ABC avec statistiques fonctionnelles (et non-fonctionnelles)

Application en statistique spatiale

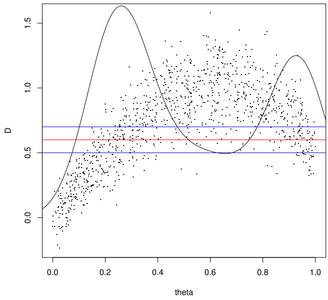
Samuel Soubeyrand
INRA PACA - BioSP

Applibugs – Lyon – 24 Juin 2016

ABC – Basic rejection algorithm

Perform the next 2 steps for i in $\{1, \ldots, I\}$, independently:

- Generate θ_i from π and simulate $\mathcal{D}_i$ from $\mathcal{M}_{\theta_i}$;
- Accept θ_i if $d_{\mathbb{D}}(\mathcal{D}, \mathcal{D}_i) \leq \epsilon$, where $d_{\mathbb{D}}$ is a distance over $\mathbb{D}$ and ϵ is a tolerance threshold for the distance between the observed data and the simulated ones.



ABC – Basic rejection algorithm

Perform the next 2 steps for i in $\{1, \ldots, I\}$, independently:

- Generate θ_i from π and simulate $\mathcal{D}_i$ from $\mathcal{M}_{\theta_i}$;
- Accept θ_i if $d_{\mathbb{D}}(\mathcal{D}, \mathcal{D}_i) \leq \epsilon$, where $d_{\mathbb{D}}$ is a distance over $\mathbb{D}$ and ϵ is a tolerance threshold for the distance between the observed data and the simulated ones.

The set $\Theta_{\epsilon,I} = \{\theta_i : d_{\mathbb{D}}(\mathcal{D}_i, \mathcal{D}) \leq \epsilon, i = 1, \ldots, I\}$ of accepted parameters forms a sample from the distribution:

$$p_{d_{\mathbb{D}},\epsilon}(\theta \mid \mathcal{D}) = \frac{\left(\int_{\mathcal{B}_{d_{\mathbb{D}}}(\mathcal{D},\epsilon)} h(x \mid \theta) dx\right) \pi(\theta)}{\int_{\Theta} \left(\int_{\mathcal{B}_{d_{\mathbb{D}}}(\mathcal{D},\epsilon)} h(x \mid \alpha) dx\right) \pi(\alpha) d\alpha}$$

with $\mathcal{B}_{d_{\mathbb{D}}}(\mathcal{D}, \epsilon)$: ball centered around $\mathcal{D}$

$x \mapsto h(x \mid \theta)$: p.d.f. of $\mathcal{D}^*$ drawn under $\mathcal{M}_\theta$

What is the target distribution in ABC–rejection?

Target distribution:

$$p_{d_{\mathbb{D}},\epsilon}(\theta \mid \mathcal{D}) = \frac{\left(\int_{\mathcal{B}_{d_{\mathbb{D}}}(\mathcal{D},\epsilon)} h(x \mid \theta)dx\right)\pi(\theta)}{\int_{\Theta}\left(\int_{\mathcal{B}_{d_{\mathbb{D}}}(\mathcal{D},\epsilon)} h(x \mid \alpha)dx\right)\pi(\alpha)d\alpha}$$

- Under regularity assumptions and an appropriate distance $d_{\mathbb{D}}$, $p_{d_{\mathbb{D}},\epsilon}(\theta \mid \mathcal{D})$ is a good approximation, when $\epsilon \to 0$, of the classical posterior distribution:

$$p(\theta \mid \mathcal{D}) = \frac{h(\mathcal{D} \mid \theta)\pi(\theta)}{\int_{\Theta} h(\mathcal{D} \mid \alpha)\pi(\alpha)d\alpha}$$

- More generally, $p_{d_{\mathbb{D}},\epsilon}(\theta \mid \mathcal{D})$ may be a good approximation, when $\epsilon \to 0$, of:

$$\frac{\left(\int_{\mathcal{V}_{d_{\mathbb{D}}}(\mathcal{D})} h(x \mid \theta)dx\right)\pi(\theta)}{\int_{\Theta}\left(\int_{\mathcal{V}_{d_{\mathbb{D}}}(\mathcal{D})} h(x \mid \alpha)dx\right)\pi(\alpha)d\alpha}$$

Construction of the distance in ABC

- ABC is carried out by defining a distance between observed and simulated data sets:

$$d_{\mathbb{D}}(\mathcal{D}, \mathcal{D}_i)$$

- Classically, the distance is based on a finite set of summary statistics (to cirvumvent the curse of dimensionality):

$$\begin{aligned} s : \mathbb{D} &\to \mathbb{S} \\ S &= s(\mathcal{D}) \\ S_i &= s(\mathcal{D}_i) \\ d_{\mathbb{D}}(\mathcal{D}, \mathcal{D}_i) &= d_{\mathbb{S}}(S, S_i) \end{aligned}$$

- The definition of $(s, d_{\mathbb{S}})$ determines the information taken into account in the ABC procedure and, consequently, the inference accuracy

Approaches for defining $(s, d_{\mathbb{S}})$

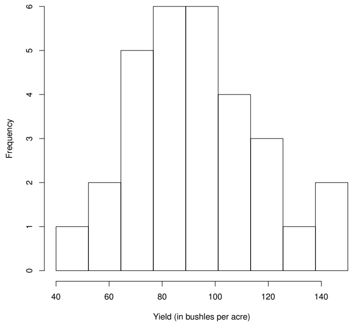
- Simply gathering a set of relevant statistics expected to be related with parameters
  - Use of the raw statistics and a mean square distance
  - Variance equalization (Beaumont et al 2002)
- "Optimization" approaches
  - Transformation into "axes" (PLS, ACP; Wegmann et al 2009)
  - Dimension reduction (or binary weighting; Barnes et al 2012, Joyce and Marjoram 2008, Nunes and Balding 2010)
  - **Optimal weighting** (Soubeyrand et al. 2013)
  - Regression-based point estimates of parameters (PEP): $\hat{\theta}(S)$ (Fearnhead and Prangle 2012, Haon-Lasportes et al. 2011)
  - Model-based PEPs (e.g. pseudo-likelihood estimates) and optimal weighting (Soubeyrand and Haon-Lasportes, 2015)

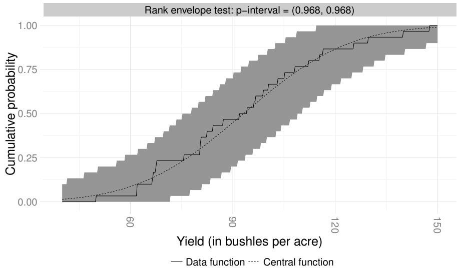
Functional statistics

- Functional statistics are convenient objects to describe variations in time, space and other ordered domains
- They are often used for:
  - describing patterns
  - testing hypotheses
  - fitting models

Example in distribution theory

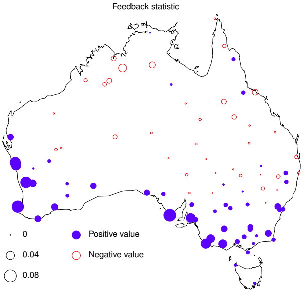
Cumulative distribution function:

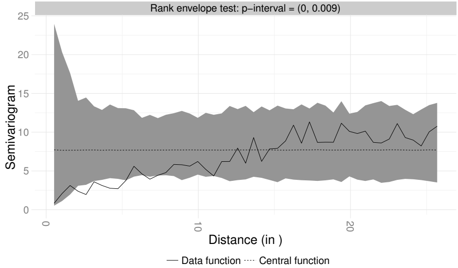




Example in geostatistics

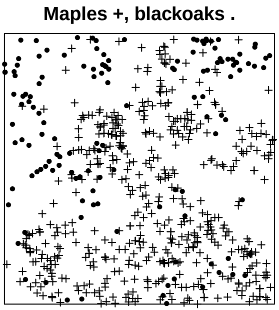
Semivariogram:

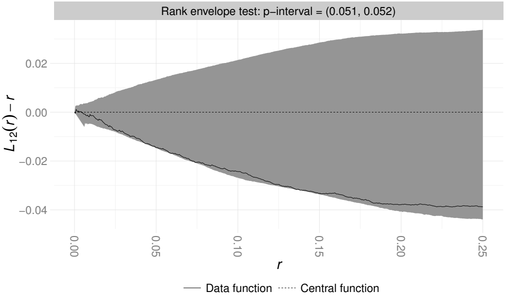




Example for spatial point processes

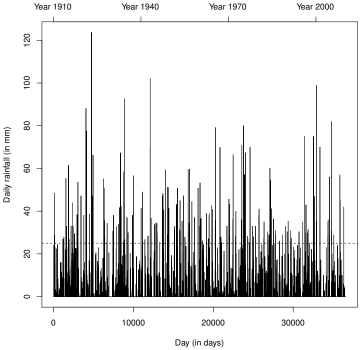
L_{12}-function:

