Inverse calibration in a Bayesian framework: ABC methods and specific challenges for computationally demanding models.

Franck Jabot
LISC - Irstea Clermont-Ferrand

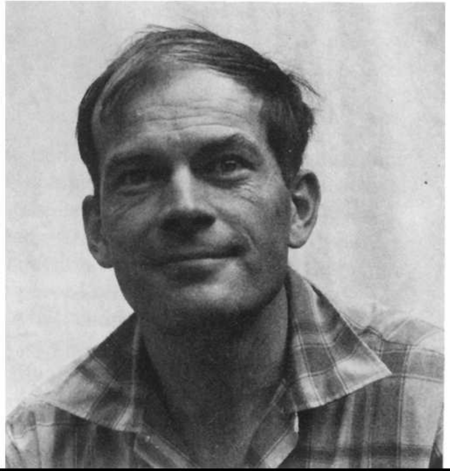
Applibugs – Lyon – 06/2016

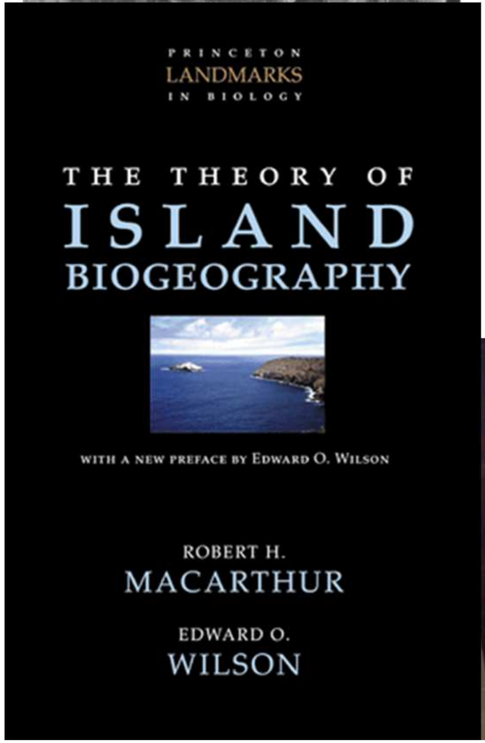
1 – Research questions

2 – Inverse methods and ABC

3 – ABC challenges for computationally-demanding models

Randomness and community dynamics





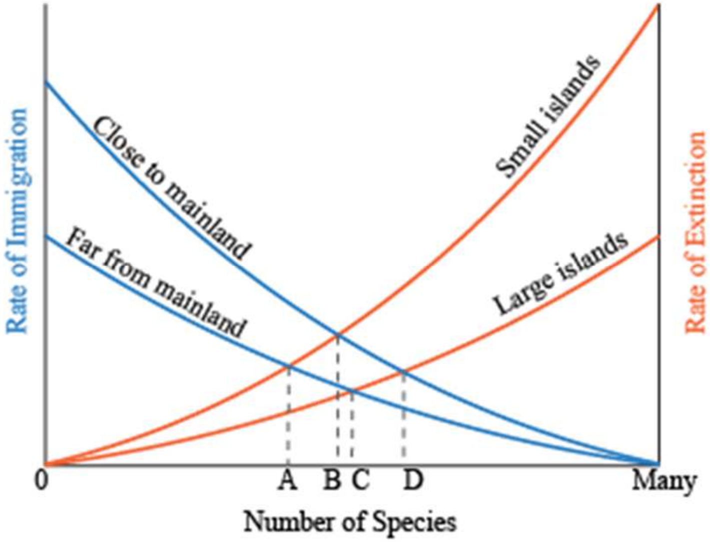
PRINCETON
LANDMARKS
IN BIOLOGY

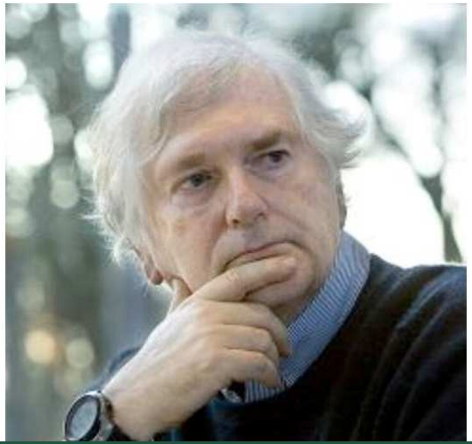
THE THEORY OF
ISLAND
BIOGEOGRAPHY

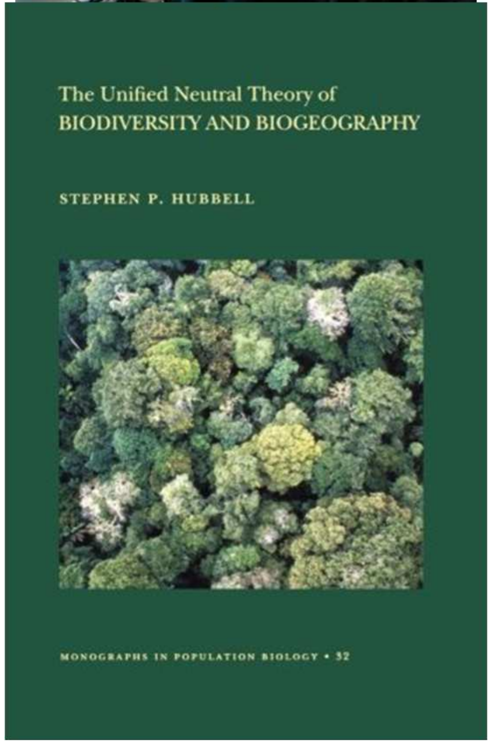
WITH A NEW PREFACE BY EDWARD O. WILSON

ROBERT H.
MACARTHUR

EDWARD O.
WILSON



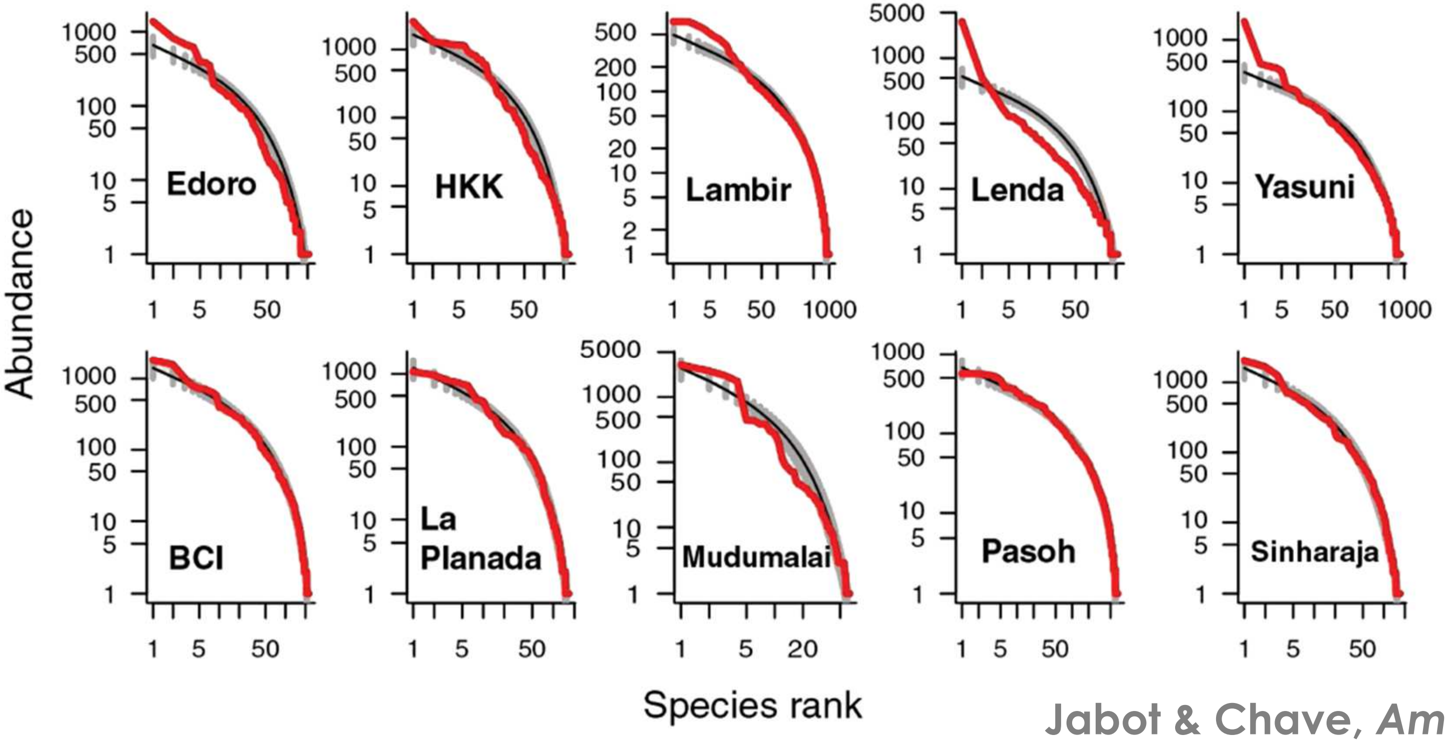




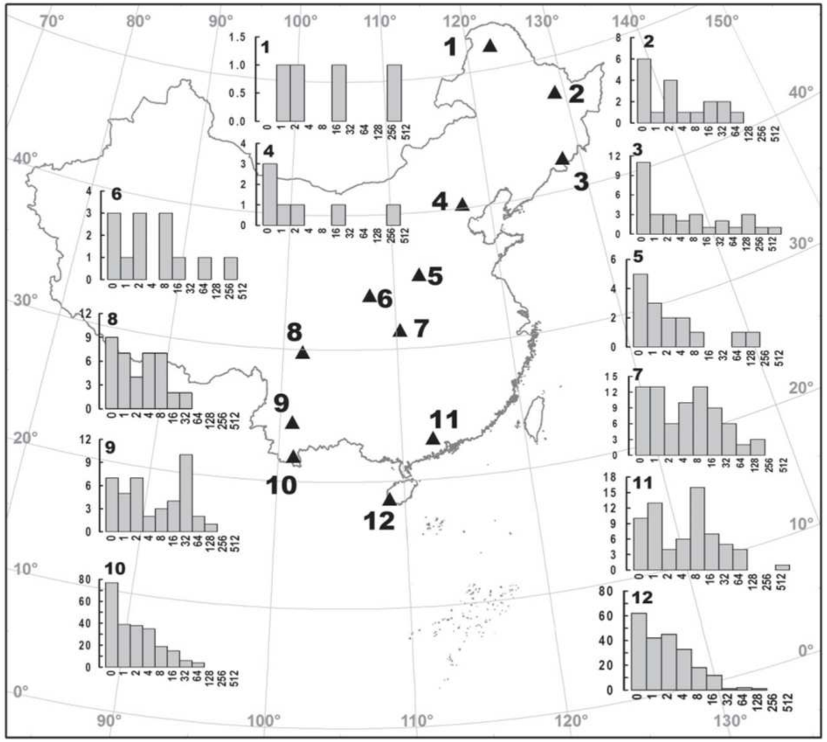
The Unified Neutral Theory of
BIODIVERSITY AND BIOGEOGRAPHY

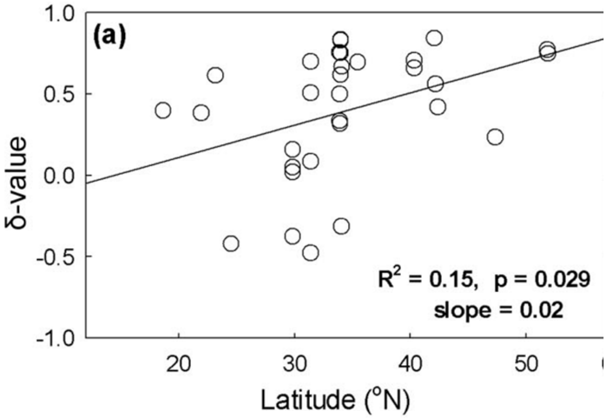
STEPHEN P. HUBBELL

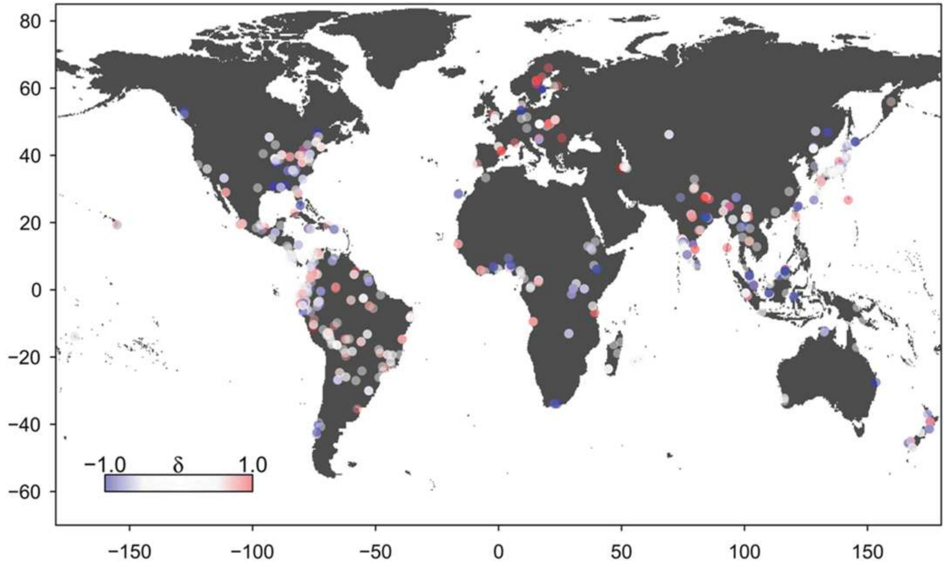
MONOGRAPHS IN POPULATION BIOLOGY • 32



Jabot & Chave, *Am. Nat.* 2011



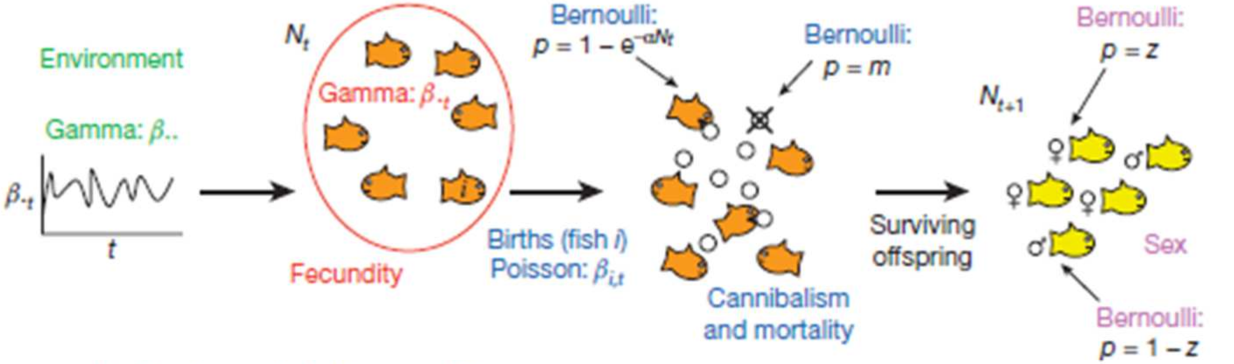




Qiao et al. *Glob. Ecol. Biogeo.* 2015

Kubota et al. *Oikos* 2016

Community dynamics



Demographic stochasticity	Environmental stochasticity	Demographic heterogeneity	Sex	Model	Abbreviation
●				Poisson	P
●	●			Negative binomial environmental	NBe
●		●		Negative binomial demographic	NBd
●	●	●		Negative binomial gamma	NBg
●			●	Poisson binomial	PB
●	●		●	Negative binomial-binomial environmental	NBBe
●		●	●	Negative binomial-binomial demographic	NBBd
●	●	●	●	Negative binomial-binomial gamma	NBBg

Mebourne & Hastings, *Nature* 2008

Community dynamics

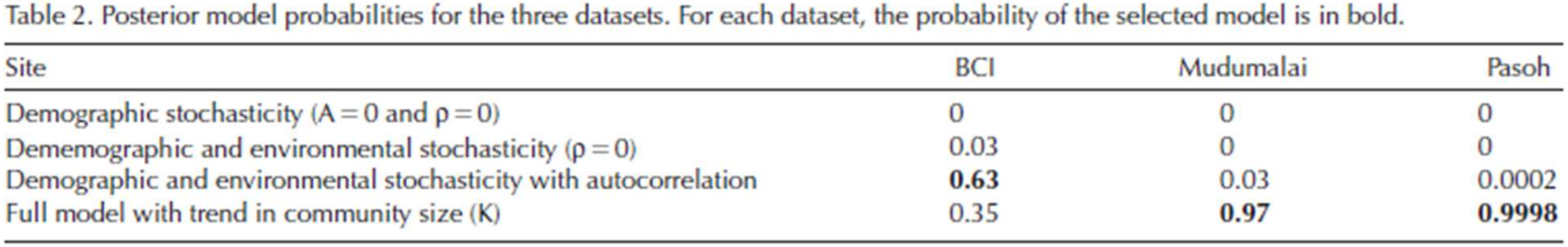
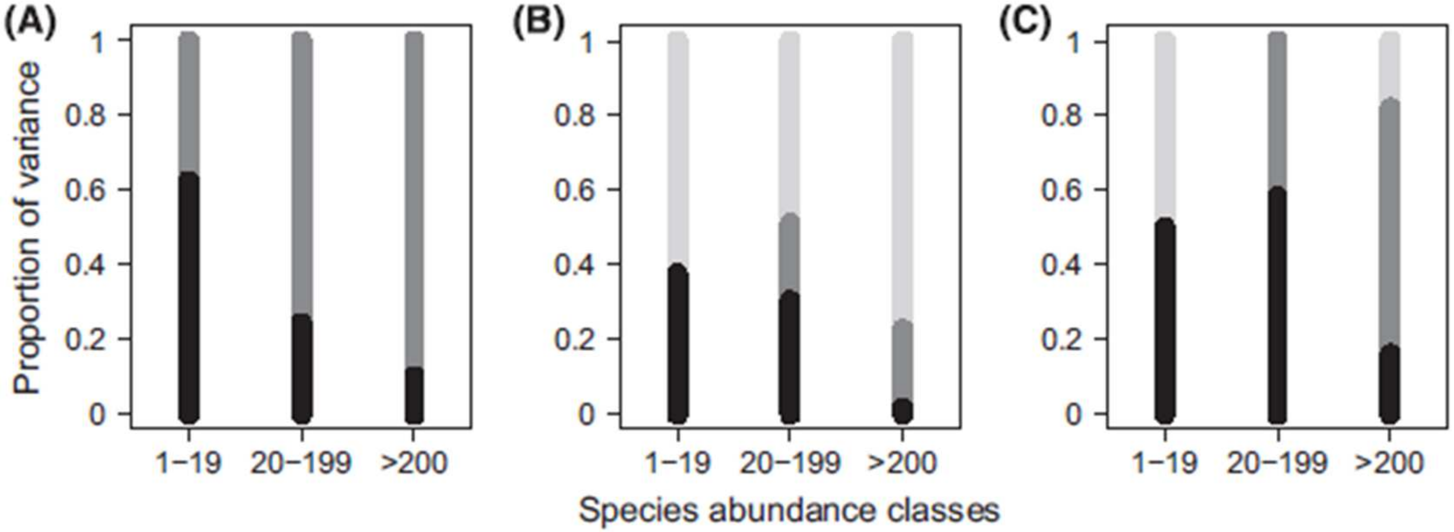


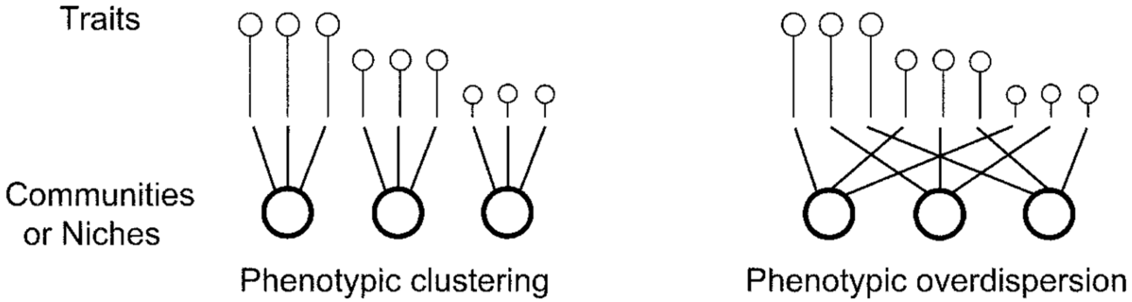
Table 2. Posterior model probabilities for the three datasets. For each dataset, the probability of the selected model is in bold.

Site	BCI	Mudumalai	Pasoh
Demographic stochasticity ($A = 0$ and $\rho = 0$)	0	0	0
Dememographic and environmental stochasticity ($\rho = 0$)	0.03	0	0
Demographic and environmental stochasticity with autocorrelation	**0.63**	0.03	0.0002
Full model with trend in community size (K)	0.35	**0.97**	**0.9998**

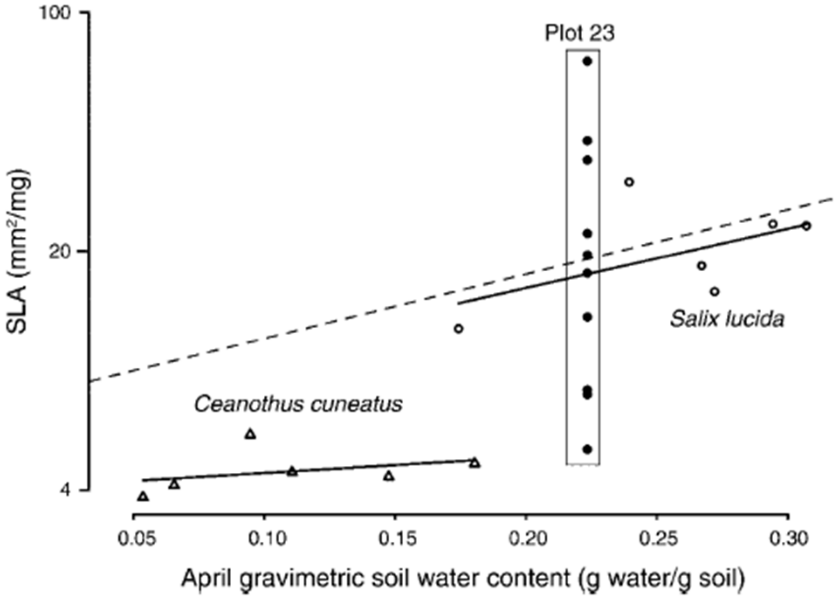


Jabot & Lohier, *Oikos* 2016

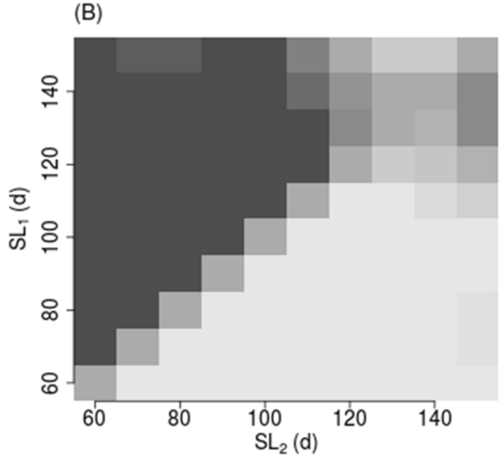
Functional integration

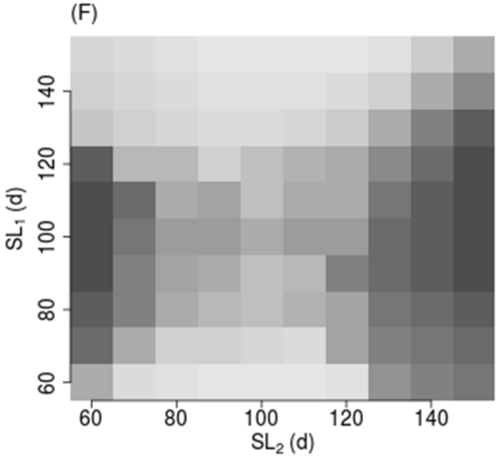


Cavender-Bares et al., *Am. Nat.* 2004

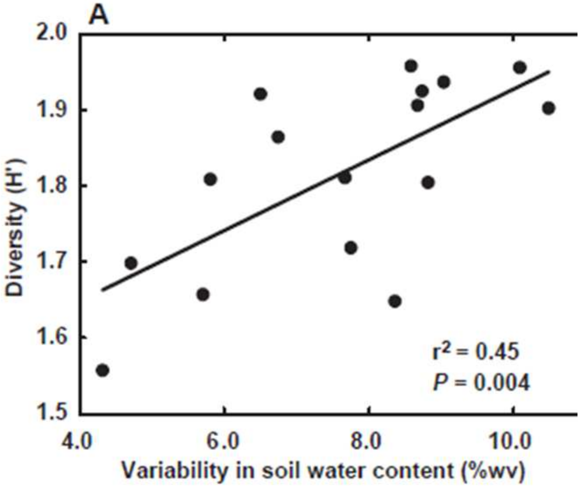


Cornwell et al., *Ecol. Monog.* 2009



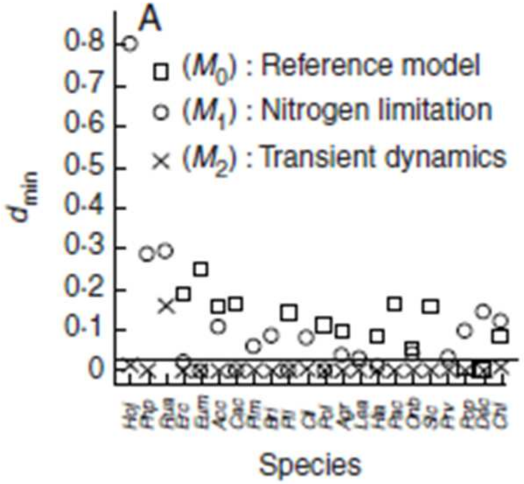


Lohier et al., *in prep.*



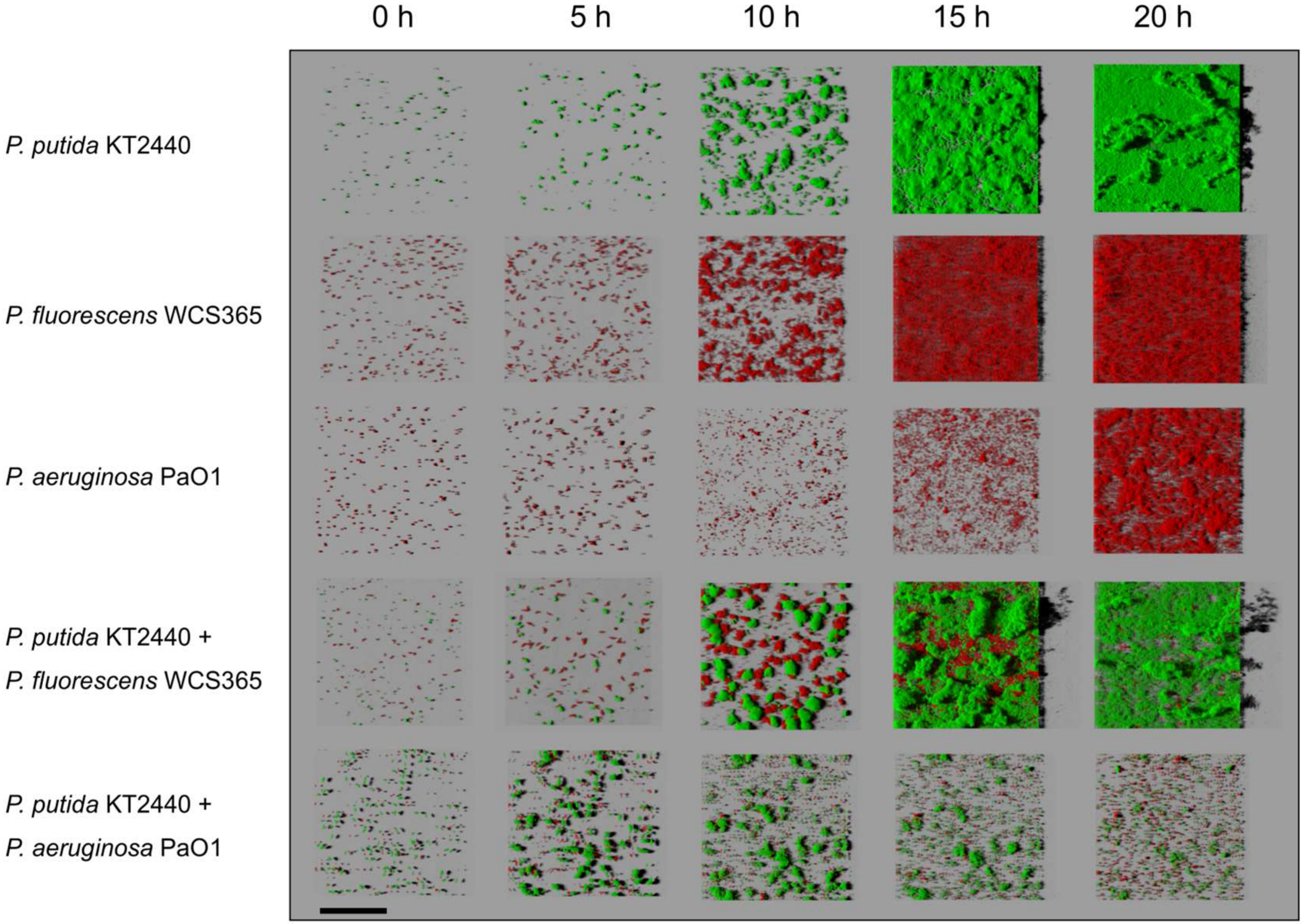
Knapp et al., *Science* 2002

Dynamics and functional integration shed light on optimal partitioning theory



Lohier et al., *Ann. Bot.* 2014

Statistical inference



Bridier et al., *Biofouling* 2014

Approximate Bayesian Computation (ABC)

Beaumont et al. *Genetics* 2002

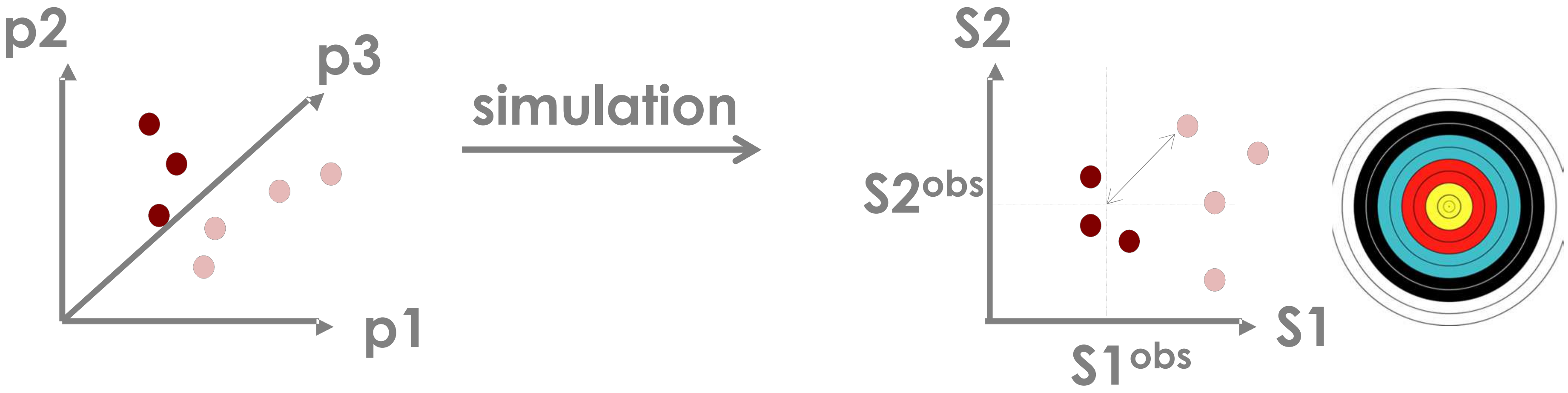
Beaumont *Ann. Rev. Ecol. Evol. Sys*. 2010

Marin et al. *Stat. Comp*. 2012

ABC in a nutshell

ABC in a nutshell

ABC in a nutshell



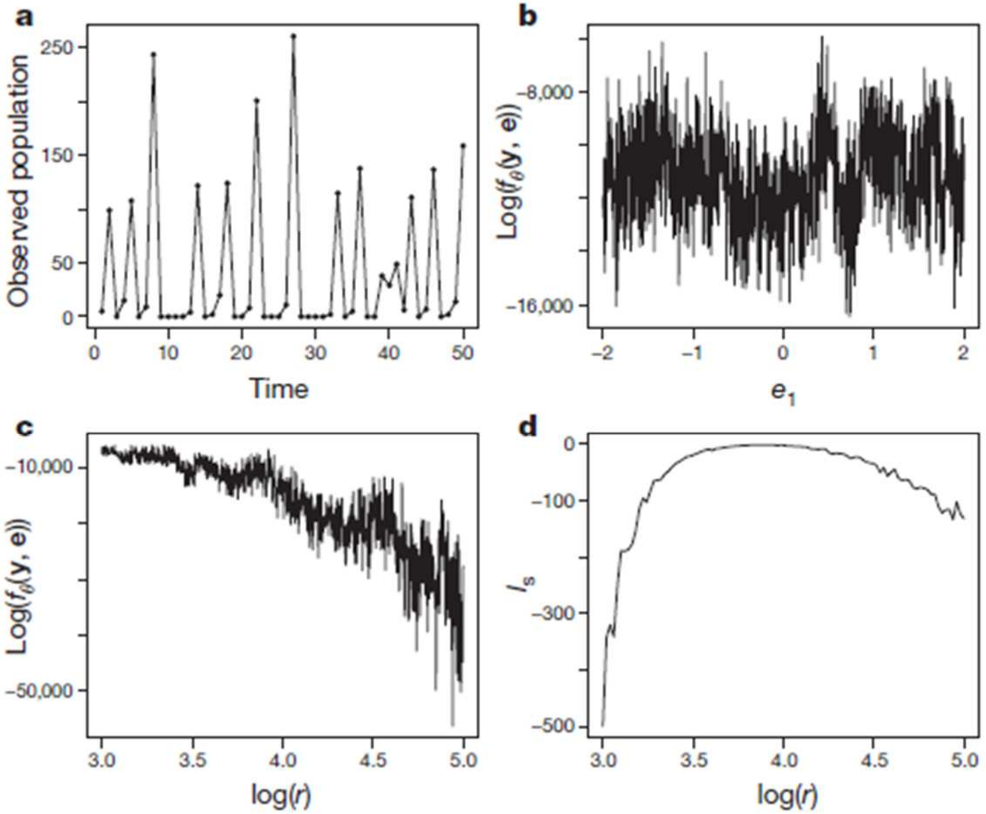
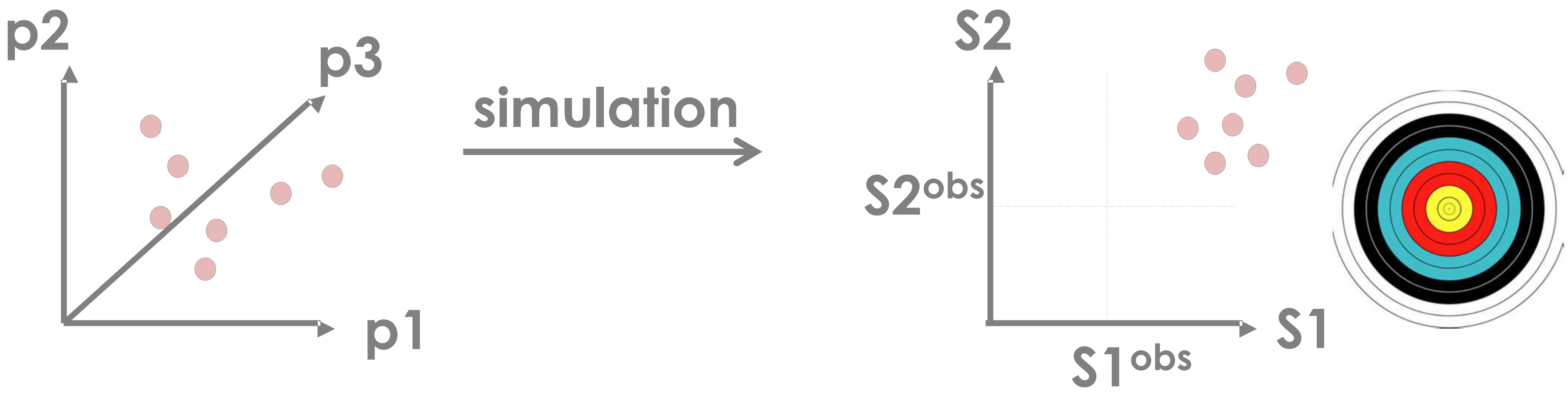


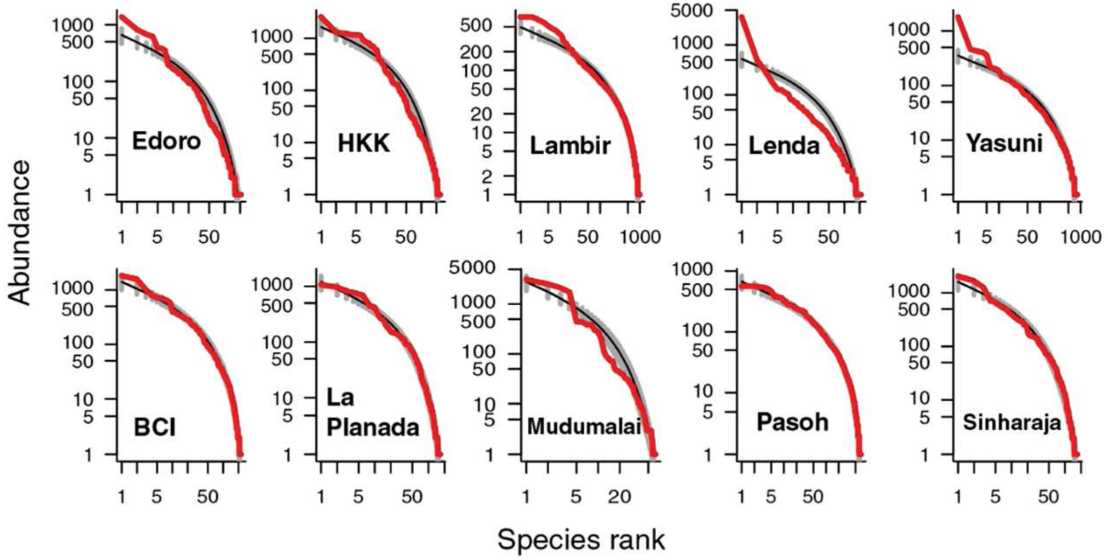
Figure 1 | Measuring fit of the Ricker model. a, Population data simulated

Wood, *Nature* 2010

Standard statistical tools

Model checking

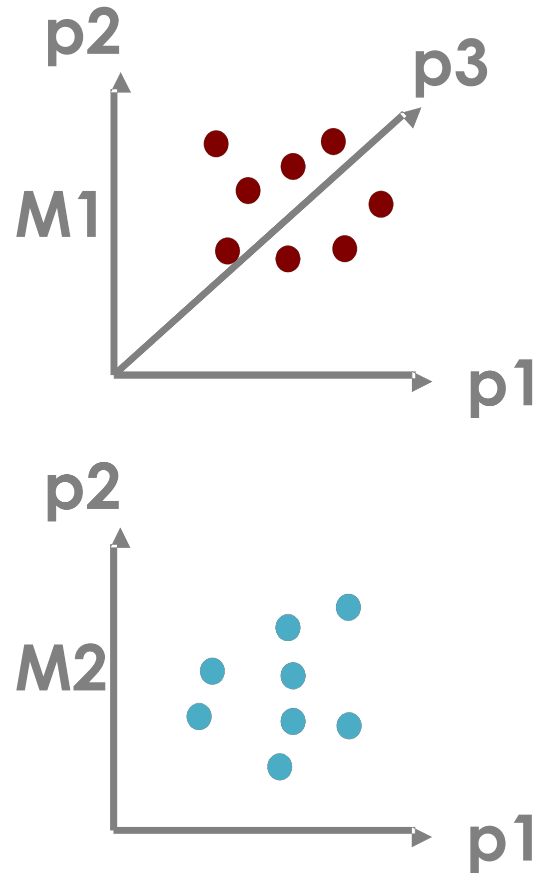


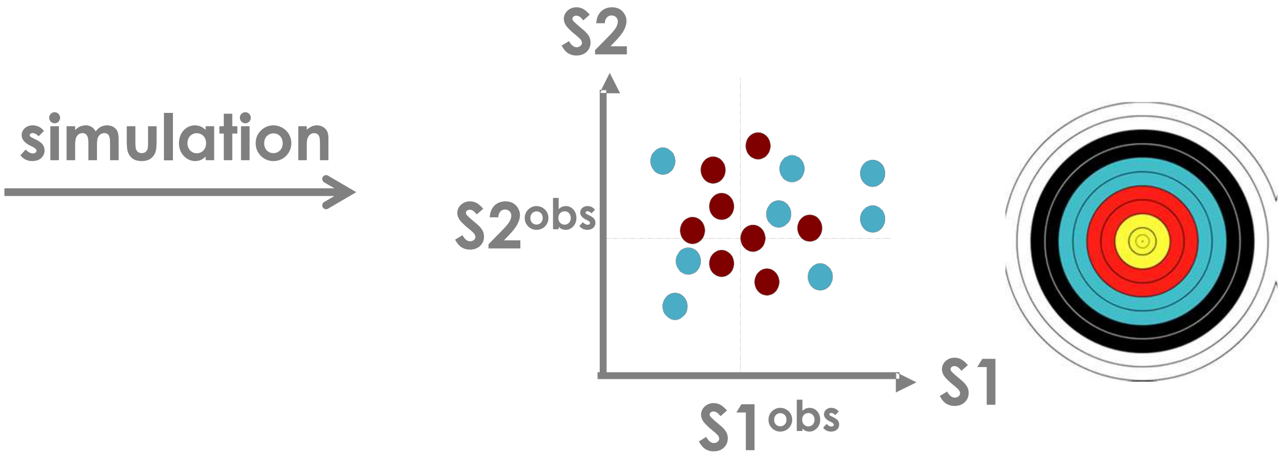


Ratmann et al. PNAS 2009

Jabot & Chave Am. Nat. 2011

Model selection





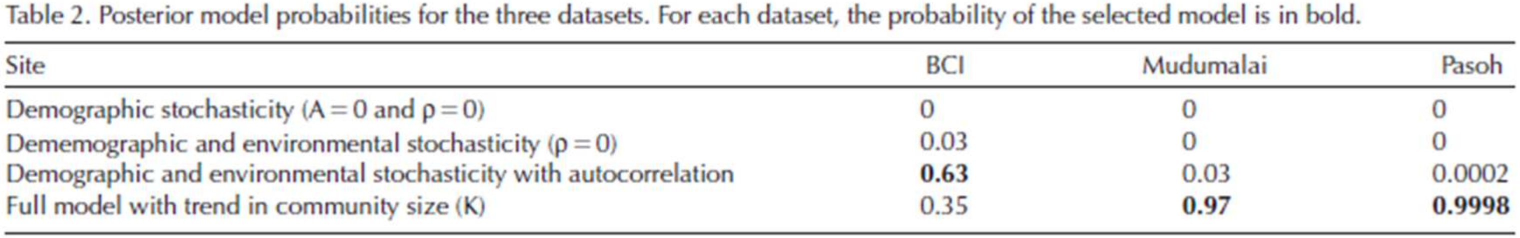


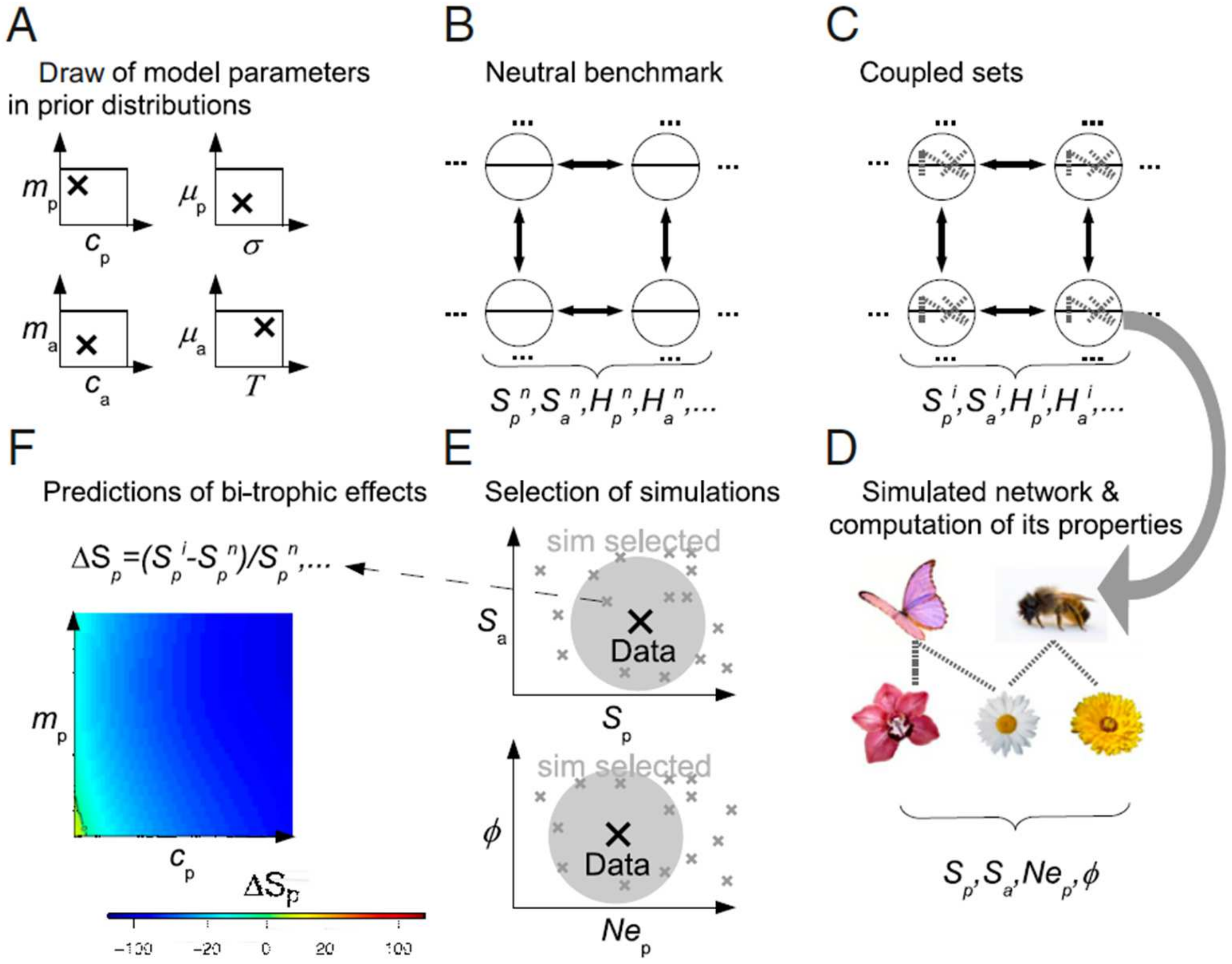
Table 2. Posterior model probabilities for the three datasets. For each dataset, the probability of the selected model is in bold.

