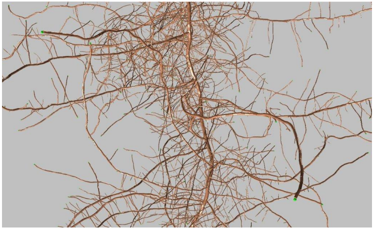
# Usefulness of sensitivity analysis for approximate bayesian computation

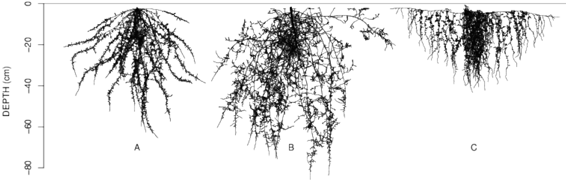
Olivier Martin, Claude Bruchou, Loic L. Pagès

# Usefullness of Sensitivity Analysis for Approximate Bayesian Computation

**O. Martin*, C. Bruchou*, L. Pagès◇**

* INRA, Avignon, Biosp ◇ INRA, Avignon, PSH






*Work supported by ANR Sim-Traces and ANR Emile*

# Overview

1. Review of ABC concepts
2. The root system model
3. Sensitivity Analysis for statistics
4. Sensitivity Analysis for MSE criterion
5. Conclusion and discussion

# 1. ABC concepts

- Approximate Bayesian Computing (ABC) is a free likelihood method to estimate model parameters
- Definition of statistics (or descriptors)
- Fast computing model

Notations:

Observed data $D$ and simulated data $D^\star$

$\theta$ is the vector of parameters with Prior $\pi(.)$

s(.): function that computes a set of statistics (descriptors)

$S = s(D)$ vector of statistics for data $D$

$S^\star = s(D^\star)$ vector of statistics for data $D^\star$

# 1. ABC: a free likelihood method

## Algorithm (Accept/Reject)

0: Suppose we have observed data $D$ and $S = s(D)$

1: Generate $\theta^\star$ from $\pi(.)$

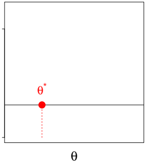
2: Generate $D^\star$ from $f(.|\theta^\star)$

3: Compute statistics $S^\star$ for $D^\star$

4: Accept $\theta^\star$ if $d_W(S, S^\star) \leq \epsilon$ and return to (1)

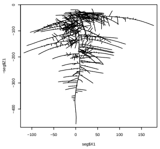
Prior

$\pi(\theta)$



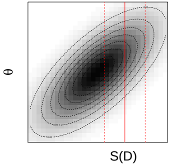

Simulation

$D^\star$



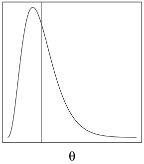
Joint Ditribution

$(\theta^\star, S^\star)$




Posterior

$\pi(\theta|d_W(S, S^\star) \leq \epsilon)$



## 1. ABC: a free likelihood method

This algorithm gives an approximation of $\pi(\theta|D)$.

Two important points for the approximation:

- The threshold $\epsilon$:

  smaller $\epsilon \rightarrow$ *better* approximation

- $D^\star$ is summarised by the statistics $S^\star$:

  *better* statistics $\rightarrow$ *better* approximation

# 2. The root system model

## Complexity of plant root system:

Functionning is linked to the dynamics of the architecture. Water and nutriment uptake depend on the root surface..

## Plant root system modelling:

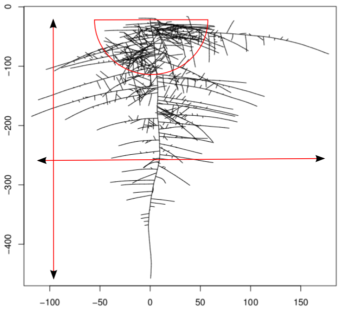
Integration of knowledge and test of new hypotheses
Summarize data into a low number of key values

## The stochastic model:

Number of parameters: 14
Output of the model: image of root system

# 3. The root system model

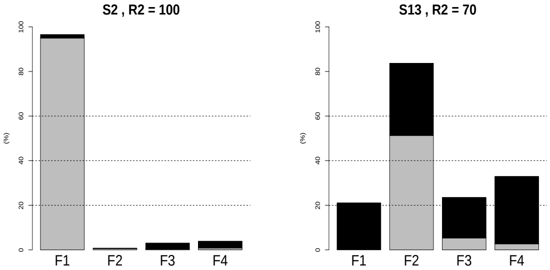
- **4 parameters** over 14 are estimated with images.
- **15 statistics** are computed: size and shape of the root system, density of pixels in different areas, ...