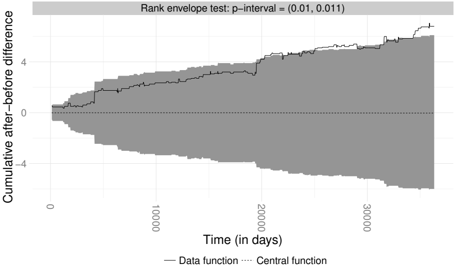




Example in time series

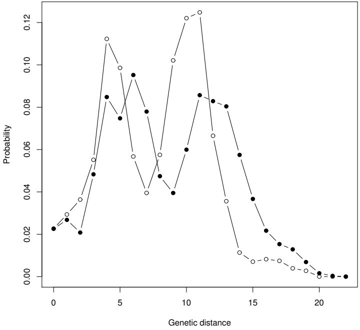
Cumulative after-before difference:

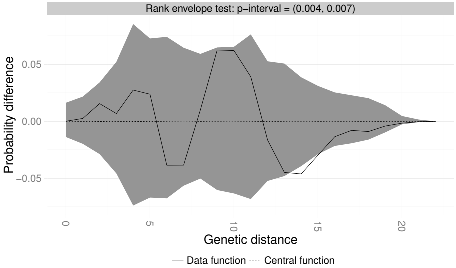




Example in genetics

Genetic distance:





ABC and Functional statistics

How to use functional statistics as summary statistics in ABC?

- Handling the infinite dimension
- Handling the dependencies along the support of the function

Contents

Exact ABC–rejection algorithm (Rubin, 1984)

A1. Carry out the next two steps, independently for i in $\{1, \ldots, I\}$,

1. Generate θ_i from π and simulate $\mathcal{D}_i$ from $\mathcal{M}_{\theta_i}$.
2. Accept θ_i if $\mathcal{D}_i = \mathcal{D}$, reject it otherwise.

- Limitation: $P(\mathcal{D}_i = \mathcal{D})$ is low for high-dimension data and zero for continuous data
- Solution:
  - Use of a tolerance threshold ($P(\mathcal{D}_i \approx \mathcal{D})$)
  - Use of summary statistics (dimension reduction)

ABC–rejection algorithm (Pritchard et al., 1999)

A2. Carry out the next three steps, independently for i in $\{1, \ldots, I\}$,

1. Generate θ_i from π and simulate $\mathcal{D}_i$ from $\mathcal{M}_{\theta_i}$.
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 $d(S_i, S) \leq \epsilon(\tau)$, where d is a distance over $\mathbb{S}$ and $\epsilon(\tau) \in \mathbb{R}_+$ is a tolerance threshold for the distance between the observed statistics $S = s(\mathcal{D})$ and the simulated ones.

$\epsilon(\tau)$ depends on the proportion τ of accepted θ_i among the I simulated parameters (τ is called the acceptance rate)

- Question: What distance d when S is a functional statistic?
- Solution: Use of an optimized weighted distance

Weighted distance for functional statistics

- Functional statistics included in:

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

- Distance between functional statistics:

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

with $w : \mathbb{R} \to \mathbb{R}_+$

- Three weight functions:
  - Constant function:

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

  - Inverse variance function (Beaumont et al., 2002):

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

  - Optimal function in $\mathbb{W} = \{w : \mathbb{R} \to \mathbb{R}_+, \int_{\mathbb{R}} w = 1\}$

ABC-rejection algorithm with functional statistics

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 $\mathcal{M}_{\theta_i}$ 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) will be used as pseudo-observed data sets (PODS);
3. Select the weight function and the acceptance rate which minimize the BMSE criterion:

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

$$\text{BMSE}_J(w, \tau) = \frac{1}{J} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(\hat{\theta}'_{jk}(w, \tau) - \theta'_{jk})^2}{\mathbb{V}(\theta'_{jk})}$$

   θ'_{jk}: k-th component of θ'_j
   $\mathbb{V}(\theta'_{jk})$: prior variance of θ'_{jk}
   $\hat{\theta}'_{jk}(w, \tau)$ point estimates (e.g. marginal posterior medians) of θ'_{jk} obtained with **A2** applied to S'_j and $\{(\theta_i, S_i) : i = 1, \dots, I\}$
4. For i in $\{1, \dots, I\}$, accept θ_i if $d(S_i, S; w_{opt}) \leq \epsilon(\tau_{opt})$.

What is the target distribution?

- The set $\Theta_{opt} = \{\theta_i : d(S_i, S; w_{opt}) \leq \epsilon(\tau_{opt}), i = 1, \ldots, I\}$ of accepted parameters forms a sample from the posterior:

$$p_{d(\cdot,\cdot;w_{opt}),\epsilon(\tau_{opt})}(\theta \mid S) = \frac{\left(\int_{\mathcal{B}_{d(\cdot,\cdot;w_{opt})}(S,\epsilon(\tau_{opt}))} f(x \mid \theta)dx\right)\pi(\theta)}{\int_\Theta \left(\int_{\mathcal{B}_{d(\cdot,\cdot;w_{opt})}(S,\epsilon(\tau_{opt}))} f(x \mid \alpha)dx\right)\pi(\alpha)d\alpha}$$

  with $x \mapsto f(x \mid \theta)$: p.d.f. of $S^* = s(\mathcal{D}^*)$ where $\mathcal{D}^*$ is drawn under $\mathcal{M}_\theta$
- Weighting the distance modifies the posterior under which the accepted parameters are drawn
- However, under regularity conditions and when $\epsilon(\tau_{opt}) \to 0$, the new posterior may be a good approximation of $p(\theta \mid S)$

Remarks

- Tuning components: $s : \mathcal{D} \mapsto S$; I (ABC simuls); J (PODS); d (weighted squared difference); BMSE; optimization algo.
- Typically, I around 10^5 or 10^6 and $J = 10^3$
- Optimization algorithm:
  - w restricted to piecewise constant functions with a finite number of jumps whose locations are known
  - Constrained Nelder-Mead algorithm
- Incorporation of a pilot ABC run for restricting the sets of PODS:

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

- Optimization of the acceptance rate when w_{cst} or w_{var} is used

$$\begin{aligned}\tau_{cst} &= \mathrm{argmin}_{\tau\in(0,1]}\mathrm{BMSE}_J(w_{cst},\tau)\\ \tau_{var} &= \mathrm{argmin}_{\tau\in(0,1]}\mathrm{BMSE}_J(w_{var},\tau)\end{aligned}$$

Application 1: simple step model

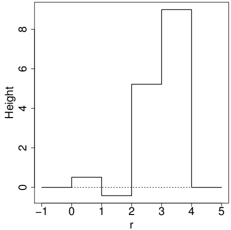
- Functional statistic:

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

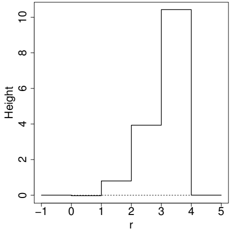
$$\varepsilon(n) \underset{indep.}{\sim} \mathcal{N}(0, \sigma(n)), \qquad n = 1, \dots, 4$$

- This function has 4 positive steps whose heights are: $\varepsilon(0)$, $\theta + \varepsilon(1)$, $4\theta + \varepsilon(2)$ and $9\theta + \varepsilon(3)$
- The first step $S(0) = \varepsilon(0)$ does not bring information on θ
- Three noise structures $(\sigma(0), \sigma(1), \sigma(2), \sigma(3))$:

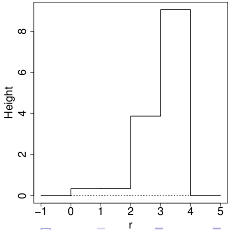
Constant
$(1, 1, 1, 1)$



Increasing
$(0.05, 0.1, 0.5, 1)$



Decreasing
$(1, 0.5, 0.1, 0.05)$



ABC tuning

- $I = 10^5$, $J = 10^3$
- Weight function:

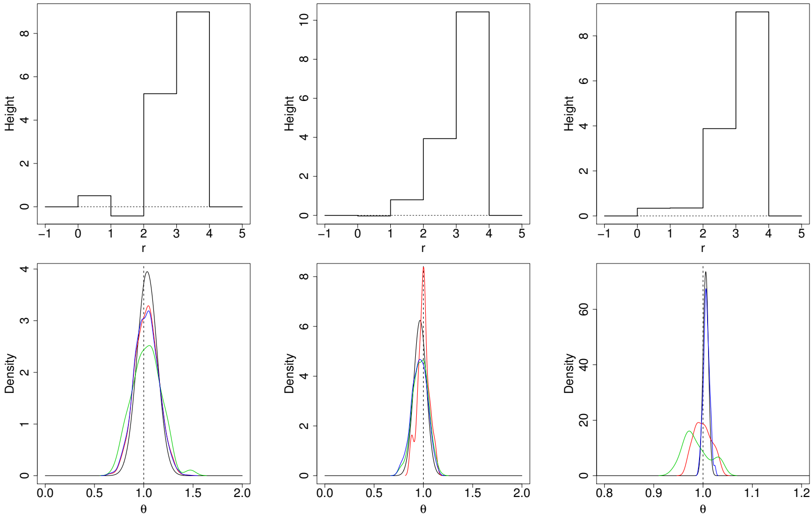
$$w(r) = \begin{cases} w_n & \text{if } r \in [n, n+1), \ \forall n \in \{0,1,2,3\} \\ 0 & \text{otherwise,} \end{cases}$$

  $w_0, w_1, w_2, w_3 \geq 0$ and $\sum_{n=0}^{3} w_n = 1$
- Distance function:

$$\begin{aligned} d(S_i, S; w) &= \int_{\mathbb{R}} w(r)\{S_i(r) - S(r)\}^2 dr \\ &= \sum_{n=0}^{3} w_n \{S_i(n) - S(n)\}^2. \end{aligned}$$

Three ABC runs

Constant noise | Increasing noise | Decreasing noise



Black: exact Bayes Red: w_{cst} Green: w_{var} Blue: w_{opt}

Series of ABC runs

Average BMSE ($\times$1000) and SD for the simple step model (based on 500 runs) and # of times that each weight fct provided the lowest BMSE:

	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

Mean values and SD of the optimum acceptance rate τ_{opt} ($\times 10^5$) and weight function w_{opt} for the simple step model computed from 500 runs:

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

Series of ABC runs

Average BMSE ($\times$1000) and SD for the simple step model (based on 500 runs) and # of times that each weight fct provided the lowest BMSE:

	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

Mean values and SD of the optimum acceptance rate τ_{opt} ($\times 10^5$) and weight function w_{opt} for the simple step model computed from 500 runs:

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

Series of ABC runs

Average BMSE ($\times$1000) and SD for the simple step model (based on 500 runs) and # of times that each weight fct provided the lowest BMSE:

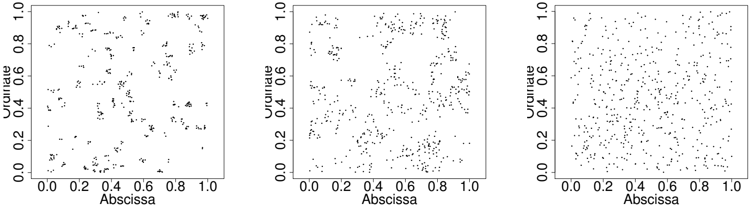
	w_{cst}	w_{var}	w_{opt}
Constant noise	9.30 (0.44)	10.02 (0.47)	9.27 (0.44)
	0	0	500
Increasing noise	4.23 (0.20)	3.90 (0.18)	3.85 (0.17)
	0	0	500
Decreasing noise	0.044 (0.002)	0.259 (0.019)	0.030 (0.001)
	0	0	500

Mean values and SD of the optimum acceptance rate τ_{opt} ($\times 10^5$) and weight function w_{opt} for the simple step model computed from 500 runs:

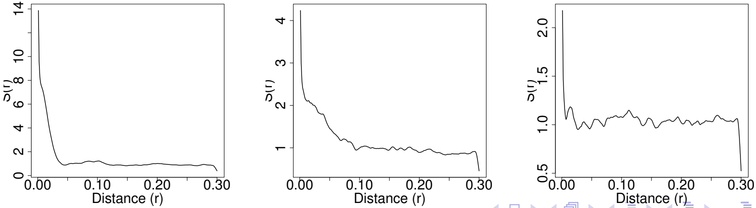
Noise	$10^5 \times \tau_{opt}$	$w_{opt}(0)$	$w_{opt}(1)$	$w_{opt}(2)$	$w_{opt}(3)$
Cst	1940 (930)	0.16 (0.11)	0.23 (0.10)	0.30 (0.09)	0.31 (0.08)
Incr.	360 (200)	0.98 (0.02)	0.02 (0.02)	0.00 (0.00)	0.00 (0.00)
Decr.	85 (38)	0.02 (0.05)	0.03 (0.02)	0.18 (0.05)	0.77 (0.06)

Application 2: modified Thomas process

- Model:
  - Parent points: homogenenous Poisson p.p. with intensity λ
  - Daugther points (the observed points): Poisson number (with mean μ) of points spread around each parent point x with a 2D-, isotropic normal distribution $\mathcal{N}(x, \sigma^2 \mathbf{Id})$



- Functional statistic: empirical pair-correlation function:



ABC tuning

- $I = 10^5$, $J = 10^3$
- Weight function with 21 jumps:

$$w(r) = \begin{cases} w_n & \text{if } r \in \left[\frac{0.3n}{20}, \frac{0.3(n+1)}{20}\right), \ \forall n \in \{0, 1, \ldots, 19\} \\ 0 & \text{if } r < 0 \text{ or } r \geq 0.3, \end{cases}$$

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

- Distance function:

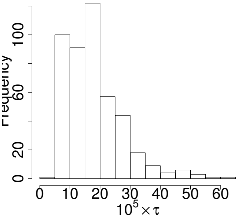
$$\begin{aligned} d(S_i, S_j'; w) &= \int_{\mathbb{R}} w(r)\{S_i(r) - S_j'(r)\}^2 dr \\ &\approx \sum_{k=1}^{249} w(0.3k/250)\{S_i(0.3k/250) - S_j'(0.3k/250)\}^2 \end{aligned}$$

Series of ABC runs

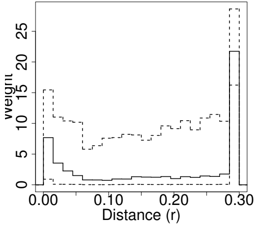
Average BMSE and SD for the modified Thomas process (based on 500 runs) and # of times that each weight function provides the lowest BMSE

	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
$10^5 \times \tau$	18 (7)	35 (13)	18 (9)

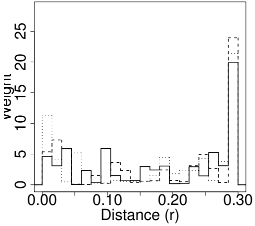
Distrib. of τ_{opt}



Pointwise median of w



3 ex. of w

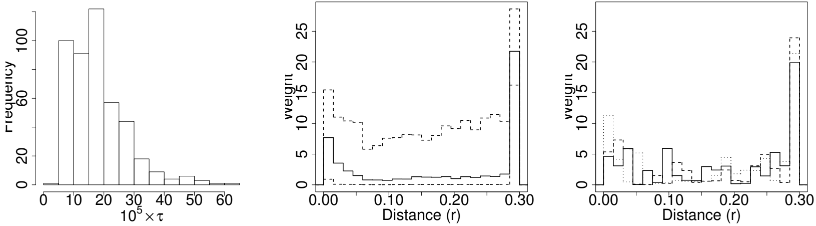


Series of ABC runs

Average BMSE and SD for the modified Thomas process (based on 500 runs) and # of times that each weight function provides the lowest BMSE

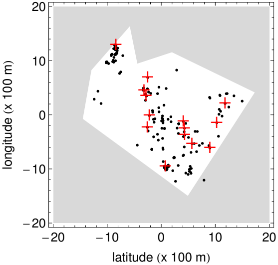
	w_{cst}	w_{var}	w_{opt}
BMSE	0.651 (0.024)	0.942 (0.031)	0.365 (0.025)
Lowest BMSE frequency	0	0	500
$10^5 \times \tau$	18 (7)	35 (13)	18 (9)

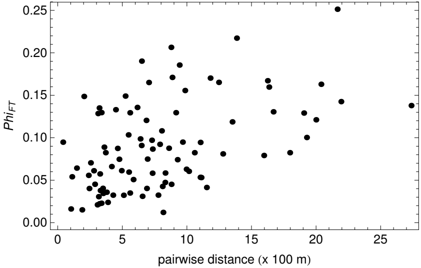
Distrib. of τ_{opt} Pointwise median of w 3 ex. of w



Application 3: dispersal model for pollen

- Data: genotypes of seeds collected from trees at known locations
- Functional statistic: genetic differentiation $\Phi^{obs}_{FT,mm'}$ between the pollen pools of mother trees m and m'
- 14 mother trees $\Rightarrow$ 91 pairs of mothers





- Model: relatively complex model including
  - a parametric dispersal kernel for pollen proportional to:

$$\exp\left\{-\left(\frac{r}{a}\right)^b\right\}$$

ABC tuning

- $I = 10^5$ or $I = 10^6$, $J = 10^3$, $\mid \mathfrak{J} \mid = 250$
- Weight function with 21 jumps:

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

$w_0, \ldots, w_{19} \geq 0$ and $\int_{\mathbb{R}} w(r)dr = \sum_{n=0}^{19}(r_{n+1} - r_n)w_n = 1$

- Distance function:

$$d(S_i, S_j'; w) = \int_{\mathbb{R}} w(r)\{S_i(r) - S_j'(r)\}^2 dr$$
$$\approx \sum_{k=1}^{91} w(\tilde{r}_k)\{S_i(\tilde{r}_k) - S_j'(\tilde{r}_k)\}^2$$

where $\{\tilde{r}_k : k = 1, \ldots, 91\}$ are the 91 inter-mother distances

BMSE and PMSE values for varying simulation number

BMSE and PMSE obtained for the estimation of the pollen dispersal parameters with $I = 10^5$ and $I = 10^6$

$I = 10^5$

	w_{cst}	w_{var}	w_{opt}	p-value
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}

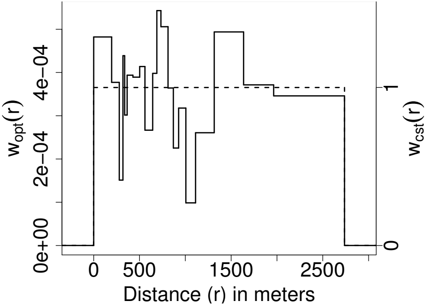
$I = 10^6$

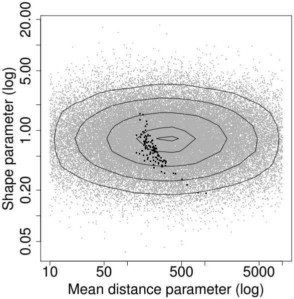
	w_{cst}	w_{var}	w_{opt}	p-value
BMSE	0.977	0.981	0.938	1.1×10^{-4}
PMSE (without pilot ABC)	0.092	0.094	0.089	0.11
PMSE (with pilot ABC)	0.090	0.094	0.083	1.8×10^{-4}

Last column: p-value of the paired t-test comparing the average MSEs obtained with w_{opt} and w_{cst}

Optimal weight function and posterior distributions

Using Algorithm **A3** with pilot ABC and $(I, J, |\mathcal{J}|) = (10^6, 10^3, 250)$:





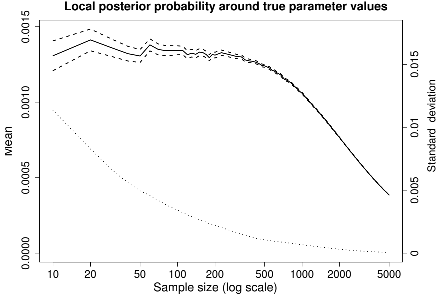
Posterior sample size: 113

Discussion

- Our approach can be applied to non-functional statistics
  - One weight per summary statistic (Application 1)
  - However, being able to sort the summary statistics with a covariate (time, distance...) allows us to reduce the number of weights to be optimized
  - Even if the dependence in the covariate is weak (Application 3)
- Trade-off between optimizing the weights and making more simulations
- Investigation around the size of the posterior sample
  - More simulations $\Rightarrow$ larger size
  - Alternative: replacing the BMSE by a criterion leading to larger sizes
  - However, the BMSE-based optimal size is appropriate for handling the bias-variance trade-off

Illustration of the bias-variance trade-off

- Model: $\mathcal{D}_1, \ldots, \mathcal{D}_{100} \underset{indep.}{\sim} \mathcal{N}\left(\binom{\mu}{\mu}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$
- Summary statistics:
  - average of the first components of the $\mathcal{D}_n$ $(n = 1, \ldots, 100)$
  - number of times that the two components of $\mathcal{D}_n$ have the same signs
- Bias-variance trade-off:



- Application of Algorithm **A3** with $(I, J) = (5 \times 10^4, 10^3)$: posterior sample size $= 585$