Calibration of a pollination model using Approximate Bayesian Computation

joint work with Ullrika Sahlin, Yann Clough and Henrik G. Smith (from Lund University)

Rencontre de la chaire Modélisation Mathématiques et Biodiversité
January 28th, 2021

Charlotte Baey

Université de Lille

Plan

1. Introduction
2. Pollination model
3. Data
4. Statistical model
5. Approximate Bayesian Computation
6. Results
7. Conclusion and discussion

Introduction

Introduction

- In 2000, the United Nations launches the *Millenium Ecosystem Assessment*
- Initial goal: “identify the impacts of ecosystem changes on human well-being, and actions needed to enhance the conservation and sustainable use of those systems”
- It popularized the term ecosystem services
  - the benefits humans obtain from ecosystems
  - e.g. : crop pollination, oxygen production by plants, carbon sequestration, ...
- One of the final recommendation was to assess these ecosystem services and include them in public policy decision making

Introduction

- One possible answer: develop models for ecosystem services, and use these models to evaluate the impact of changes (climate change, agricultural practices, management interventions, ...) on the ecosystems
- These models are often complex, and rarely calibrated on experimental data (rely on expert judgment, literature data, ...)

Introduction

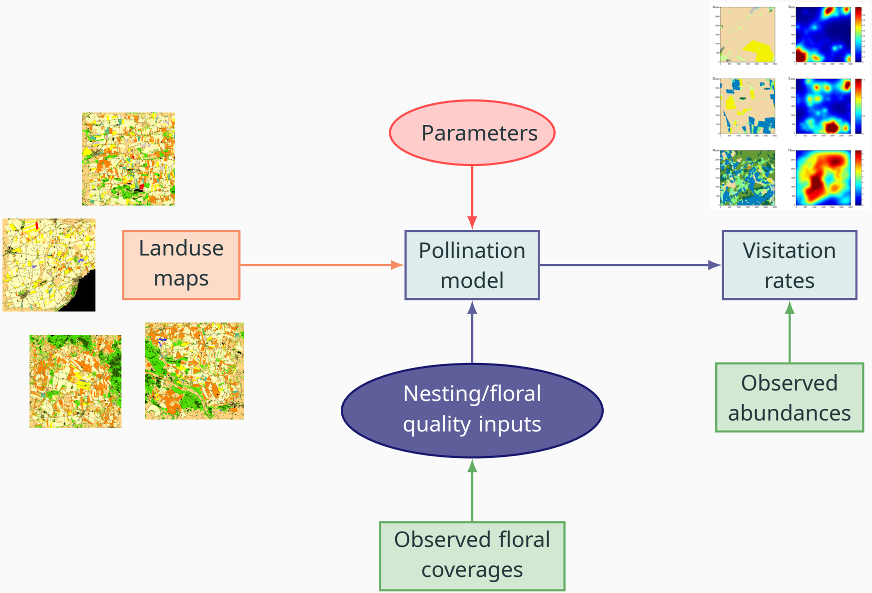
- One possible answer: develop models for ecosystem services, and use these models to evaluate the impact of changes (climate change, agricultural practices, management interventions, ...) on the ecosystems
- These models are often complex, and rarely calibrated on experimental data (rely on expert judgment, literature data, ...)

Objectives

- Build a mechanistic model for pollination at the landscape level
- Propose a methodology to calibrate the model on experimental data

Pollination model

General structure



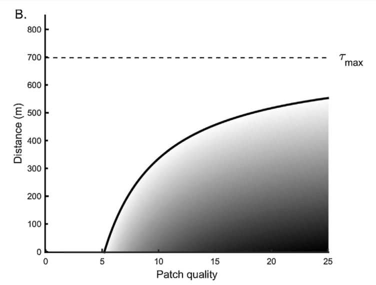
Central-place foragers model (Olsson et Bolin, 2014, Olsson et al. 2015)

Based on a fitness isocline curve

$$\tau_f = \tau_0 \left(1 - \frac{f_0}{f}\right),$$

with:

- f_0 minimum floral value that will be visited by bees
- τ_0 the maximum distance travelled by a bee



(Olsson et al. 2015)

CPF model - suitability of a nest

- For a bee nesting in patch i and floral patch j at distance d_{ij} :

$$\Delta_{ij} = \tau_0 \left(1 - \frac{f_0}{f_j}\right) - d_{ij},$$

$\rightarrow$ distance spare flying for patch j compared to what they were willing to fly for a patch of that quality.

CPF model - suitability of a nest

- For a bee nesting in patch *i* and floral patch *j* at distance d_{ij} :

$$\Delta_{ij} = \tau_0 \left(1 - \frac{f_0}{f_j}\right) - d_{ij},$$

$\rightarrow$ distance spare flying for patch *j* compared to what they were willing to fly for a patch of that quality.

CPF model - suitability of a nest

- For a bee nesting in patch i and floral patch j at distance d_{ij} :

$$\Delta_{ij} = \tau_0 \left(1 - \frac{f_0}{f_j}\right) - d_{ij},$$

$\rightarrow$ distance spare flying for patch j compared to what they were willing to fly for a patch of that quality.

- Suitability or fitness of a nest in patch i:

$$s_i = \sum_j \Delta_{ij} \mathbf{1}_{\Delta_{ij} > 0},$$

$\rightarrow$ distance a bee will spare flying when its nest is surrounded by patches of good quality.

CPF model - optimization of foraging

- A bee in a nest with high suitability exploit less patches further away compared to a bee in a nest with low suitability

CPF model - optimization of foraging

- A bee in a nest with high suitability exploit less patches further away compared to a bee in a nest with low suitability

CPF model - optimization of foraging

- A bee in a nest with high suitability exploit less patches further away compared to a bee in a nest with low suitability
- We define a new maximum distance the bee is prepared to travel from patch i:

$$\tau_i = \frac{\tau_0}{1 + \exp((\sqrt{s_i} - a)/b)}. \tag{1}$$

$\rightarrow$ patch-specific maximum distance.

CPF model - optimization of foraging

- A bee in a nest with high suitability exploit less patches further away compared to a bee in a nest with low suitability
- We define a new maximum distance the bee is prepared to travel from patch i:

$$\tau_i = \frac{\tau_0}{1 + \exp((\sqrt{s_i} - a)/b)}. \tag{1}$$

  $\rightarrow$ patch-specific maximum distance.
- We define:

$$\Delta^*_{ij} = \tau_i \left(1 - \frac{f_0}{f_j}\right) - d_{ij}, \tag{2}$$

  $\rightarrow$ contribution from patch j to fitness of the bees in patch i.

- The number of foraging bees from nest in patch *i* to floral resources in patch *j* is

$$r_{i \to j} = q_i \frac{\Delta_{ij}^*}{\sum_{j=1}^{J} \Delta_{ij}^*}.$$

where q_i is habitat quality (usually, the number of bees nesting there).

CPF model - visitation rate

- The number of foraging bees from nest in patch i to floral resources in patch j is

$$r_{i \to j} = q_i \frac{\Delta^*_{ij}}{\sum_{j=1}^{J} \Delta^*_{ij}}.$$

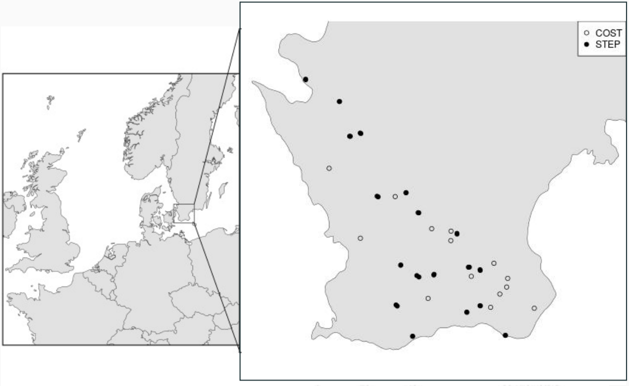
  where q_i is habitat quality (usually, the number of bees nesting there).

- The (instantaneous) total number of bees visiting patch i is:

$$\nu_i = \sum_{j=1}^{J} r_{j \to i},$$

Data

Data

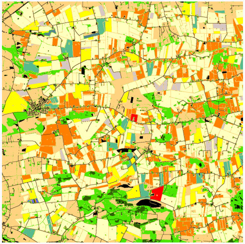


- Two studies on pollinator abundances in southern Sweden
- Data collected in four different years, several times a year (covering 3 different periods of bumblebees life cycle)
- Number of bees flying or foraging in a given transect for a given period of time was recorded

Model inputs

For each sampling site *i*, each year *j* and each period *k*:

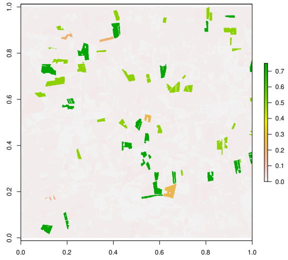
A landscape map



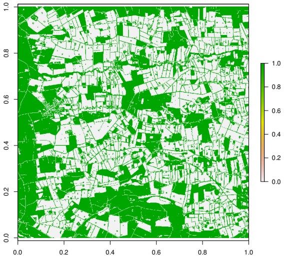
$\Downarrow$

denoted by $\mathcal{M}_{ijk}$

A floral quality map



A nesting map



informed by expert judgement or literature data

Statistical model

Statistical model - Bayesian formulation

- y_{ijk}: observed nb of bees on site i, year j and period k.

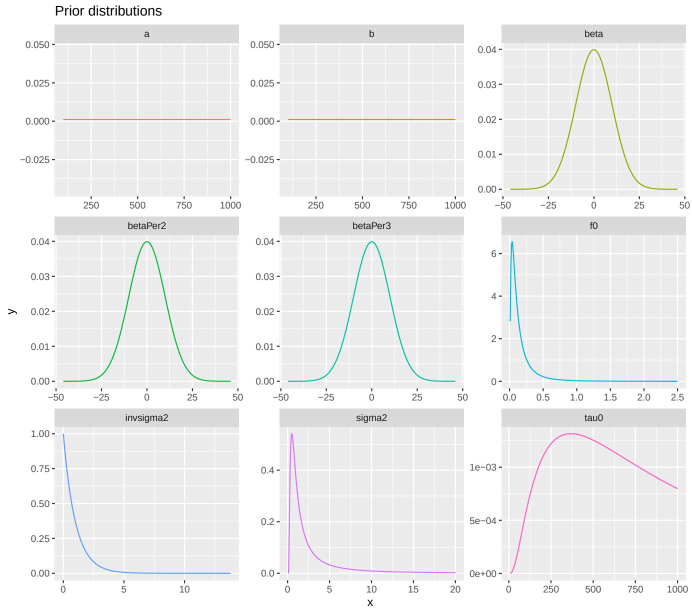
Statistical model - Bayesian formulation

- y_{ijk}: observed nb of bees on site i, year j and period k.
- **Likelihood**
$$\begin{cases} y_{ijk} \mid \lambda_{ijk}, \theta & \sim \mathcal{P}(c_i \cdot \lambda_{ijk}) \\ \log \lambda_{ijk} & = \log \nu_i(\theta, \mathcal{M}_{jk}) + \beta_k + \varepsilon_{ijk} \\ \varepsilon_{ijk} & \sim \mathcal{N}(0, \sigma^2). \end{cases}$$
  - c_i a known scaling parameter,
  - λ_{ijk} the real intensity of the visitation rates,
  - β_k a period-specific parameter
- Complete vector of parameters $\psi = (\tau_0, f_0, a, b, \beta_1, \ldots, \beta_K, \sigma^2)$

Statistical model - Bayesian formulation

- y_{ijk}: observed nb of bees on site i, year j and period k.
- **Likelihood**
$$\begin{cases} y_{ijk} \mid \lambda_{ijk}, \theta & \sim \mathcal{P}(c_i \cdot \lambda_{ijk}) \\ \log \lambda_{ijk} & = \log \nu_i(\theta, \mathcal{M}_{jk}) + \beta_k + \varepsilon_{ijk} \\ \varepsilon_{ijk} & \sim \mathcal{N}(0, \sigma^2). \end{cases}$$
  - c_i a known scaling parameter,
  - λ_{ijk} the real intensity of the visitation rates,
  - β_k a period-specific parameter
- Complete vector of parameters $\psi = (\tau_0, f_0, a, b, \beta_1, \ldots, \beta_K, \sigma^2)$
- **Priors**
$$\tau_0 \sim \mathcal{LN}_{[0,1000]}(\log(1000), 1) \quad f_0 \sim \mathcal{LN}(\log(0.1), 1)$$
$$a \sim \mathcal{U}([100, 1000]) \quad b \sim \mathcal{U}([100, 1000])$$
$$\beta_k \sim \mathcal{N}(\mu_k, \sigma_k^2), \quad k = 1, \ldots, K$$
$$\sigma^2 \sim \mathcal{IG}(\xi, \eta)$$

Statistical model - Bayesian formulation



Approximate Bayesian Computation

Bayesian estimation

- In a Bayesian context, we are now interested in the **posterior** distribution of the parameters, i.e. in $\pi(\psi \mid y)$

Bayesian estimation

- In a Bayesian context, we are now interested in the **posterior** distribution of the parameters, i.e. in $\pi(\psi \mid y)$
- We have :

$$\pi(\psi \mid y) \propto \underbrace{f(y \mid \psi)}_{\text{likelihood}} \underbrace{p(\psi)}_{\text{prior}}$$

Bayesian estimation

- In a Bayesian context, we are now interested in the **posterior** distribution of the parameters, i.e. in $\pi(\psi \mid y)$
- We have :

$$\pi(\psi \mid y) \propto \underbrace{f(y \mid \psi)}_{\text{likelihood}} \; \underbrace{p(\psi)}_{\text{prior}}$$

- But here the likelihood is intractable:

Bayesian estimation

- In a Bayesian context, we are now interested in the **posterior** distribution of the parameters, i.e. in $\pi(\psi \mid y)$
- We have :

$$\pi(\psi \mid y) \propto \underbrace{f(y \mid \psi)}_{\text{likelihood}} \underbrace{p(\psi)}_{\text{prior}}$$

- But here the likelihood is intractable:

Bayesian estimation

- In a Bayesian context, we are now interested in the **posterior** distribution of the parameters, i.e. in $\pi(\psi \mid y)$
- We have :

$$\pi(\psi \mid y) \propto \underbrace{f(y \mid \psi)}_{\text{likelihood}} \underbrace{p(\psi)}_{\text{prior}}$$

- But here the likelihood is intractable:

$$\begin{aligned} f(y \mid \psi) &= \int f(y, \lambda \mid \psi) d\lambda = \int f(y \mid \lambda, \psi) f(\lambda \mid \psi) d\lambda \\ &= \prod_{ijk} \frac{1}{\sqrt{2\pi}\sigma y_{ijk}!} \int_0^{+\infty} e^{-\lambda} \lambda^{y_{ijk}-1} \exp\left(- \frac{(\log \lambda - \log \nu_i(\theta, \mathcal{M}_{ijk}) - \beta_k)^2}{2\sigma^2} \right) d\lambda \end{aligned}$$

Bayesian estimation

- In a Bayesian context, we are now interested in the **posterior** distribution of the parameters, i.e. in $\pi(\psi \mid y)$
- We have :

$$\pi(\psi \mid y) \propto \underbrace{f(y \mid \psi)}_{\text{likelihood}} \; \underbrace{p(\psi)}_{\text{prior}}$$

- But here the likelihood is intractable:

$$f(y \mid \psi) = \int f(y, \lambda \mid \psi) d\lambda = \int f(y \mid \lambda, \psi) f(\lambda \mid \psi) d\lambda$$

$$= \prod_{ijk} \frac{1}{\sqrt{2\pi}\sigma y_{ijk}!} \int_0^{+\infty} e^{-\lambda} \lambda^{y_{ijk}-1} \exp\left(- \frac{(\log \lambda - \log \nu_i(\theta, \mathcal{M}_{ijk}) - \beta_k)^2}{2\sigma^2} \right) d\lambda$$

- how to handle this situation?

Bayesian estimation

- In a Bayesian context, we are now interested in the **posterior** distribution of the parameters, i.e. in $\pi(\psi \mid y)$
- We have :

$$\pi(\psi \mid y) \propto \underbrace{f(y \mid \psi)}_{\text{likelihood}} \underbrace{p(\psi)}_{\text{prior}}$$

- But here the likelihood is intractable:

$$\begin{aligned} f(y \mid \psi) &= \int f(y, \lambda \mid \psi) d\lambda = \int f(y \mid \lambda, \psi) f(\lambda \mid \psi) d\lambda \\ &= \prod_{ijk} \frac{1}{\sqrt{2\pi}\sigma y_{ijk}!} \int_0^{+\infty} e^{-\lambda} \lambda^{y_{ijk}-1} \exp\left(-\frac{(\log\lambda - \log\nu_i(\theta, \mathcal{M}_{ijk}) - \beta_k)^2}{2\sigma^2} \right) d\lambda \end{aligned}$$

- how to handle this situation?
  - change the likelihood so that it becomes tractable? $\rightarrow$ can introduce biais, do not reflect the "true" process as we think it is generated

Bayesian estimation

- In a Bayesian context, we are now interested in the **posterior** distribution of the parameters, i.e. in $\pi(\psi \mid y)$
- We have :

$$\pi(\psi \mid y) \propto \underbrace{f(y \mid \psi)}_{\text{likelihood}} \underbrace{p(\psi)}_{\text{prior}}$$

- But here the likelihood is intractable:

$$f(y \mid \psi) = \int f(y, \lambda \mid \psi) d\lambda = \int f(y \mid \lambda, \psi) f(\lambda \mid \psi) d\lambda$$

$$= \prod_{ijk} \frac{1}{\sqrt{2\pi}\sigma y_{ijk}!} \int_0^{+\infty} e^{-\lambda} \lambda^{y_{ijk}-1} \exp\left(-\frac{(\log \lambda - \log \nu_i(\theta, \mathcal{M}_{ijk}) - \beta_k)^2}{2\sigma^2} \right) d\lambda$$

- how to handle this situation?
  - change the likelihood so that it becomes tractable? $\rightarrow$ can introduce biais, do not reflect the “true” process as we think it is generated
  - use an estimation method which can deal with untractable likelihood $\rightarrow$ approximate Bayesian computation (ABC)

Approximation Bayesian computation (ABC)

- Introduced at the end of the 1990 in the area of population genetics

Approximation Bayesian computation (ABC)

- Introduced at the end of the 1990 in the area of population genetics
- Initial ABC algorithm: “rejection sampling” (Tavaré et al. 1997)

Approximation Bayesian computation (ABC)

- Introduced at the end of the 1990 in the area of population genetics
- Initial ABC algorithm: “rejection sampling” (Tavaré et al. 1997)

Approximation Bayesian computation (ABC)

- Introduced at the end of the 1990 in the area of population genetics
- Initial ABC algorithm: “rejection sampling” (Tavaré et al. 1997)

ABC rejection sampling

Input: a threshold ε and a distance d on the set of observations

For $m = 1, \dots, M$:

1. draw a sample $\psi^{(m)}$ from the prior distribution
2. generate a set of observations $y^{(m)}$ using $p(y \mid \psi)$
3. if $d(y_{obs}, y^{(m)}) \leq \varepsilon$, keep $\psi^{(m)}$
4. **Output**: a sample of size M_ε with all the accepted sets of parameters $\psi^{(m)}$

Approximation Bayesian computation (ABC)

- Introduced at the end of the 1990 in the area of population genetics
- Initial ABC algorithm: "rejection sampling" (Tavaré et al. 1997)

ABC rejection sampling

Input: a threshold ε and a distance d on the set of observations

For $m = 1, \ldots, M$:

1. draw a sample $\psi^{(m)}$ from the prior distribution
2. generate a set of observations $y^{(m)}$ using $p(y \mid \psi)$
3. if $d(y_{obs}, y^{(m)}) \leq \varepsilon$, keep $\psi^{(m)}$
4. **Output**: a sample of size M_ε with all the accepted sets of parameters $\psi^{(m)}$

- The accepted values follow the ABC posterior distribution $\pi_\varepsilon(\psi \mid y)$:

$$\pi_{ABC}(\psi \mid y_{obs}) \propto \int f(y|\psi)p(\psi)\mathbb{1}_{A_\varepsilon}(y)dy,$$

where $A_\varepsilon = \{y; d(y_{obs}, y) < \varepsilon\}$.

Approximate Bayesian computation (ABC)

- When $\varepsilon \to 0$, the ABC posterior converges to the true posterior distribution

Approximate Bayesian computation (ABC)

- When $\varepsilon \to 0$, the ABC posterior converges to the true posterior distribution
- On the contrary, when $\varepsilon \to \infty$, the ABC posterior converges to the prior distribution

Approximate Bayesian computation (ABC)

- When $\varepsilon \to 0$, the ABC posterior converges to the true posterior distribution
- On the contrary, when $\varepsilon \to \infty$, the ABC posterior converges to the prior distribution
- In the meantime, when the dimension increases, $d(y_{obs}, y^{(m)})$ tends to be very large $\to$ curse of dimensionality

Approximate Bayesian computation (ABC)

- When $\varepsilon \to 0$, the ABC posterior converges to the true posterior distribution
- On the contrary, when $\varepsilon \to \infty$, the ABC posterior converges to the prior distribution
- In the meantime, when the dimension increases, $d(y_{obs}, y^{(m)})$ tends to be very large $\to$ curse of dimensionality
- Several extensions to the original algorithm have been proposed:

Approximate Bayesian computation (ABC)

- When $\varepsilon \to 0$, the ABC posterior converges to the true posterior distribution
- On the contrary, when $\varepsilon \to \infty$, the ABC posterior converges to the prior distribution
- In the meantime, when the dimension increases, $d(y_{obs}, y^{(m)})$ tends to be very large $\to$ curse of dimensionality
- Several extensions to the original algorithm have been proposed:
  - introduction of summary statistics $s(\cdot)$ of dimension $q < n \to$ samples from $\pi(\psi \mid s_{obs})$ instead of the posterior $\pi(\psi \mid y_{obs})$

Approximate Bayesian computation (ABC)

- When $\varepsilon \to 0$, the ABC posterior converges to the true posterior distribution
- On the contrary, when $\varepsilon \to \infty$, the ABC posterior converges to the prior distribution
- In the meantime, when the dimension increases, $d(y_{obs}, y^{(m)})$ tends to be very large $\to$ curse of dimensionality
- Several extensions to the original algorithm have been proposed:
  - introduction of summary statistics $s(\cdot)$ of dimension $q < n \to$ samples from $\pi(\psi \mid s_{obs})$ instead of the posterior $\pi(\psi \mid y_{obs})$
  - replace crude rejection by kernel smoothing $\to$ each sample is used, with a weight $w_m = K(d(y_{obs}, y^{(m)}))$

ABC - choice of summary statistics

- Choice of the summary statistics is crucial: a poor specification leads to a poor approximation of the posterior, but choosing **sufficient statistics** has no impact, i.e. $\pi(\psi \mid s_{obs}) = \pi(\psi \mid y_{obs})$.

Objective

find low-dimensional summary and highly informative summary statistics

ABC - choice of summary statistics

- Choice of the summary statistics is crucial: a poor specification leads to a poor approximation of the posterior, but choosing **sufficient statistics** has no impact, i.e. $\pi(\psi \mid s_{obs}) = \pi(\psi \mid y_{obs})$.
- A statistic is said to be sufficient w.r.t. a parameter θ if it carries all the necessary knowledge to perform inference for θ. More precisely, s is sufficient iff:

$$\mathbb{P}(y \in A \mid s(y) \in B, \theta) = \mathbb{P}(y \in A \mid s(y) \in B)$$

  ex. : for a Gaussian i.i.d. sample $\mathcal{N}(\mu, \sigma^2)$, the sample mean is sufficient for μ.

Objective

find low-dimensional summary and highly informative summary statistics

Methods based on regression adjustment

- **First idea:** build a relationship between the parameter values and the summary statistics values, e.g. via regression techniques.

$$\psi_i^{(m)} = m_i(s(y^{(m)}) - s_{obs}) + \sigma(s(y^{(m)}))\varepsilon_{im}, \quad i = 1, \dots, p$$

  Then, samples from $\pi_{ABC}(\psi \mid s_{obs})$ are obtained via:

$$\psi_i^{*(m)} = \hat{m}_i(s_{obs}) + \left(\psi_i^{(m)} - \hat{m}(s(y^{(m)}))\right) \frac{\hat{\sigma}(s_{obs})}{\hat{\sigma}(s(y^{(m)}))}$$

- **First idea:** build a relationship between the parameter values and the summary statistics values, e.g. via regression techniques.

$$\psi_i^{(m)} = m_i(s(y^{(m)}) - s_{obs}) + \sigma(s(y^{(m)}))\varepsilon_{im}, \quad i = 1, \dots, p$$

  Then, samples from $\pi_{ABC}(\psi \mid s_{obs})$ are obtained via:

$$\psi_i^{*(m)} = \hat{m}_i(s_{obs}) + \left(\psi_i^{(m)} - \hat{m}(s(y^{(m)}))\right) \frac{\hat{\sigma}(s_{obs})}{\hat{\sigma}(s(y^{(m)}))}$$

- Estimation via weighted least squares, where weights are given by a chosen smoothing kernel K

Methods based on regression adjustment

- **First idea:** build a relationship between the parameter values and the summary statistics values, e.g. via regression techniques.

$$\psi_i^{(m)} = m_i(s(y^{(m)}) - s_{obs}) + \sigma(s(y^{(m)}))\varepsilon_{im}, \quad i = 1, \dots, p$$

  Then, samples from $\pi_{ABC}(\psi \mid s_{obs})$ are obtained via:

$$\psi_i^{*(m)} = \hat{m}_i(s_{obs}) + \left(\psi_i^{(m)} - \hat{m}(s(y^{(m)}))\right) \frac{\hat{\sigma}(s_{obs})}{\hat{\sigma}(s(y^{(m)}))}$$

- Estimation via weighted least squares, where weights are given by a chosen smoothing kernel K
- Local regression around $s_{obs} \rightarrow$ lowers the effect of the distance between $s(y^{(m)})$ and s_{obs}.

Methods based on regression adjustment

- **First idea:** build a relationship between the parameter values and the summary statistics values, e.g. via regression techniques.

$$\psi_i^{(m)} = m_i(s(y^{(m)}) - s_{obs}) + \sigma(s(y^{(m)}))\varepsilon_{im}, \quad i = 1, \ldots, p$$

  Then, samples from $\pi_{ABC}(\psi \mid s_{obs})$ are obtained via:

$$\psi_i^{*(m)} = \hat{m}_i(s_{obs}) + \left(\psi_i^{(m)} - \hat{m}(s(y^{(m)}))\right) \frac{\hat{\sigma}(s_{obs})}{\hat{\sigma}(s(y^{(m)}))}$$

- Estimation via weighted least squares, where weights are given by a chosen smoothing kernel K
- Local regression around $s_{obs} \to$ lowers the effect of the distance between $s(y^{(m)})$ and s_{obs}.
- Can be combined with dimension reduction methods to further reduce the summary statistics dimension

Methods based on regression adjustment

Some methods based on regression adjustment (shortnames for methods compared in this work are marked in blue):

Regression adjustment methods

- local linear heteroscedastic model (Beaumont et al. 2002) [LocLH]

Methods based on regression adjustment

Some methods based on regression adjustment (shortnames for methods compared in this work are marked in blue):

Regression adjustment methods

- local linear heteroscedastic model (Beaumont et al. 2002) [LocLH]
- local nonlinear heteroscedastic model (Blum and François 2010) [LocNLH]

Methods based on regression adjustment

Some methods based on regression adjustment (shortnames for methods compared in this work are marked in blue):

Regression adjustment methods

- local linear heteroscedastic model (Beaumont et al. 2002) [LocLH]
- local nonlinear heteroscedastic model (Blum and François 2010) [LocNLH]
- adaptive nonlinear heteroscedastic model (Blum and François 2010) [ANLH]
  $\rightarrow$ two-step procedure:

Methods based on regression adjustment

Some methods based on regression adjustment (shortnames for methods compared in this work are marked in blue):

Regression adjustment methods

- local linear heteroscedastic model (Beaumont et al. 2002) [LocLH]
- local nonlinear heteroscedastic model (Blum and François 2010) [LocNLH]
- adaptive nonlinear heteroscedastic model (Blum and François 2010) [ANLH]
  $\rightarrow$ two-step procedure:
  1. perform a LocNLH regression and estimate the distribution support D of the adjusted values

Methods based on regression adjustment

Some methods based on regression adjustment (shortnames for methods compared in this work are marked in blue):

Regression adjustment methods

- local linear heteroscedastic model (Beaumont et al. 2002) [LocLH]
- local nonlinear heteroscedastic model (Blum and François 2010) [LocNLH]
- adaptive nonlinear heteroscedastic model (Blum and François 2010) [ANLH]
  $\rightarrow$ two-step procedure:
  1. perform a LocNLH regression and estimate the distribution support D of the adjusted values
  2. perform a second LocNLH regression using parameters values samples from p_D, the conditional prior of the parameters given that they fall in D

Methods based on regression adjustment

Some methods based on regression adjustment (shortnames for methods compared in this work are marked in blue):

Regression adjustment methods

- local linear heteroscedastic model (Beaumont et al. 2002) [LocLH]
- local nonlinear heteroscedastic model (Blum and François 2010) [LocNLH]
- adaptive nonlinear heteroscedastic model (Blum and François 2010) [ANLH]
  $\rightarrow$ two-step procedure:
  1. perform a LocNLH regression and estimate the distribution support D of the adjusted values
  2. perform a second LocNLH regression using parameters values samples from p_D, the conditional prior of the parameters given that they fall in D
- penalized regression (Wegmann et al. 2016)

Methods based on regression adjustment

Some methods based on regression adjustment (shortnames for methods compared in this work are marked in blue):

Regression adjustment methods

- local linear heteroscedastic model (Beaumont et al. 2002) [LocLH]
- local nonlinear heteroscedastic model (Blum and François 2010) [LocNLH]
- adaptive nonlinear heteroscedastic model (Blum and François 2010) [ANLH]
  $\rightarrow$ two-step procedure:
  1. perform a LocNLH regression and estimate the distribution support D of the adjusted values
  2. perform a second LocNLH regression using parameters values samples from p_D, the conditional prior of the parameters given that they fall in D
- penalized regression (Wegmann et al. 2016)
- best subset selection (via criteria such as AIC or BIC for example) (Numes and Balding 2010, Blum et al. 2013)

Methods based on quantile regression

- As outputs of the previous approaches, we get a sample of the ABC posterior

Methods based on quantile regression

- As outputs of the previous approaches, we get a sample of the ABC posterior
- However, we are sometimes only interested in some quantities from this posterior distribution (e.g. quantiles, mean, ...)

Methods based on quantile regression

- As outputs of the previous approaches, we get a sample of the ABC posterior
- However, we are sometimes only interested in some quantities from this posterior distribution (e.g. quantiles, mean, ...)

$\rightarrow$ what if we try to approximate these quantities using ABC instead of the whole posterior ?

- As outputs of the previous approaches, we get a sample of the ABC posterior
- However, we are sometimes only interested in some quantities from this posterior distribution (e.g. quantiles, mean, ...)

$\rightarrow$ what if we try to approximate these quantities using ABC instead of the whole posterior ?

Methods based on quantile regression

- As outputs of the previous approaches, we get a sample of the ABC posterior
- However, we are sometimes only interested in some quantities from this posterior distribution (e.g. quantiles, mean, ...)

$\rightarrow$ what if we try to approximate these quantities using ABC instead of the whole posterior ?

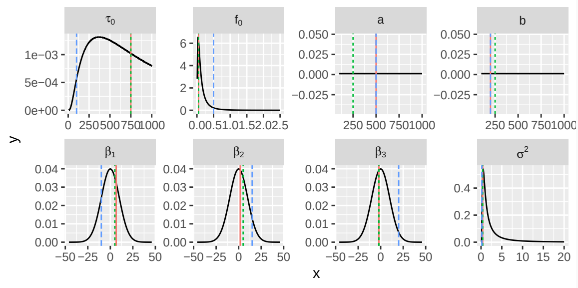
Some examples

- Raynal et al. 2016 suggested the use of random forests combined with quantile regression [qRF]
- Other machine learning algorithms could be used, e.g. gradient boosting methods, also combined with quantile regression [qGBM]

Results

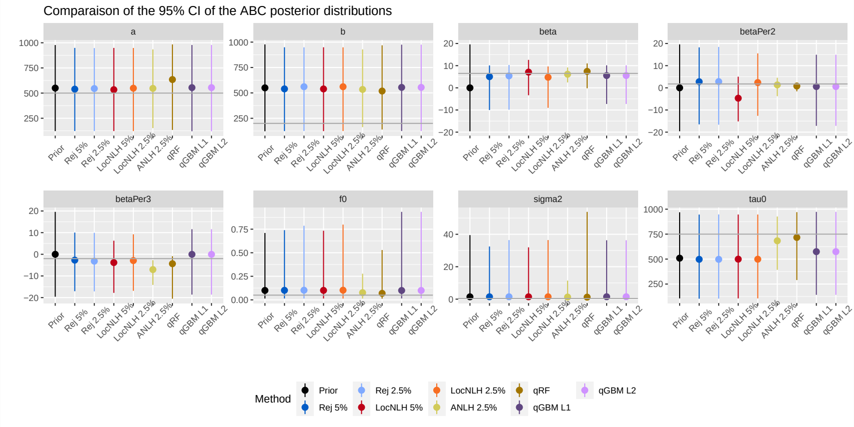
Simulated data

- We simulated three datasets corresponding to different parameter values

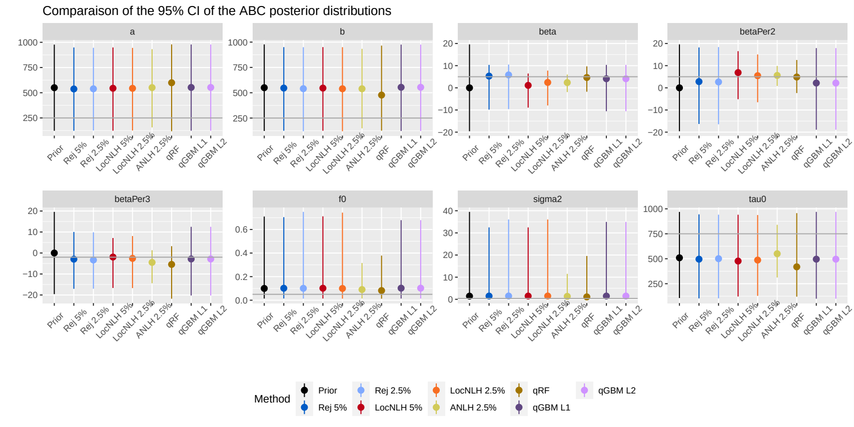


- We generated $M = 100\,000$ parameter samples from the prior and M corresponding simulated datasets
- Summary statistics were defined as : min, max, mean, quantiles of order 25%, 50% (median), 75%, and number of 0 observed per landuse type, and per combination of landuse type and flowering period
  $\rightarrow$ from 790 observations to 112 summary statistics

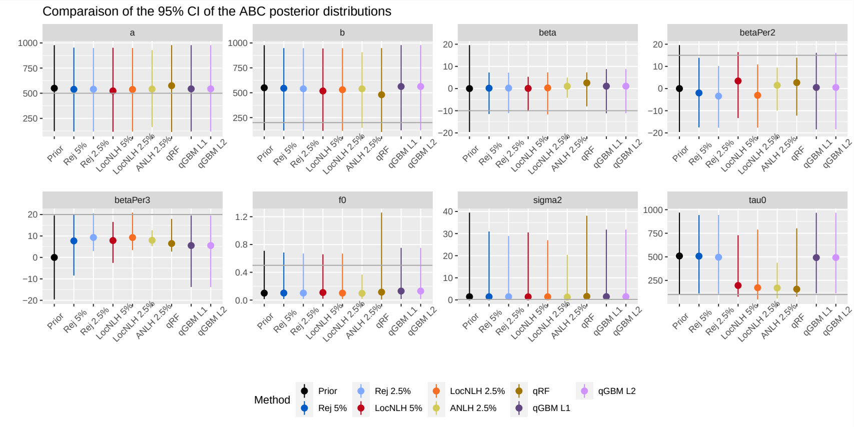
Results (no local linear approaches)



Results (no local linear approaches)



Results (no local linear approaches)

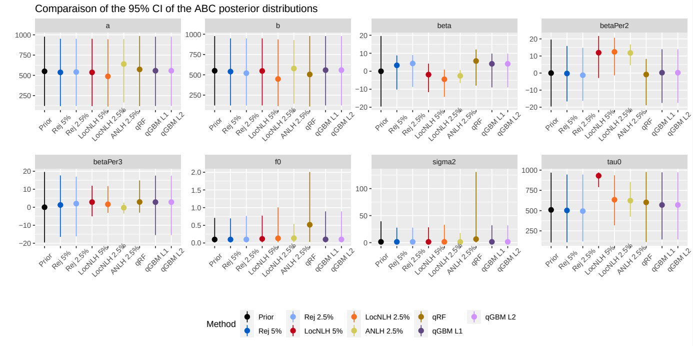


Results – best results

Posterior mean - posterior median for CPF model parameters

Param	True	LocNLH 5%	LocNLH 2.5%	ANLH 5%	qRF
τ_0	750	512 / 499	510 / 498	679 / 684	**691 / 717**
	750	501 / 476	512 / 486	**556 / 550**	454 / 419
	100	249 / 196	230 / 173	**192** / 170	214 / **158**
f_0	0.05	0.167 / 0.101	0.174 / 0.103	**0.095** / 0.076	0.102 / **0.069**
	0.05	0.166 / 0.102	0.164 / 0.100	**0.111** / 0.090	**0.111 / 0.082**
	0.05	0.172 / 0.109	0.167 / 0.101	**0.124 / 0.097**	0.212 / 0.114
a	500	**532 / 534**	539 / 548	545 / 547	592 / 635
	250	**537 / 544**	538 / 543	549 / 551	588 / 599
	500	**526 / 522**	534 / 537	545 / 540	559 / 575
b	200	538 / 537	548 / 559	541 / 534	**526 / 518**
	250	542 / 545	538 / 538	542 / 539	**498 / 477**
	200	519 / 517	526 / 529	533 / 539	**510 / 480**

Real data



Param	MAP	95% CI
τ_0	623	[425 ; 855]
f_0	0.137	[0.021 ; 0.540]
a	640	[166 ; 948]
b	578	[144 ; 928]

Param	MAP	95% CI
β_1	-2.54	[-6.53 ; 0.624]
β_2	11.8	[4.54 ; 16.8]
β_3	-0.270	[-3.54 ; 2.56]
σ^2	1.40	[0.253 ; 17.8]

Conclusion and discussion

Conclusion

- Preliminary results suggest that some parameters are difficult to infer
- Methods based on nonlinear local regression perform better
- Improvements are needed to enhance predictive quality
- Results are conditional on the floral and nesting maps

Perspectives

- Evaluate the performances of the methods on more simulated data (computation of RMSE)
- Another choice for the summary statistics?
- Influence of the origin of the data (STEP study vs COST study)?
- Use the estimated ABC posterior distribution to tune likelihood-free MCMC algorithms (initialization of the chain, choice of the proposal distribution) (e.g. Wegmann 2009)
- Evaluate the influence of the input maps