# 3. Sensitivity analysis of statistics

- Can parameters be estimated with the statistics ?
- Anova: 4 factors with 5 levels, interaction of order 3

Gray: Principal
Black: Interaction




- About 8-10 statistics over the 15 seem to be sufficient to estimate parameters

## 4. Sensitivity analysis of MSE

- Find the best weights $W$ of $d_W$ to minimize MSE criterion ?
- Point estimate: $\hat{\theta} = Mean\{\theta^\star : d_W(S, S^\star) \leq \epsilon\}$ with

$$d_W^2(S, S^\star) = \sum_{i=1}^{N_S=15} w_i (S_i - S_i^\star)^2 \text{ and } w_i > 0, \sum_{i=1}^{N_S} w_i = 1.$$

- Criterion to evaluate point estimate $\hat{\theta}$:

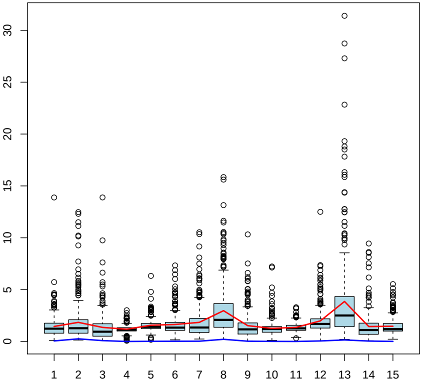
$$MSE_\theta(W) = \sum_{k=1}^{N_\theta=4} \frac{(\hat{\theta}^{(k)} - \theta^{(k)})^2}{\sigma^2_{\theta^{(k)}}}$$

# 4. Sensitivity analysis of MSE

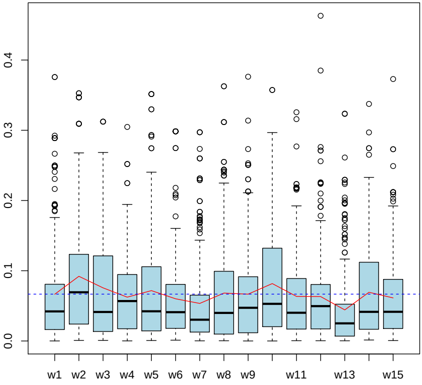
- Generate uniformly a $R$-sample of weights $W^r, r = 1, ..., R$ with $W^r = (w_1^r, ..., w_{N_S}^r)$ and $\sum_{i=1}^{N_S} w_i^r = 1$
- Generate a $N$-sample $\theta_l, l = 1, ..., N$ from $\pi(\theta)$.
- For each $\theta_l, l = 1, ..., N$
  - Compute $MSE_{\theta_l}(W^r)$, r=1,...,R
  - Fit a canonical polynomial of degree 2:
    $MSE_{\theta_l}(W) = P_l(W) + e$, $l = 1, ..., N$
    $$\text{with } P_l(W) = \sum_{i=1}^{N_S} \delta_{ii} w_i^2 + \sum_{i=1}^{N_S} \sum_{i<j}^{N_S} \delta_{ij} w_i w_j$$
  - Sensitivity indices by comparing nested polynomials models.

# 4. Sensitivity analysis of MSE

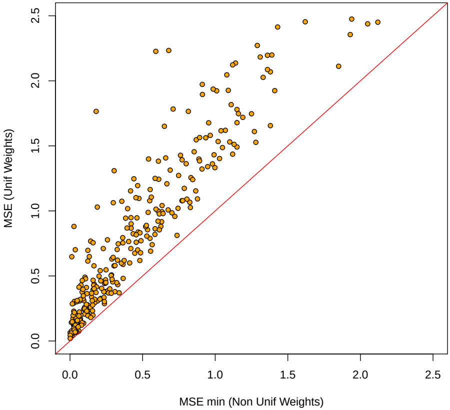
Sensitivity indices

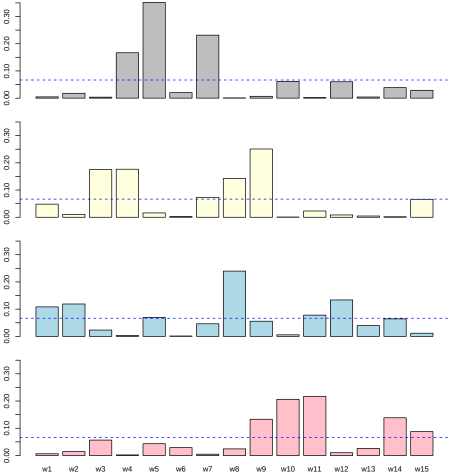



Minimum weights



# 4. Sensitivity analysis of MSE



# 5. Conclusion

## Conclusion

- Difficult to find an optimal distance (for all $\theta$)
- Interaction between weights associated to statistics
- ABC with three steps:
  1. Pilot ABC ($\rightarrow$ first approximation $\tilde{\theta}$)
  2. Determine optimal weights associated to $\tilde{\theta}$
  3. ABC with the optimal weights ($\rightarrow$ second approximation $\hat{\theta}$)

## Future work

- Optimal weights determined by global optimum of $P_W$
- Study based on the expectations of the statistics (rather one observation)

# References

Beaumont, M., Zhang, W., Balding, D.J. (2002). Approximate bayesian computation in population genetics. Genetics.

Cornell, J. (2002). Experiments with mixtures. J. Wiley and sons, N. Y., 3rd edition.

Joyce, P., Marjoram, P. (2008). Approximately sufficient statistics and bayesian computation. Stat. Appl. Genet. Mol. Biol.

Pagès, L. (2011). Links between root developmental traits and foraging performance. Plant, Cell and Environment.

# Factor F3 and F4: