is a Bayesian MSE (BMSE): the square error is integrated over Θ with respect to the prior distribution π . This approach, detailed below in algorithm **A3**, is analogue to minimizing the mean square error of prediction where θ is the random variable to be predicted McCulloch and Searle (2001, chap. 9).

The optimized weight function w_{opt} as well as an optimized acceptance rate τ_{opt} are determined within the following ABC-rejection algorithm:

A3. Carry out the next four steps,

1. For i in $\{1, \dots, I\}$, independently generate θ_i from π , simulate \mathcal{D}_i from \mathcal{M}_{θ_i} and compute the functional statistic $S_i = s(\mathcal{D}_i)$;
2. For j in $\{1, \dots, J\}$, independently generate θ'_j from π , simulate \mathcal{D}'_j from $\mathcal{M}_{\theta'_j}$ and compute the functional statistic $S'_j = s(\mathcal{D}'_j)$; (θ'_j, S'_j) , $j=1, \dots, J$, will be used as pseudo-observed data sets (PODS);
3. Select the weight function and the acceptance rate which minimize the following BMSE criterion:

$$\text{BMSE}_j(w, \tau) = \frac{1}{J} \sum_{j=1}^J \sum_{k=1}^K \frac{(\hat{\theta}'_{jk}(w, \tau) - \theta'_j)^2}{\text{Var}(\hat{\theta}'_{jk})}. \quad (4)$$

In Eq. (4), θ'_{jk} , $k=1, \dots, K$, are the K components of θ'_j ($\Theta \subset \mathbb{R}^K$, $K \geq 1$); $\text{Var}(\theta'_{jk})$ is the prior variance of θ'_{jk} depending only on π and allows the scaling of the parameter components; the point estimates $\hat{\theta}'_{jk}(w, \tau)$ are the marginal posterior medians of θ'_{jk} :

$$\hat{\theta}'_{jk}(w, \tau) = \text{Median}\{\theta'_{ik} : d(S_i, S'_j; w) \leq \varepsilon(\tau), i=1, \dots, I\},$$

obtained by applying the last step of algorithm **A2** with S'_j for the observed statistic, S_i for the simulated statistics, $d(\cdot, \cdot; w)$ for the distance and τ for the acceptance rate. The BMSE is minimized over the space function $\mathbb{W} = \{w: \mathbb{R} \rightarrow \mathbb{R}_+, \int_{\mathbb{R}} w = 1\}$ and the interval $(0, 1]$:

$$(w_{opt}, \tau_{opt}) = \text{argmin}_{w, \tau \in \mathbb{W} \times (0, 1]} \text{BMSE}_J(w, \tau). \quad (5)$$

4. For i in $\{1, \dots, I\}$, accept θ_i if $d(S_i, S; w_{opt}) \leq \varepsilon(\tau_{opt})$.

The set of accepted parameters $\Theta_{opt} = \{\theta_i : d(S_i, S; w_{opt}) \leq \varepsilon(\tau_{opt}), i=1, \dots, I\}$ forms a sample from the posterior distribution (1) with $\varepsilon(\tau) = \varepsilon(\tau_{opt})$ and with $B(S, \varepsilon(\tau))$ equal to the ball with center S and radius $\varepsilon(\tau_{opt})$ in the space \mathbb{S} with distance $d(\cdot, \cdot; w_{opt})$. Thus, weighting the distance modifies the posterior under which the accepted parameters are drawn. However, when $\varepsilon(\tau_{opt})$ tends to zero, the new posterior distribution [like the one given in Eq. (1)] may be a good approximation of $p(\theta|S)$ given in Eq. (2); see Appendix A.

Note that the BMSE in Eq. (4) is the Monte-Carlo approximation of the exact BMSE equal to $\sum_{k=1}^K E\{(\hat{\theta}'_{jk}(w, \tau) - \theta'_{jk})^2\} / \text{Var}(\theta'_{jk})$. Besides, other criteria than the BMSE may be used to select w and τ , e.g. mean square errors or mean absolute errors based on the posterior mode, the posterior mean or posterior quantiles.

To be able to compare the ABC procedures using the three weight functions w_{cst} , w_{var} and w_{opt} , the acceptance rate τ should also be optimized when w_{cst} and w_{var} are used. Thus, the algorithm **A3** is modified for w_{cst} and w_{var} by carrying out the minimization of the BMSE with respect to the acceptance rate τ only [this approach is analogue to the proposal of Csilléry et al. (2011) based on cross-validation]. When w_{cst} or w_{var} is used, the minimization program (5) is respectively replaced by:

$$\tau_{cst} = \text{argmin}_{\tau \in (0, 1]} \text{BMSE}_J(w_{cst}, \tau) \quad (6)$$

or

$$\tau_{var} = \text{argmin}_{\tau \in (0, 1]} \text{BMSE}_J(w_{var}, \tau) \quad (7)$$

2.3 Using a pilot ABC run

After a first (pilot) run of algorithm **A3** which yields a pilot posterior sample, namely Θ_{pilot} , one may proceed to a second selection of the weight function and the acceptance rate by restricting the computation of the MSE to simulations close to Θ_{pilot} . This approach may be implemented as follows (without supplementary simulation):

A4. Carry out the next three steps,

1. Select the set \mathcal{J} with size $|\mathcal{J}| < J$ formed by the indices $j \in \{1, \dots, J\}$ corresponding to the $|\mathcal{J}|$ smallest distances between θ'_j and Θ_{pilot} , this distance being defined by:

$$\min \left\{ \sum_{k=1}^K \frac{(\theta'_{jk} - \theta_{pilot,k})^2}{\text{Var}(\theta'_{jk})} : \theta_{pilot} \in \Theta_{pilot} \right\},$$

where $\theta_{pilot,k}$, $k=1, \dots, K$, are the K components of θ_{pilot} .

2. Select the weight function and the acceptance rate which minimize the following partial MSE (PMSE) criterion:

$$\text{PMSE}_{\mathcal{J}}(w, \tau) = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} \sum_{k=1}^K \frac{(\hat{\theta}'_{jk}(w, \tau) - \theta'_{jk})^2}{\text{Var}(\theta'_{jk})}. \quad (8)$$

Terms in Eq. (8) are the same than those in Eq. (4) except that the sum is restricted to \mathcal{J} . The PMSE is minimized over the space function \mathbb{W} and the interval $(0,1]$:

$$(w_{opt}^{pilot}, \tau_{opt}^{pilot}) = \text{argmin}_{w, \tau \in \mathbb{W} \times (0,1]} \text{PMSE}_{\mathcal{J}}(w, \tau). \quad (9)$$

3. For i in $\{1, \dots, I\}$, accept θ_i if $d(S_i, S; w_{opt}^{pilot}) \leq \varepsilon(\tau_{opt}^{pilot})$.

The set of accepted parameters forms a sample from the posterior distribution (1) with $\varepsilon(\tau) = \varepsilon(\tau_{opt}^{pilot})$ and with $B(S, \varepsilon(\tau))$ equal to the ball with center S and radius $\varepsilon(\tau_{opt}^{pilot})$ in the space \mathbb{S} with distance $d(\cdot, \cdot; w_{opt}^{pilot})$. Here also weighting the distance modifies the posterior under which the accepted parameters are drawn. However, when $\varepsilon(\tau_{opt}^{pilot})$ tends to zero, the new posterior distribution [like the one given in Eq. (1)] may be a good approximation of $p(\theta|S)$ given in Eq. (2); see Appendix A.

The pilot posterior sample can be the set of accepted parameters resulting from algorithm **A3** based on w_{cst} , w_{var} or w_{opt} . It can even be a combination of these three posterior samples.

2.4 Optimization algorithm

In the examples tackled below, the class of weight functions \mathbb{W} in which the optimized function w_{opt} is selected is restricted to sets of positive piecewise constant functions with a finite number of jumps, with known jump locations $r_0 < \dots < r_N$ ($N > 0$) and with integral over \mathbb{R} equal to 1. This implies that $w \in \mathbb{W}$ takes the value zero over the infinite intervals $(-\infty, r_0)$ and $[r_N, \infty)$. Let w_0, \dots, w_{N-1} denote the values of w in the intervals $[r_0, r_1), \dots, [r_{N-1}, r_N)$.

The optimization program (5) [resp. (9)] consists in minimizing the BMSE (resp. the PMSE) with respect to a finite number of jump levels, namely w_0, \dots, w_{N-1} , and the acceptance rate τ with the constraints $w_0 \geq 0, \dots, w_{N-1} \geq 0$, $\sum_{n=0}^{N-1} (r_{n+1} - r_n) w_n = 1$ and $\tau \in (0, 1]$. To carry out this optimization we used the Nelder-Mead algorithm (Nelder and Mead, 1965) modified to take into account the constraints on w_0, \dots, w_{N-1} , τ : when the constraints were not satisfied in a reflection (or expansion) step, the reflection (or expansion) coefficient was divided by $m = 2, 3, \dots$ until the constraints were satisfied.

When w_{cst} and w_{var} are used, only the acceptance rate has to be selected. This corresponds to a simple search of a minimum in the interval $(0, 1]$.

3 Application to a simple step model

3.1 Model

We built a model which randomly generates the heights of four steps; three realizations of this model are shown in top panels of Figure 1. The heights, which are independently and normally distributed with varying means and varying variances, are used as summary statistics (in this case the summary statistics are the raw data).

More formally, we study the step functional statistics S equal to zero over \mathbb{R} except over the interval $[0, 4)$:

$$S(r) = \begin{cases} \theta \lfloor r \rfloor^2 + \varepsilon(\lfloor r \rfloor) & \text{if } r \in [0, 4) \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

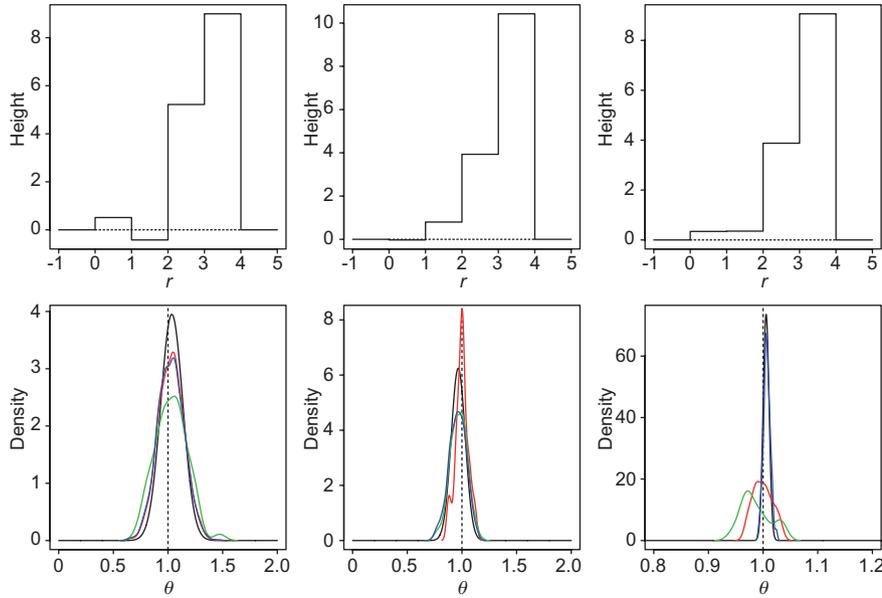


Figure 1 Realizations (top panels) of the simple step model and corresponding posterior distributions (PDs, bottom panels). The realizations were obtained with the constant (left), increasing (center) and decreasing (right) noise structures. The PDs were obtained with the exact approach (black curve) and with algorithm **A3** applied with the three weight functions w_{cst} (red curve), w_{var} (green curve) and w_{opt} (blue curve). The last three PDs were drawn using a Gaussian kernel smoothing applied to the posterior sample provided by the algorithm. The vertical dashed line is the true value of θ , namely 1.

where $\lfloor r \rfloor$ denotes the floor value of r , θ is a real parameter and $\varepsilon(0), \varepsilon(1), \varepsilon(2), \varepsilon(3)$ are independently drawn from centered normal distributions with standard deviations $\sigma(0), \sigma(1), \sigma(2), \sigma(3)$. Thus, the heights of the four steps are $\varepsilon(0), \theta + \varepsilon(1), 4\theta + \varepsilon(2)$ and $9\theta + \varepsilon(3)$.

Let the prior distribution for θ be the uniform distribution over $[0, 2]$.

We study three noise structures, that is to say three vectors of values for the standard deviations:

1. Constant noise: $[\sigma(0), \sigma(1), \sigma(2), \sigma(3)] = (1, 1, 1, 1)$
2. Increasing noise: $[\sigma(0), \sigma(1), \sigma(2), \sigma(3)] = (0.05, 0.1, 0.5, 1)$
3. Decreasing noise: $(\sigma(0), \sigma(1), \sigma(2), \sigma(3)) = (1, 0.5, 0.1, 0.05)$.

We aim to see which of the weight functions leads to the lowest BMSE and, for the optimized function, how the statistics are weighted with respect to the structure of the noise. We especially aim to see how the statistics $S(0)$ which does not depend on θ is weighted.

3.2 ABC tuning

For each noise structure of the model described above, we independently ran 500 times the first two steps of algorithm **A3**, with $I=10^5$ and $J=10^3$. Then, for each weight function, we minimized the BMSE; see Eqs. (5), (6) and (7).

We selected the optimized weight function among the step functions satisfying:

$$w(r) = \begin{cases} w_n & \text{if } r \in [n, n+1), \forall n \in \{0, 1, 2, 3\} \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

$w_0, w_1, w_2, w_3 \geq 0$ and $\sum_{n=0}^3 w_n = 1$. The distance between functional statistics S_i and S_j is simply:

$$d(S_i, S'_j; w) = \int_{\mathbb{R}} w(r) \{S_i(r) - S'_j(r)\}^2 dr$$

$$= \sum_{n=0}^3 w_n \{S_i(n) - S'_j(n)\}^2.$$

3.3 Results

The average BMSEs obtained with the three weight functions and the three noise structures are shown in Table 1. The gain in using w_{opt} for the constant noise structure is not large but is larger for the other noise structures. Besides, w_{cst} is better than w_{var} for the increasing noise structure whereas the opposite is true for the decreasing noise structure. This indicates that the choice of one of these functions depends on the model. In comparison, the performance of the optimized function w_{opt} is more robust: Using w_{opt} led for each of the 1500 runs to the lowest BMSE. This was expected because, in this application, w_{cst} and w_{var} are particular cases of w_{opt} . In the two next applications, only w_{cst} is a particular case of w_{opt} .

Table 2 provides statistics about the optimized acceptance rate and weight function. The size ($10^5 \times \tau_{opt}$) of the posterior sample is much larger for the constant noise structure than for the varying noise structures. This may explain the low reduction of the BMSE in the case of the constant noise structure, see Table 1. For the increasing noise structure, the optimization program yields high weight to $S(0)$ (with no information about θ) because it does not cost a lot to use a statistics which has a very low variability. Besides, among $S(1)$, $S(2)$ and $S(3)$ the algorithm gives more weight to the less varying statistics, namely $S(1)$. For the decreasing noise structure, the algorithm gives low weight for the strongly varying and useless statistics $S(0)$. Moreover, among $S(1)$, $S(2)$ and $S(3)$ the algorithm gives more weight to the less varying statistics, namely $S(3)$.

3.4 Comparison with the exact posterior distribution

For the simple step model, we can compare the ABC-based posterior distributions and the exact posterior distribution. The likelihood of the model is:

$$L(\theta) = \phi_{0, \sigma(0)}(S(0)) \phi_{\theta, \sigma(1)}(S(1)) \phi_{4\theta, \sigma(2)}(S(2)) \phi_{9\theta, \sigma(3)}(S(3)),$$

where $\phi_{\mu, \sigma}(\cdot)$ is the probability density function of the normal distribution with mean μ and standard deviation σ . Thus, the exact posterior distribution of θ can be easily computed:

$$p_{exact}(\theta|S) = \frac{L(\theta) \mathbf{1}(0 \leq \theta \leq 2)}{\int_0^2 L(\alpha) d\alpha},$$

where $\mathbf{1}$ is the indicator function [$\mathbf{1}(E)=1$ if event E holds, zero otherwise].

Table 1 BMSE statistics for the simple step model. For each noise structure and each weight function: mean value (and standard deviation) of $1000 \times \text{BMSE}$ based on 500 runs (1st line); number of times among the 500 runs that each weight function provided the lowest BMSE (2nd line).

	w_{cst}	w_{var}	w_{opt}
Constant noise	9.30 (0.44)	10.02 (0.47)	9.27 (0.44)
	0	0	500
Increasing noise	4.23 (0.20)	3.90 (0.18)	3.85 (0.17)
	0	0	500
Decreasing noise	0.044 (0.002)	0.259 (0.019)	0.030 (0.001)
	0	0	500

Table 2 Mean values (and standard deviations) of the optimum acceptance rate τ_{opt} and weight function w_{opt} for the simple step model with respect to the three noise structures (based on 500 runs). The quantity $10^5 \times \tau_{opt}$ is the size of the posterior sample based on the $l=10^5$ simulations.

Noise	$10^5 \times \tau_{opt}$	$w_{opt}(0)$	$w_{opt}(1)$	$w_{opt}(2)$	$w_{opt}(3)$
Constant	1940 (930)	0.16 (0.11)	0.23 (0.10)	0.30 (0.09)	0.31 (0.08)
Increasing	360 (200)	0.98 (0.02)	0.02 (0.02)	0.00 (0.00)	0.00 (0.00)
Decreasing	85 (38)	0.02 (0.05)	0.03 (0.02)	0.18 (0.05)	0.77 (0.06)

Bottom panels of Figure 1 show, for the three realizations provided in top panels, the posterior distributions obtained with the exact approach and with algorithm **A3** applied with the three weight functions w_{cst} , w_{var} and w_{opt} . Qualitatively, we can see that algorithm **A3** with w_{opt} leads to a reasonable approximation of the exact posterior distribution. Besides, these plots corroborates what can be seen in Table 1:

1. For the constant noise structure (1st line in Table 1 and left plots in Figure 1), the mean BMSEs with w_{cst} and w_{opt} are close as well as the corresponding posterior distributions;
2. For the increasing noise structure (2nd line in Table 1 and center plots in Figure 1), the mean BMSEs with w_{var} and w_{opt} are close as well as the corresponding posterior distributions;
3. For the decreasing noise structure (3rd line in Table 1 and right plots in Figure 1), the mean BMSE with w_{opt} is much lower than the mean BMSEs with w_{cst} and w_{var} , and the posterior distribution with w_{opt} is a better approximation of the exact posterior distribution.

These are qualitative statements based on three simulations but these statements are consistent when a series of simulations/estimations is performed (data not shown).

4 Application to a spatial point process

4.1 Model and methods

A Neyman-Scott point process (Illian et al., 2008, chap. 6) is made of daughter points forming clusters around parent points. The parent points are generated from a stationary Poisson point process with intensity λ . The daughter points in a cluster are random in number with mean μ and are independently and identically distributed around their parent. The Neyman-Scott process only consists of the daughter points and is a model for clustered point patterns. It can be viewed as a model for the dispersal of a species over one generation.

Here, we consider a particular case of the Neyman-Scott process: the planar case of the modified Thomas process (Illian et al., 2008, chap. 6). The daughter points are distributed around each parent point under a bivariate isotropic normal distribution with dispersion parameter σ . The modified Thomas process, say χ , is observed in the square window $[0, 1] \times [0, 1]$. The intensity λ of the parents is supposed to be known: $\lambda=100$ but the parameters μ and σ have independent uniform prior distributions over the intervals $[1, 10]$ and $[0.01, 0.1]$, respectively.

Figure 2 (top) shows three realizations of the modified Thomas process with three values of the dispersal parameter σ .

4.2 ABC tuning

Functional summary characteristics, such as the pair correlation function, are commonly used to estimate the parameters of point processes; see Illian et al. (2008, chap. 7). For instance, the pair correlation function

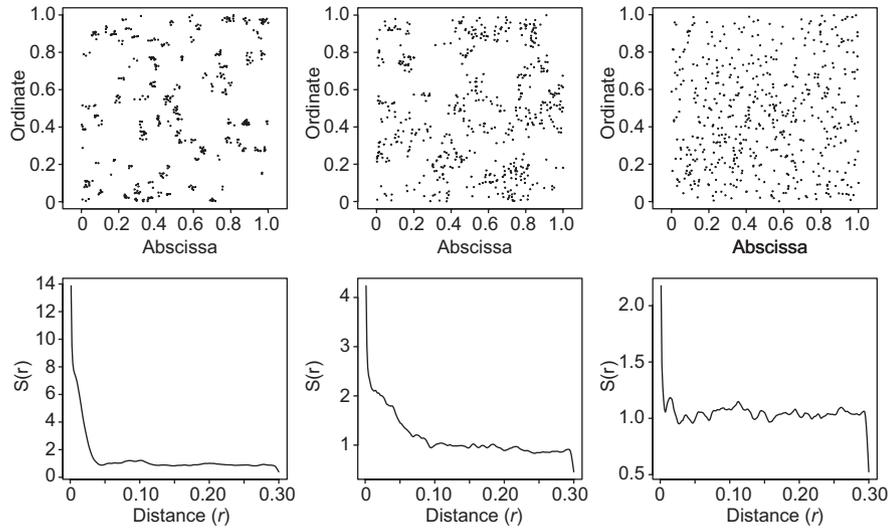


Figure 2 Realizations of the modified Thomas process and the corresponding estimated pair correlation functions S . $(\lambda, \mu, \sigma)=(100, 5, 0.01)$ in the left panels, $(100, 50, 0.025)$ in the middle panels and $(100, 5, 0.1)$ in the right panels.

especially brings information about the degree of clustering in point patterns. Below, the functional statistics under consideration is the empirical pair correlation function, thereafter noted $S(r)$, given by the kernel smoothing method with the Ripley's correction (Stoyan and Stoyan, 1994, pp. 284–285). We restricted $S(\cdot)$ to the interval $(0, 0.3)$ where the endpoint 0.3 is about the fifth of the maximum inter-point distance in the point patterns studied in this section; see Illian et al. (2008, p. 96) for recommendations about the choice of the endpoint. Examples of $S(\cdot)$ functions are drawn in Figure 2 (bottom).

Using the modified Thomas process and the prior distributions described above, we independently ran 500 times the first two steps of algorithm **A3**, with $I=10^5$ and $J=10^3$. Then, for the three weight functions w_{cst} , w_{var} and w_{opt} , we minimized the BMSE; see Eqs. (5), (6) and (7).

The optimized weight function w_{opt} was selected among the step functions with 21 jumps satisfying:

$$w(r) = \begin{cases} w_n & \text{if } r \in [0.3n/20, 0.3(n+1)/20), \forall n \in \{0, 1, \dots, 19\} \\ 0 & \text{if } r < 0 \text{ or } r \geq 0.3, \end{cases} \quad (12)$$

$$w_0, \dots, w_{19} \geq 0 \text{ and } \int_{\mathbb{R}} w(r) dr = \sum_{n=0}^{19} (0.3/20) w_n = 1.$$

The distance between functional statistics S_i and S'_j was approximated by a finite sum of 249 terms:

$$\begin{aligned} d(S_i, S'_j; w) &= \int_{\mathbb{R}} w(r) \{S_i(r) - S'_j(r)\}^2 dr \\ &\approx \sum_{k=1}^{249} w(0.3k/250) \{S_i(0.3k/250) - S'_j(0.3k/250)\}^2. \end{aligned}$$

Here, w_{var} which can take 249 different values, is not a particular case of w_{opt} which can only take 20 different values. Thus, if equalizing the variance of the statistics is a valuable approach for this application, then w_{var} could lead to better results than w_{opt} .

4.3 Results

The BMSEs obtained with the three weight functions are shown in Table 3. The decrease of the BMSE with w_{opt} is significant and w_{opt} led to the lowest BMSE 500 times over 500 runs.

Table 3 BMSE statistics for the modified Thomas process: mean value (and standard deviation) of the BMSE for each weight function, based on 500 runs (1st line); number of times among the 500 runs that each weight function provides the minimum BMSE (2nd line).

	w_{cst}	w_{var}	w_{opt}
BMSE	0.651 (0.024)	0.942 (0.031)	0.365 (0.025)
Lowest BMSE frequency	0	0	500

The optimum size of the posterior sample (number of accepted parameters among $I=10^5$ parameters) generally takes low value (mean: 17.9, SD: 9.3); see Figure 3, left. It does not really differ from the optimum size obtained with w_{cst} (mean: 18.0, SD: 7.2) but is generally smaller than the optimum size obtained with w_{var} (mean: 35.3, SD: 13.1).

The optimum weight function w_{opt} varies between runs (Figure 3, middle and right) but the median optimum function has a U-shape, giving more weight to $S(r)$ when r is small or close to 0.3. Despite the variability of w_{opt} between runs, this weight function strongly decreases the BMSE and, consequently, improves the accuracy of the inference based on ABC.

5 Application to a spatial point process marked with genetic data

Here we aim to estimate, for a tree species, the pollen dispersal function with mean dispersal distance δ and shape parameter b based on data classically used in the TwoGener approach (Austerlitz and Smouse, 2002).

5.1 Data

Samples of n_m seeds are collected on each tree m of a set of M trees, called mother-trees, with known locations $\{z_m^{mother} : m=1, \dots, M\}$ and known genotypes $\{G_m^{mother} : m=1, \dots, M\}$ at a set of L microsatellite markers. Seeds are genotyped at the same loci and we know the allele frequencies in the whole population $\{f_{la} : l=1, \dots, L, a=1, \dots, A_l\}$ possibly estimated from a set of adult trees, including or not the mother-trees; A_l is the number of alleles at locus l . Knowing the genotype of a seed and the genotype of its mother, the paternal genetic contribution can be retrieved (possibly in a probabilistic way when the seed and mother are heterozygous for the same alleles; for details see Smouse et al., 2001). The genetic differentiation $\Phi_{FT,mm'}^{obs}$ between the pollen pools of two mother-trees m and m' is estimated from the paternal contributions of the seeds sampled on these two trees by applying an analysis of molecular variance (AMOVA). We also assume that we know the locations $\{z_d^{donors} : d=1, \dots, D\}$ of a sample of D pollen donors, including the mother-trees in hermaphroditic species.

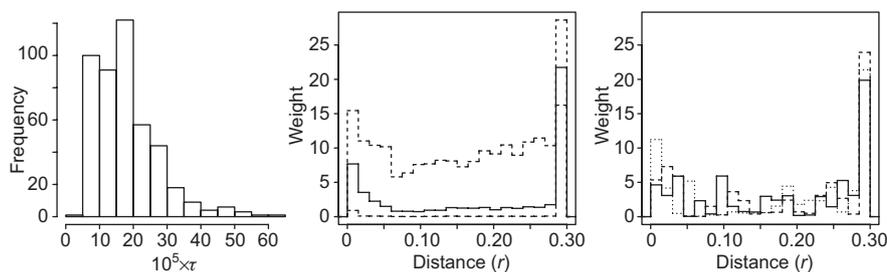


Figure 3 Modified Thomas process. Distribution of the optimum acceptance rate τ_{opt} (left; τ_{opt} is multiplied by 10^5 to get the posterior sample size). Median and quantiles of order 0.025 and 0.975 of the optimum weight function w_{opt} (middle) and three examples of w_{opt} which were obtained among the 500 runs.

As an illustrative application, we used a *Sorbus torminalis* dataset sampled in 1999 and previously analyzed in Austerlitz et al. (2004) and Oddou-Muratorio et al. (2005). In this study, 653 seeds were collected on 14 mother-trees (11–100 seeds/tree, mean: 46.6 seeds/tree) in a 472 ha plot where we know exhaustively all pollen sources (i.e. 185 adult trees for which locations are known); see Figure 4, top left. Using six microsatellite markers with six to 21 alleles per locus, we estimated all $\Phi_{FT,mm'}^{obs}$, plotted on Figure 4 (top right) as a function of pairwise spatial distances.

5.2 Model

For a given density of pollen sources outside the study plot, λ , and a given set of dispersal parameters, (δ, b) , the pairwise differentiation indices $\Phi_{FT,mm'}$ were obtained as follows. Locations for pollen donors outside the study plot $\{z_d^{out} : d=1, \dots, D^{out}\}$ were drawn from a homogeneous Poisson point process of density λ in a disk of radius 16 km (i.e. representing the whole forest) from which the study area (white area on Figure 4, top left) was removed. The set of all pollen sources then gathered the known trees inside the study plot and the simulated trees outside the study plot $\{z_d : d=1, \dots, D+D^{out}\} = \{z_1^{donors}, \dots, z_D^{donors}, z_1^{out}, \dots, z_{D^{out}}^{out}\}$. For each source tree that was not a mother-tree we simulated a genotype by drawing two alleles independently at each locus l with allele frequencies $\{f_{la} : a=1, \dots, A_l\}$.

The proportion of pollen from the source d in the pollen pool of the mother-tree m was computed as:

$$\pi_{md} = \frac{\exp\left(\frac{r_{md}^b}{a^b}\right)}{\sum_{d' \neq m} \exp\left(\frac{r_{md'}^b}{a^b}\right)},$$

where r_{md} is the distance between mother m and pollen source d , b is the shape parameter, $a = \delta \Gamma(2/b) / \Gamma(3/b)$ is the natural parameter of the exponential-power dispersal kernel used here (see Austerlitz et al., 2004) and $\delta = a \Gamma(3/b) / \Gamma(2/b)$ is the mean dispersal distance.

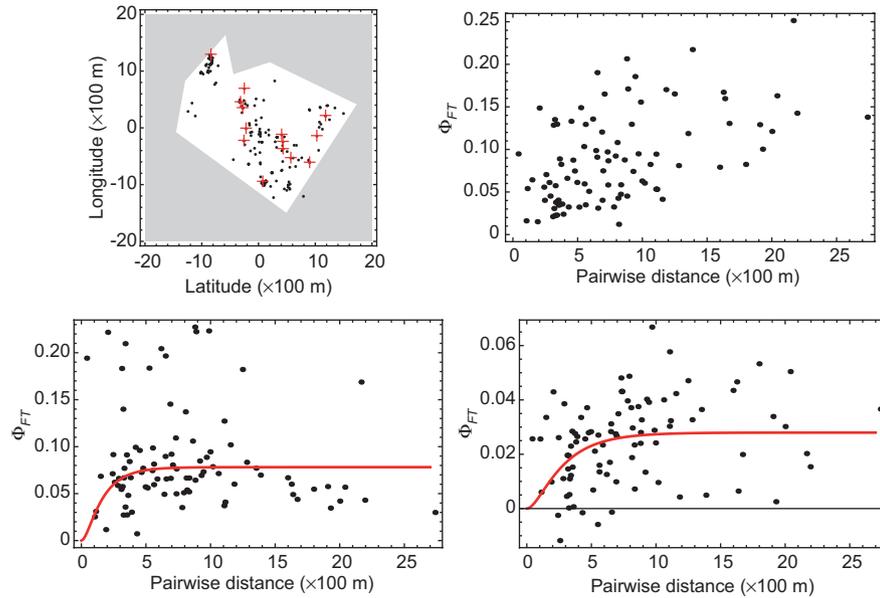


Figure 4 Top: Data for the estimation of the pollen dispersal function. Top left: Spatial distribution of the sampled mother-trees (crosses) and known additional pollen sources (dots) in the *S. torminalis* dataset. Top right: $\Phi_{FT,mm'}^{obs}$ as a function of the pairwise distance $r_{mm'}$ between mother trees estimated from the observed genotypes using an AMOVA. Bottom (left and right): Simulated clouds of points $(r_{mm'}, \Phi_{FT,mm'}^{sim})$ for two different sets of parameters and expected $\Phi_{FT,mm'}^{sim}$ (curves) assuming a Poisson process like in Austerlitz and Smouse (2002).

For each of the n_m seeds sampled on mother-tree m we drew a father-tree d randomly with proportions $\{\pi_{md}: d=1, \dots, D+D^{out}\}$ and a genotype by drawing at each locus one allele evenly among the two that each parent carries. The pairwise genetic differentiation $\Phi_{FT,mm'}^{sim}$ for mother-trees m and m' were estimated from an AMOVA applied to the retrieved paternal contributions of the sampled seeds, similarly to $\Phi_{FT,mm'}^{obs}$; two examples of simulations are provided in Figure 4, bottom.

For the prior distributions, we used a uniform distribution over $[0, 10^{-4}]$ for parameter λ , a log-normal distribution with parameters (mean: 5.75, SD: 1.73) for δ and a log-normal distribution with parameters (mean: -0.25, SD: 0.7) for b .

5.3 ABC tuning

The functional statistics which are used in this application are the $\Phi_{FT,mm'}$ computed at all the pairwise distances $r_{mm'}$ separating the 14 mother-trees of the data set. This set of pairwise distances is denoted by $\{\tilde{r}_k: k=1, \dots, 91\}$.

We ran algorithm **A3** with two simulation numbers: $(I, J)=(10^5, 10^3)$ and $(I, J)=(10^6, 10^3)$. Then, we minimized the BMSE for the three weight functions w_{cst} , w_{var} and w_{opt} ; see Eqs. (5), (6) and (7). The BMSE was computed with respect to parameters δ and b but not λ because very few information about λ is contained in the Φ_{FT} statistics that we used (data not shown). Thus, λ was randomly drawn in the interval $[0, 10^{-4}]$ to take into account the uncertainty about the density of trees outside the study plot, but was not estimated to avoid noisy results. Our interest is focused on dispersal parameters δ and b .

The optimized weight function w_{opt} was selected among the step functions with 21 jumps satisfying:

$$w(r) = \begin{cases} w_n & \text{if } r \in [r_n, r_{n+1}), \forall n \in \{0, 1, \dots, 19\} \\ 0 & \text{if } r < 0 \text{ or } r \geq r_{20}, \end{cases} \quad (13)$$

$w_0, \dots, w_{19} \geq 0$ and $\int_{\mathbb{R}} w(r) dr = \sum_{n=0}^{19} (r_{n+1} - r_n) w_n = 1$. The sequence $0=r_0, \dots, r_N$ was chosen so that the number of sampled inter-tree distances per interval $[r_n, r_{n+1})$ was approximately constant (equal to 4 or 5); r_N was set to the maximum inter-tree distance.

The distance between functional statistics S_i and S'_j was approximated by a finite sum of 91 terms:

$$\begin{aligned} d(S_i, S'_j; w) &= \int_{\mathbb{R}} w(r) \{S_i(r) - S'_j(r)\}^2 dr \\ &\approx \sum_{k=1}^{91} w(\tilde{r}_k) \{S_i(\tilde{r}_k) - S'_j(\tilde{r}_k)\}^2. \end{aligned}$$

Like in the previous application, w_{var} which can take 91 different values, is not a particular case of w_{opt} which can only take 20 different values. Thus, if equalizing the variance of the statistics is a valuable approach for this application, then w_{var} could lead to better results than w_{opt} .

5.4 Results

The BMSE obtained with the three weight functions and the two simulation numbers are provided in Table 4. The optimal weight function allows a decrease of about 4–5% of the BMSE with respect to the constant weight function (the variance weight function leads to larger BMSEs than the two other weight functions). This moderate decrease is however significant as indicated by the low p -values of the paired t-test carried out to compare the BMSEs obtained with w_{opt} and w_{cst} (the BMSEs can be compared with paired t-tests because we used the same $J=10^3$ simulations in every case).

The gain in using the optimized weight function is moderate, certainly because the functional statistics Φ_{FT} is particularly noisy compared, for instance, to the pair correlation function used for inferring parameters of the modified Thomas process.

Table 4 BMSE and PMSE obtained for the estimation of the pollen dispersal function with $I=10^5$ and $I=10^6$ simulations and with the three weight functions. In each row, the fifth column gives the p -value of the paired t-test comparing the average MSEs obtained with w_{opt} and w_{cst} .

	w_{cst}	w_{var}	w_{opt}	p -value
$I=10^5$				
BMSE	1.009	1.051	0.974	7.9×10^{-4}
PMSE (without pilot ABC)	0.101	0.102	0.100	0.57
PMSE (with pilot ABC)	0.097	0.099	0.087	5.4×10^{-5}
$I=10^6$				
BMSE	0.977	0.981	0.938	1.1×10^{-4}
PMSE (without pilot ABC)	0.092	0.094	0.089	0.11
PMSE (with pilot ABC)	0.090	0.094	0.083	1.8×10^{-4}

However, even if the gain is moderate, optimizing the weight function in the ABC procedure with only $I=10^5$ simulations led to a BMSE value (0.974) slightly lower than the BMSE value (0.977) obtained with ten times more simulations (i.e. $I=10^6$) and the constant weight function. Therefore, using w_{opt} is especially advantageous when the simulations are very time-consuming.

For this application, we also ran algorithm **A4**. For the pilot posterior sample Θ_{pilot} , we used the union of the three posterior samples obtained with the three weight functions and $I=10^5$. The partial MSEs (PMSE) were computed for the $|\mathcal{J}|=250$ closest parameter vectors from Θ_{pilot} . Table 4 provides the PMSEs obtained with and without pilot ABC. These values were computed for the same set \mathcal{J} of parameters so that they can be compared. We can see that, whatever the weight function, using the pilot ABC allows to reduce the PMSE. The reduction is about 10% in the case of w_{opt} and is significant (p -values of paired t-tests: 6.2×10^{-5} for $I=10^5$ and 0.029 for $I=10^6$). Besides, algorithm **A4** has the same properties than algorithm **A3**: The optimal weight function allows a significant decrease of the PMSE (about 10%); using $I=10^5$ simulations and the optimal weight function led to a lower PMSE than the one obtained with $I=10^6$ simulations and the constant weight function.

Figure 5 (top) shows the three weight functions obtained with $I=10^6$ and algorithm **A4** including the pilot ABC run. w_{opt}^{pilot} is fluctuating around the constant function and seems to not accumulate information over too

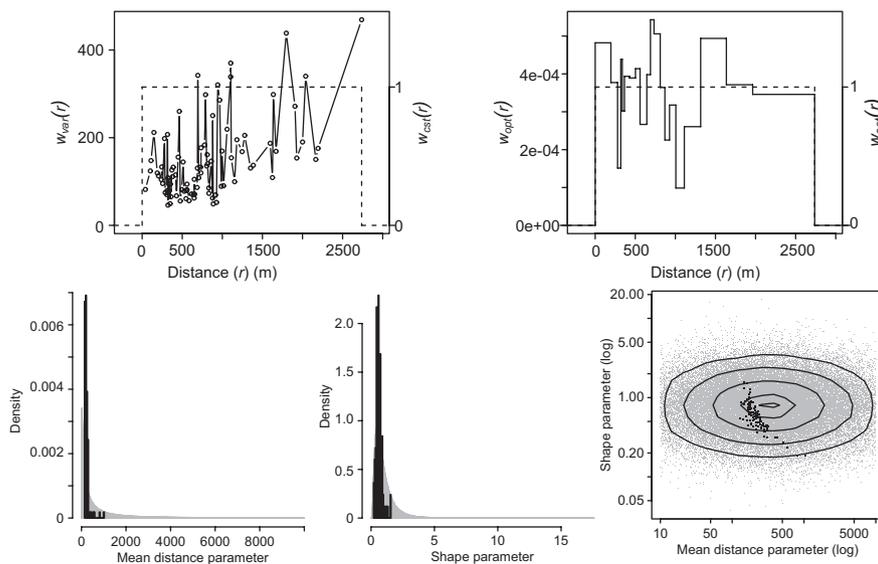


Figure 5 Estimation of the pollen dispersal function with algorithm **A4** including a pilot ABC run. Top: Constant weight function (dashed lines), variance weight function (left, dots linked by lines) and optimal weight function (right, solid line). Bottom: Marginal prior distributions (gray histogram, $I=10^6$ values) and marginal posterior distributions (black histogram, 113 values) for the mean distance parameter (left) and the shape parameter (center). Bottom right: Joint prior distribution (contour lines and gray dots; only 10^5 points are drawn) and joint posterior distribution (black dots) of the parameters.

long ranges of distances. Figure 5 (bottom) shows the (marginal and joint) prior and posterior distributions of the two parameters (the posterior sample has size 113). The joint posterior distribution clearly shows that data contain information over the dispersal parameters included in the model.

5.5 Comparison with other approaches

We carried out a comparison between algorithm **A4** and the ABC-rejection algorithm **A2** applied after a transformation of the set of summary statistics proposed in previous works. We tested three transformations: the minimum entropy approach and the two-stage procedure of Nunes and Balding (2010) and the PLS transformation of Wegmann et al. (2009). The three algorithms are denoted **A2-ME**, **A2-TS** and **A2-PLS**.

Details about the implementation of these algorithms and about the definition of the comparison criteria are provided in Appendix B. Here we only notice that in Nunes and Balding (2010) and Wegmann et al. (2009), the acceptance rate is not optimized but fixed. So, we tested two different values for the acceptance rate: $\tau=10^{-4}$ and 10^{-3} (for ABC with $I=10^6$ simulations).

Table 5 shows the results of the comparison. Algorithm **A4** is the most efficient approach with respect to the BMSE and the PMSE (criteria used in **A4**); Algorithm **A2-PLS** with $\tau=10^{-4}$ reaches the same PMSE for parameter δ but not for parameter b . Based on the mean square root of the sum of squared errors denoted MRSSE (criterion used in **A2-TS**), algorithms **A4** and **A2-TS** are the most efficient for parameter δ but **A2-TS** with $\tau=10^{-4}$ outperforms the other algorithms. We also investigated the coverage properties of the algorithm as suggested by (Wegmann et al., 2009). The coverage of δ is a little bit larger than expected with algorithm **A2-PLS**, $\tau=10^{-3}$, but the coverage of b is correct for all the algorithms.

Thus, even if the spatial dependence in the summary statistics is weak in this application (see Figure 4, top right), the overall performance of algorithm **A4** which does exploit the dependence between the statistics is satisfactory (despite the inaccuracy of the estimation of b with respect to the MRSSE).

Table 5 Comparison between algorithms **A4** and **A2** with transformation of the set of summary statistics for estimating parameters δ and b of the pollen dispersal function. Three transformations are included in algorithm **A2**: the minimum entropy approach (**A2-ME**) and the two-stage procedure (**A2-TS**) of Nunes and Balding (2010), and the PLS transformation (**A2-PLS**) of Wegmann et al. (2009). **A2-ME**, **TS** and **PLS** were carried out with two different values of the acceptance rate: $\tau=10^{-4}$ and $\tau=10^{-3}$.

Algorithm	τ	BMSE	PMSE $\times 10^{-3}$	Coverage	MRSSE
Parameter δ					
A4	τ_{opt}^{pilot}	0.408*	3.26*	0.981	0.037*
A2-ME	10^{-4}	0.467	6.14	0.955	0.052
A2-ME	10^{-3}	0.462	6.20	0.955	0.055
A2-TS	10^{-4}	0.451	3.55	0.962	0.034*
A2-TS	10^{-3}	0.445	4.02	0.977	0.035*
A2-PLS	10^{-4}	0.460	3.26*	0.970	0.058
A2-PLS	10^{-3}	0.484	4.78	0.989 ⁺	0.077
Parameter b					
A4	τ_{opt}^{pilot}	0.812*	0.193*	0.962	0.702
A2-ME	10^{-4}	0.972	0.371	0.947	0.717
A2-ME	10^{-3}	0.951	0.362	0.970	0.703
A2-TS	10^{-4}	0.862	0.234	0.936	0.475*
A2-TS	10^{-3}	0.860	0.245	0.974	0.539
A2-PLS	10^{-4}	0.886	0.234	0.959	0.537
A2-PLS	10^{-3}	0.925	0.272	0.966	0.583

The implementation of **A2-ME**, **TS** and **PLS** and the comparison criteria are described in Appendix B. In columns BMSE, PMSE and MRSSE, figures with stars are significantly lower than figures without stars (based on one-sided paired t-test with risk level 0.01). In columns Coverage, figures with the mark + indicates a coverage of the 95%-posterior interval which is significantly different from 95% (two-sided binomial test with risk level 0.01).

6 Discussion

We proposed a modification of ABC algorithms relying on a weighted distance between functional summary statistics and we optimized the weight function by minimizing a loss function computed on pseudo-observed datasets. This approach allows improved accuracy of estimation carried out with an ABC-rejection approach. We applied it to two series of 500 simulations (simple step model and modified Thomas process) and fit to a dispersal model based on pairwise genetic distances measured between individuals spread in space (real spatial point process marked with genetic data). This application allowed us to infer the parameters of the dispersal function of the wildservice tree (*S. torminalis*).

6.1 Performance of the approach

The application of the method to the pollen dispersal model (third example) showed that, with the optimized weights, we can run ten times fewer simulations (10^5 instead of 10^6) and reach an estimation accuracy equal to the one obtained with the constant weights. This result is particularly useful when simulations are very time consuming. For 10^5 simulations, the optimization took a few hours with a desk computer. Nevertheless, simulations can be parallelized with the ABC-rejection algorithm whereas parallelizing the Nelder-Mead algorithm is more difficult.

We showed that the use of a pilot ABC run lead to even better estimations (with respect to the BMSE and the PMSE). The risk with such an iterative procedure is that the pilot study results in an overly narrow region in the space of parameters.

6.2 Other approaches for transforming the summary statistics

Other approaches for transforming the raw summary statistics have been proposed (see e.g. Joyce and Marjoram, 2008; Wegmann et al., 2009; Nunes and Balding, 2010; Haon-Lasportes et al., 2011; Jung and Marjoram, 2011; Fearnhead and Prangle, 2012). In our application to the pollen dispersal model, we compared our approach with those of Nunes and Balding (2010) and Wegmann et al. (2009). Based on four comparison criteria (BMSE, PMSE, coverage rate and MRSSE), the performance of our approach was satisfactory.

The two-stage procedure of Nunes and Balding (2010) and our algorithm **A4** have comparable stages: a pilot study is used in both procedures (based on entropy in the former and BMSE in the latter) and a transformation of statistics is performed by optimizing a criterion (MRSSE in the former and PMSE in the latter) computed from outputs of the pilot ABC. The criteria used in both approaches are a matter of choice (see Section 6.3 below) and could be replaced.

Despite the similarity described above, a major difference exists between Nunes and Balding (2010) approach and ours. Our approach can be viewed as a dimension reduction of the summary statistics followed by a quantitative weighting. Indeed, the raw summary statistics are replaced by N subsums of the summary statistics and the subsums are weighted. The minimum entropy and the two-stage approaches of Nunes and Balding (2010) only perform a dimension reduction which can also be viewed like a binary weighting. The PLS transformation of Wegmann et al. (2009) also performs a binary weighting applied to the PLS axes. In this approach, the binary weighting is based on how much each axis explains the variability of the parameters.

6.3 Selection of the acceptance rate

The role of the acceptance rate τ in the agreement between the ABC-based posterior distribution and the exact posterior distribution given the statistics is studied and discussed in Blum (2010a) and Jung and Mar-

joram (2011). Choosing the value of τ consists in balancing between bias and variance of the ABC-based posterior distribution (as for the bandwidth of kernel estimators in nonparametric statistics). In practice, the acceptance rate was optimized in our approach by minimizing the BMSE and PMSE criteria with the Nelder-Mead algorithm. Jung and Marjoram (2011) also used a MSE criterion to optimize τ whereas Wilson et al. (2009, Supplementary Data) selected τ by calibrating the coverages of the parameter posterior intervals and Blum (2010b) optimized τ by maximizing the evidence function. Further studies could be carried out to compare these approaches.

6.4 Possible extensions of the approach

In the proposed method, several tuning parameters remain to be fixed by the user. The acceptance rate and the jump levels of the weight function are optimized in our algorithm whereas the numbers of simulations I and J and the jump locations of the weight function are preliminary determined by the user. When a pilot ABC run is used, the proportion of simulations which form the pilot sample has also to be determined. A further advance would be to find a way to make an automatic choice of these tuning parameters.

Note that the piecewise constant weight function and the Nelder-Mead algorithm could be replaced by other classes of functions (e.g. regular parametric functions, spline-based functions or kernel-based functions; see Ruppert et al., 2003) and optimizers (e.g. simulated annealing and genetic algorithms; see Kirkpatrick et al., 1983; Jung and Marjoram, 2011). Similarly, other choices could be made for the weighted distance in Eq. (3) (which could be replaced, for instance, by the integral of weighted absolute values) and the optimization criteria BMSE and PMSE in Eqs. (4) and (8) (which could be replaced, for example, by the mean square root of the sum of squared errors; see Nunes and Balding, 2010). Numerical comparisons would be useful to quantify the influence of these choices on estimation accuracy.

The approach that we proposed could be adapted to ABC-with-regression algorithms (Beaumont et al., 2002; Blum and François, 2010; Leuenberger and Wegmann, 2010) and to sequential ABC algorithms (Marjoram et al., 2003; Beaumont et al., 2009). For the latter algorithms, a constant weight function could be used during a first stage of the algorithm and then the weight function could be updated when enough simulations are available. This would lead to an adaptive algorithm (Haario et al., 2001).

Moreover, in spatial statistics, a single data set is commonly characterized by various functional statistics reflecting various properties (Illian et al., 2008). Our method could be easily adapted to incorporate several functional statistics in the distance between observed and simulated data sets.

6.5 Summary statistics depending on an auxiliary variable

The optimization of weights defining a distance between statistics can also be carried out for ABC methods relying on non-functional statistics as proposed by Jung and Marjoram (2011). These authors used a genetic algorithm to optimize the weights of scalar statistics as well as the acceptance rate (see above). They applied their method to inference problems with small numbers of statistics and shown that the optimized weights allowed a reduction of a mean square error computed for posterior means of parameters.

However, being able to sort the summary statistics with respect to a distance or a time or any other auxiliary variable, allows the reduction of the number of weights to be optimized and, consequently, to have less weights than the number of statistics. Thus, the proposed approach is particularly valuable when a large number of statistics correlated with respect to an auxiliary variable are available. In comparison to other methods that improve the distance between observed and simulated statistics in the ABC framework (e.g. Joyce and Marjoram, 2008; Wegmann et al., 2009; Nunes and Balding, 2010; Haon-Lasportes et al., 2011; Barnes et al., 2012; Fearnhead and Prangle, 2012), our method does exploit knowledge about the structure of dependence among statistics. However, when there is no dependence with respect to an auxiliary variable in the summary statistics, other methods might outperform ours.

Finally, situations with large numbers of statistics correlated with respect to auxiliary variables are frequently encountered in spatial studies where statistical analyses are carried out with spatially explicit (genetic) models. Therefore, our approach should especially help in developing the use of ABC in spatial studies like those presented in this article (up to now, ABC has been lightly applied in such studies).

Acknowledgements: We thank the reviewers for their useful suggestions and comments. This research was supported by the French research agency ANR (EMILE project).

7 Appendix

A Approximation of the posterior distribution of θ given S

This appendix shows in a simple case that $p_\varepsilon(\theta|S)$ given by Eq. (1) approximates the posterior distribution conditional on the statistics $p(\theta|S)$ given by Eq. (2). This simple case is based on regularity assumptions about the conditional probability distribution function $S \mapsto \tilde{f}(S|\theta)$ of S given θ which cannot be checked in usual applications of ABC where \tilde{f} is generally analytically intractable.

Here, (i) θ and S are fixed, (ii) the space \mathbb{S} of statistics is \mathbb{R} , (iii) the conditional probability distribution function $S \mapsto \tilde{f}(S|\theta)$ of S given θ is three times differentiable over \mathbb{R} , (iv) the absolute values of its second and third derivatives are π -integrable over Θ , (v) its third derivative is a Lipschitz function ($\exists A \in [0, \infty)$, $\forall (u, S) \in \mathbb{S}$, $|\tilde{f}'''(u|\theta) - \tilde{f}'''(S|\theta)| \leq A|u - S|$).

From the Taylor's theorem,

$$\begin{aligned} \int_{B(S, \varepsilon)} \tilde{f}(z|\theta) dz &= \int_{S-\varepsilon}^{S+\varepsilon} \tilde{f}(z|\theta) dz \\ &= 2\varepsilon \tilde{f}(S|\theta) + \frac{\varepsilon^3}{3} \tilde{f}'''(S|\theta) \\ &\quad + \int_{S-\varepsilon}^{S+\varepsilon} \int_{S-\varepsilon}^{S+\varepsilon} \frac{(z-u)^2}{2} \tilde{f}'''(u|\theta) du dz \\ &= 2\varepsilon \tilde{f}(S|\theta) + R(\varepsilon, S, \theta). \end{aligned}$$

Using assumption (v), the absolute value of the remainder term $R(\varepsilon, S, \theta)$ is bounded from above by:

$$|R(\varepsilon, S, \theta)| \leq \varepsilon^3 \{ |\tilde{f}'''(S|\theta)|/3 + 4|f'''(S|\theta)| + 4A\varepsilon \}.$$

Let $r(\varepsilon, S, \theta) = R(\varepsilon, S, \theta)/\varepsilon^3$. Using assumption (iv), the upper bound of $|R(\varepsilon, S, \theta)|$ is π -integrable over Θ as well as $\theta \mapsto r(\varepsilon, S, \theta)$ and, consequently,

$$\begin{aligned} p_\varepsilon(\theta|S) &= \frac{\left(\int_{B(S, \varepsilon)} \tilde{f}(z|\theta) dz \right) \pi(\theta)}{\int_{\Theta} \left(\int_{B(S, \varepsilon)} \tilde{f}(z|\alpha) dz \right) \pi(\alpha) d\alpha} \\ &= \frac{\tilde{f}(S|\theta) \pi(\theta) + \varepsilon^2 r(\varepsilon, S, \theta) \pi(\theta)}{\int_{\Theta} \tilde{f}(S|\alpha) \pi(\alpha) d\alpha + \varepsilon^2 \int_{\Theta} r(\varepsilon, S, \theta) \pi(\alpha) d\alpha} \\ &= \frac{\tilde{f}(S|\theta) \pi(\theta)}{\int_{\Theta} \tilde{f}(S|\alpha) \pi(\alpha) d\alpha} (1 + O(\varepsilon^2)) \\ &= p(\theta|S) (1 + O(\varepsilon^2)). \end{aligned}$$

Therefore, when ε tends to zero, $p_\varepsilon(\theta|S)$ approximates the posterior distribution conditional on the statistics $p(\theta|S)$.

B Implementation of algorithms A2-ME, TS and PLS and comparison criteria

In Subsection 5.5, algorithm **A2** was applied to three different sets of summary statistics: (i) a subset of the raw statistics obtained with the minimum entropy approach of Nunes and Balding (2010), (ii) a subset of the raw statistics obtained with the two-stage procedure of Nunes and Balding (2010) and (iii) a subset of the axes obtained after a PLS regression between the parameters and the statistics like in Wegmann et al. (2009). The three algorithms are denoted **A2-ME**, **A2-TS** and **A2-PLS**. Note that these approaches can be carried out with the *abctools* R package proposed by Nunes and Prangle and the *ABCtoolbox* software of (Wegmann et al., 2010).

For the ME (resp. TS) selection of the statistics, the exhaustive search of the subset of statistics which minimizes the entropy (resp. mean square root of the sum of squared errors, denoted MRSSE) over all the possible subsets was not feasible (for 91 statistics there are almost 2.5×10^{27} possible subsets). Therefore, we replaced the exhaustive search by an iterative search: we used a forward search based on the entropy (resp. MRSSE). At iteration one, the subset is made of the statistic which leads to the lowest entropy (resp. MRSSE). Then, at each of the following iterations, the current subset is completed by the statistic which leads, when it is merged to the current subset of statistics, to the lowest entropy (resp. MRSSE). The iterative search is stopped when the entropy (resp. MRSSE) is no more decreasing by the addition of any of the remaining statistics. For **A2-ME**, 5 (resp. 7) statistics were selected with $\tau=10^{-4}$ (resp. $\tau=10^{-3}$). For **A2-TS**, 11 (resp. 14) statistics were selected with $\tau=10^{-4}$ (resp. $\tau=10^{-3}$).

It has to be noted that, in **A2-TS**, the MRSSE that we used includes a standardization to rescale the components of the parameter vector (as suggested by Nunes and Balding, 2010) and satisfies:

$$\text{MRSSE} = \frac{1}{|\mathcal{J}^{EM}|} \sum_{j \in \mathcal{J}^{EM}} \left(\frac{1}{n_{acc}} \sum_{m=1}^{n_{acc}} \sum_{k=1}^K \frac{\{\Theta_m^{EM}(jk) - \theta_{jk}^{EM}\}^2}{\text{Var}(\theta_{jk})} \right)^{1/2}.$$

where \mathcal{J}^{EM} is the set of 100 PODS selected in stage one of the two-stage procedure, $\{\Theta_m^{EM}(j) = (\Theta_m^{EM}(j,1), \dots, \Theta_m^{EM}(j,K))': m=1, \dots, n_{acc}\}$ is the posterior sample (set of accepted parameter vectors) obtained for each PODS in \mathcal{J}^{EM} obtained after stage one of the two-stage procedure. The size of the posterior sample is $n_{acc} = I\tau = 10^6 \tau$ where τ is the acceptance threshold (10^{-4} or 10^{-3}).

For **A2-PLS**, we fitted the PLS regression to 10^4 simulations and kept the minimum number of axes explaining at least 99% of the variance of the parameters. This led us to keep 24 axes among 91 possible axes.

We computed marginal criteria measuring the accuracy of the estimation for each parameter component. The marginal BMSE was computed with 1000 new pseudo-observed data sets (PODS) not used in the implementation of the algorithms which are compared. Among the 1000 new PODS, only 266 were used to compute the marginal PMSE, the marginal coverage of the 95%-posterior intervals of δ and b and the marginal mean square root of the sum of squared errors [MRSSE; criterion used in the two-stage approach of Nunes and Balding (2010)]. The 266 PODS were obtained as follows: For each value of τ (10^{-4} and 10^{-3}), we selected the 250 PODS with the summary statistics reduced by the ME approach which are the closest from the observed summary statistics reduced by the ME approach (the closeness is quantified with the Euclidean distance). Then we merged the two sets of PODS and obtained a set of 266 different PODS. This selection of PODS is similar to stage one in the TS procedure of Nunes and Balding (2010).

Let (θ_j^*, S_j^*) , $j=1, \dots, 10^3$ denote the 1000 new PODS. The marginal BMSE satisfies:

$$\text{BMSE}^{marg} = \frac{1}{10^3} \sum_{j=1}^{10^3} \frac{(\hat{\theta}_{jk}^* - \theta_{jk}^*)^2}{\text{Var}(\theta_{jk}^*)},$$

where $\hat{\theta}_{jk}^*$, $k=1, \dots, K$, are the marginal posterior medians of the K components of $\theta_j^* = (\theta_{j1}^*, \dots, \theta_{jK}^*)$, the marginal posterior medians being obtained with either **A4**, **A2-ME**, **A2-TS** or **A2-PLS**.

Let \mathcal{J}^* denote the set of 266 selected PODS. The marginal PMSE satisfies:

$$\text{PMSE}^{\text{marg}} = \frac{1}{|\mathcal{J}^*|} \sum_{j \in \mathcal{J}^*} \frac{(\hat{\theta}_{jk}^* - \theta_{jk}^*)^2}{\text{Var}(\theta_{jk}^*)}.$$

The marginal coverage of any parameter by the corresponding marginal 95%-posterior interval is:

$$\text{Coverage}^{\text{marg}} = \frac{1}{|\mathcal{J}^*|} \sum_{j \in \mathcal{J}^*} \mathbf{1}(\theta_{jk}^* \leq \bar{\theta}_{jk}^*),$$

where $\underline{\theta}_{jk}^*$ and $\bar{\theta}_{jk}^*$ are the posterior quantiles of order 0.025 and 0.975 of the k -th component of θ_j^* , the posterior quantiles being obtained with either **A4**, **A2-ME**, **A2-TS** or **A2-PLS**, and $\mathbf{1}$ is the indicator function [$\mathbf{1}(E)=1$ if event E holds, zero otherwise]. The marginal MRSSE satisfies:

$$\text{MRSSE}^{\text{marg}} = \frac{1}{|\mathcal{J}^*|} \sum_{j \in \mathcal{J}^*} \left(\frac{1}{n_{\text{acc}}} \sum_{m=1}^{n_{\text{acc}}} \frac{\{\Theta_m^*(jk) - \theta_{jk}^*\}^2}{\text{Var}(\theta_{jk}^*)} \right)^{1/2}.$$

where $\{\Theta_m^*(j) = (\Theta_m^*(j,1), \dots, \Theta_m^*(j,K))': m=1, \dots, n_{\text{acc}}\}$ denotes the posterior sample (set of accepted parameter vectors) obtained by applying either **A4**, **A2-ME**, **A2-TS** or **A2-PLS** to the j -th PODS (to infer θ_j^*). The size of the posterior sample is $n_{\text{acc}} = I\tau = 10^6 \tau$ where $\tau = \tau_{\text{opt}}^{\text{pilot}}$ in algorithm **A4** and $\tau = 10^{-4}$ or 10^{-3} in algorithms **A2-ME**, **TS** and **PLS**.

References

- Austerlitz, F. and P. E. Smouse (2002) "Two-generation analysis of pollen flow across a landscape. iv. estimating the dispersal parameter," *Genetics*, 161, 355.
- Austerlitz, F., C. W. Dick, C. Dutech, E. K. Klein, S. Oddou-Muratorio, P. E. Smouse and V. L. Sork (2004) "Using genetic markers to estimate the pollen dispersal curve," *Mol. Ecol.*, 13, 937–954.
- Barnes, C. P., S. Filippi, M. P. H. Stumpf and T. Thorne (2012) "Considerate approaches to constructing summary statistics for ABC model selection," *Stat. Comput.*, 22, 1181–1197.
- Beaumont, M. A. (2010) "Approximate bayesian computation in evolution and ecology," *Annu. Rev. Ecol. Evol. Syst.*, 41, 379–406.
- Beaumont, M. A., W. Zhang and D. J. Balding (2002) "Approximate bayesian computation in population genetics," *Genetics*, 162, 2025–2035.
- Beaumont, M. A., J.-M. Cornuet, J.-M. Marin and C. Robert (2009) "Adaptivity for ABC algorithms: the ABC-PMC scheme," *Biometrika* (in press), 96, 983–990.
- Blum, M. G. B. (2010a) "Approximate bayesian computation: a nonparametric perspective," *J. Am. Stat. Assoc.*, 205, 1178–1187.
- Blum, M. G. B. (2010b) Choosing the summary statistics and the acceptance rate in approximate bayesian computation. In: Lechevallier, Y., Saporta, G. (Eds.), *Proceedings of COMPSTAT'2010*. Physica-Verlag, pp. 47–56.
- Blum, M. G. B. and O. François (2010) "Non-linear regression models for approximate bayesian computation," *Stat. Comput.*, 20, 63–73.
- Blum, M. G. B., M. A. Nunes, D. Prangle and S. A. Sisson (2012) "A comparative review of dimension reduction methods in approximate bayesian computation," *Arxiv preprintarXiv: 1202.3819*.
- Carpentier, F. (2010) *Modélisations de la dispersion du pollen et estimation à partir de marqueurs génétiques*, Ph.D. thesis, Université Montpellier 2.
- Chilés, J.-P. and P. Delfiner (1999) *Geostatistics. Modeling Spatial Uncertainty*. New York: Wiley.
- Cressie, N. A. C. (1991) *Statistics for Spatial Data*. New York: Wiley.
- Csilléry, K., M. G. B. Blum, O. E. Gaggiotti and O. François (2010) "Approximate bayesian computation (ABC) in practice," *Trends Ecol. Evol.*, 25, 410–418.
- Csilléry, K., O. François and M. Blum (2011) "Abc: an R package for Approximate Bayesian Computation (ABC)," *Arxiv preprint arXiv:1106.2793*.
- Fearnhead, P. and D. Prangle (2012) "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation," *J. R. Stat. Soc. B*, 74, 419–474.
- Haario, H., E. Saksman and J. Tamminen (2001) "An adaptive metropolis algorithm," *Bernoulli*, 7, 223–242.

- Haon-Lasportes, E., F. Carpentier, O. Martin, E. K. Klein and S. Soubeyrand (2011) Conditioning on parameter point estimates in approximate bayesian computation. Research Report. INRA, Biostatistics and Spatial Processes Research Unit.
- Hardy, O. J. (2003) "Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers," *Mol. Ecol.*, 12, 1577–1588.
- Illian, J., A. Penttinen, H. Stoyan and D. Stoyan (2008) *Statistical Analysis and Modelling of Spatial PointPatterns*. New York: Wiley.
- Joyce, P. and P. Marjoram (2008) "Approximately sufficient statistics and bayesian computation," *Stat. Appl. Genet. Mol. Biol.*, 7, 1–16.
- Jung, H. and P. Marjoram (2011) "Choice of summary statistic weights in approximate bayesian computation," *Stat. Appl. Genet. Mol. Biol.*, 10, 1–23.
- Kirkpatrick, S., C. D. Gelatt Jr. and M. P. Vecchi (1983) "Optimization by simulated annealing," *Science* 220, 671–680.
- Leuenberger, C. and D. Wegmann (2010) "Bayesian computation and model selection without likelihoods," *Genetics*, 184, 243–252.
- Marin, J. M., P. Pudlo, C. P. Robert and R. Ryder (2011) "Approximate bayesian computational methods," *J. Stat. Comput.*, 22, 1167–1180.
- Marjoram, P., V. Plagnol and S. Tavaré (2003) "Markov chain Monte Carlo without likelihoods," *PNAS*, 100, 15324–15328.
- McCulloch, C. E. and S. R. Searle (2001) *Generalized, Linear, and Mixed Models*. New York: Wiley.
- Nelder, J. A. and R. Mead (1965) "A simplex method for function minimization," *Comput. J.*, 7, 308–313.
- Nunes, M. A. and D. J. Balding (2010) "On optimal selection of summary statistics for approximate bayesian computation," *Stat. Appl. Genet. Mol. Biol.*, 9, 1–14.
- Oddou-Muratorio, S., E. K. Klein and F. Austerlitz (2005) "Pollen flow in the wildservice tree, *sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent–offspring analysis," *Mol. Ecol.*, 14, 4441–4452.
- Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun and M. W. Feldman (1999) "Population growth of human y chromosomes: a study of y chromosome microsatellites," *Mol. Biol. Evol.*, 16, 1791–1798.
- Robledo-Arnuncio, J. J. and F. Austerlitz (2006) "Pollen dispersal in spatially aggregated populations," *The American Naturalist*, 168, 500–511.
- Robledo-Arnuncio, J. J., F. Austerlitz and P. E. Smouse (2006) "A new method of estimating the pollen dispersal curve independently of effective density," *Genetics*, 173, 1033–1045.
- Rohatgi, V. K. (2003) *Statistical Inference*. Mineola, NY: Dover Publications.
- Rousset, F. (1997) "Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance," *Genetics*, 145, 1219.
- Rousset, F. (2000) "Genetic differentiation between individuals," *J. Evol. Biol.*, 13, 58–62.
- Rousset, F. and R. Leblois (2007) "Likelihood and approximate likelihood analyses of genetic structure in a linear habitat: performance and robustness to model mis-specification," *Mol. Biol. Evol.*, 24, 2730–2745.
- Rubin, D. B. (1984) "Bayesianly justifiable and relevant frequency calculations for the applied statistician," *Ann. Stat.*, 12, 1151–1172.
- Ruppert, D., M. P. Wand and R. J. Carroll (2003) *Semiparametric Regression*. Cambridge: Cambridge University Press.
- Smouse, P. E., R. J. Dyer, R. D. Westfall and V. L. Sork (2001) "Two-generation analysis of pollen flow across a landscape .i. malegamete heterogeneity among females," *Evolution*, 55, 260–271.
- Stoyan, D. and H. Stoyan (1994) *Fractals, Random Shapes and Pointfields: Methods of Geometrical Statistics*. New York: Wiley.
- Wegmann, D., C. Leuenberger and L. Excoffier (2009) "Efficient Approximate Bayesian Computation coupled with Markov chain Monte Carlo without likelihood," *Genetics*, 182, 1207–1218.
- Wegmann, D., C. Leuenberger, S. Neuenschwander and L. Excoffier (2010) "Abctoolbox: a versatile toolkit for approximate bayesian computations," *BMC Bioinformatics*, 11, 116. INRA, UR546 Biostatistics and Spatial Processes, F-84914 Avignon, France
- Wilson, D. J., E. Gabriel, A. J. H. Leatherbarrow, J. Cheesbrough, S. Gee, E. Bolton, A. Fox, C. A. Hart, P. J. Diggle and P. Fearnhead (2009) "Rapid evolution and the importance of recombination to the gastroenteric pathogen *campylobacter jejuni*," *Mol. Biol. Evol.*, 26, 385–397.