Site	BCI	Mudumalai	Pasoh
Demographic stochasticity ($A = 0$ and $\rho = 0$)	0	0	0
Dememographic and environmental stochasticity ($\rho = 0$)	0.03	0	0
Demographic and environmental stochasticity with autocorrelation	**0.63**	0.03	0.0002
Full model with trend in community size (K)	0.35	**0.97**	**0.9998**

Beaumont 2008

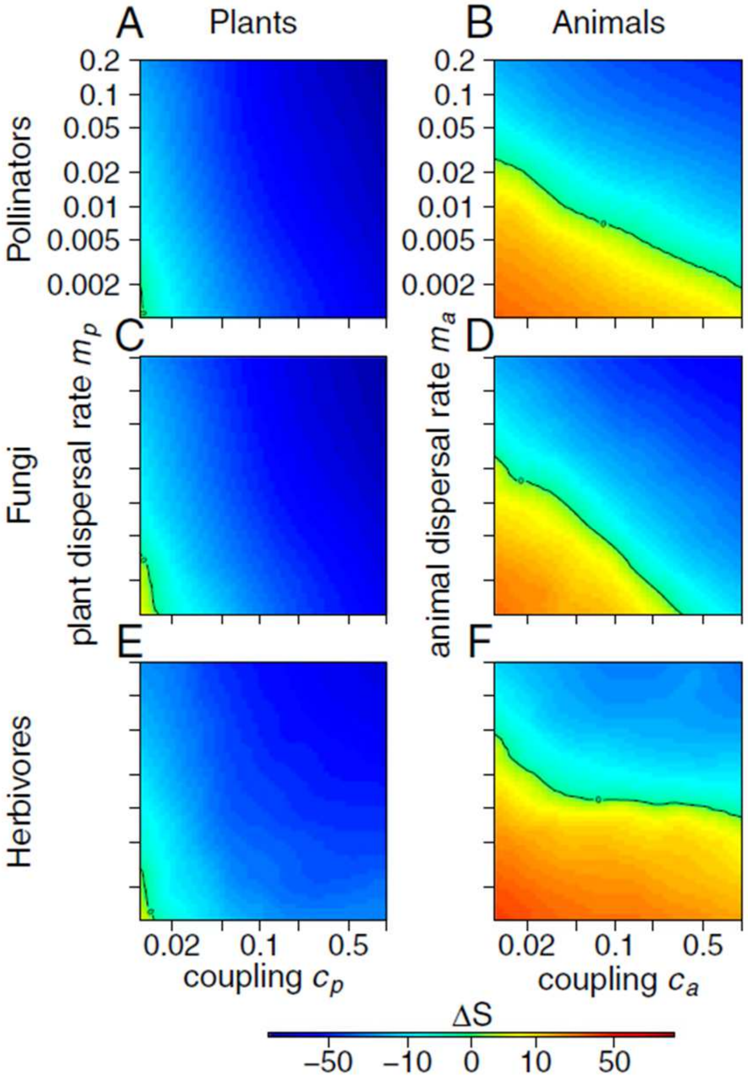
Jabot & Lohier, *Oikos* 2016

Model averaging



Jabot & Bascompte, *PNAS* 2012

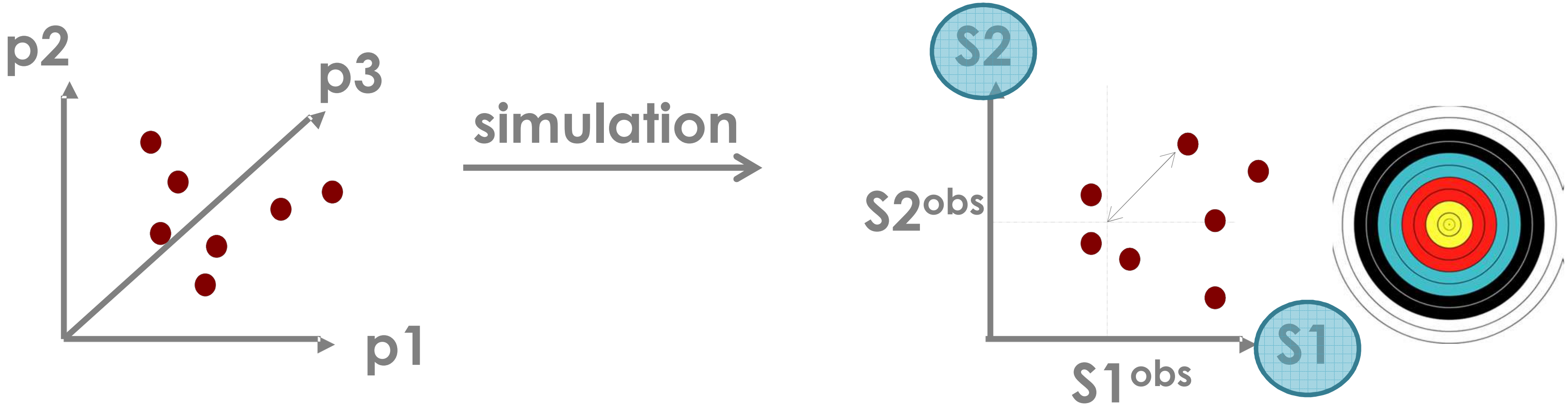
Model averaging



Jabot & Bascompte, *PNAS* 2012

Methodological works on ABC

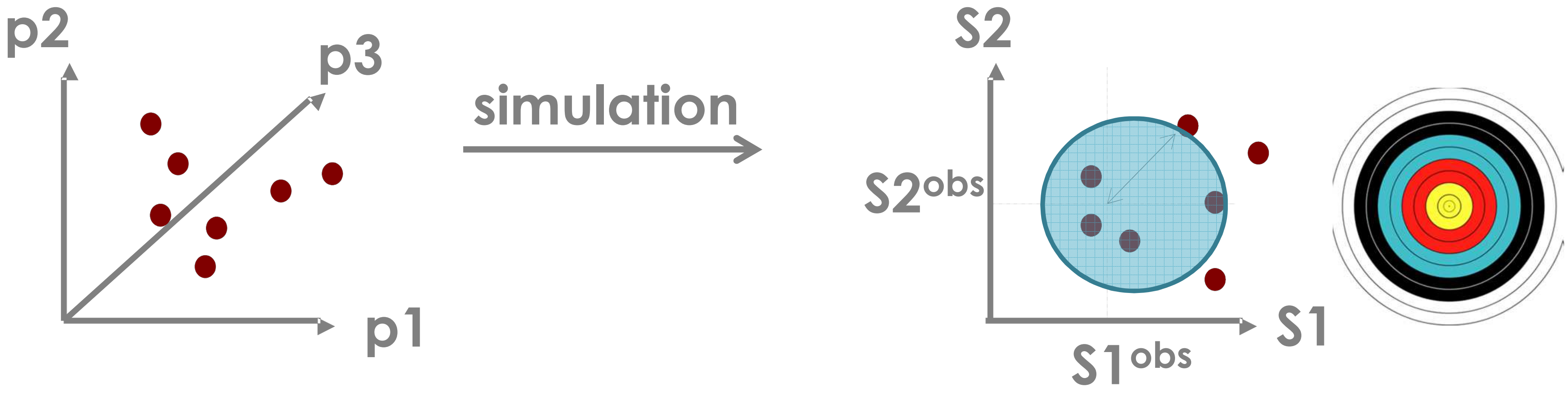
Methodological works on ABC



Choice of summary statistics:

Fearnhead & Prangle *J. Roy. Stat. Soc.* 2012
Blum et al. *Stat. Sci.* 2013

Methodological works on ABC



Choice of retained simulations (postprocessing treatments):

Beaumont et al. *Genetics* 2002
Blum & François *Stat. Comp.* 2010

-> R package abc, Csilléry et al. *Meth. Ecol. Evol.* 2012

Methodological works on ABC

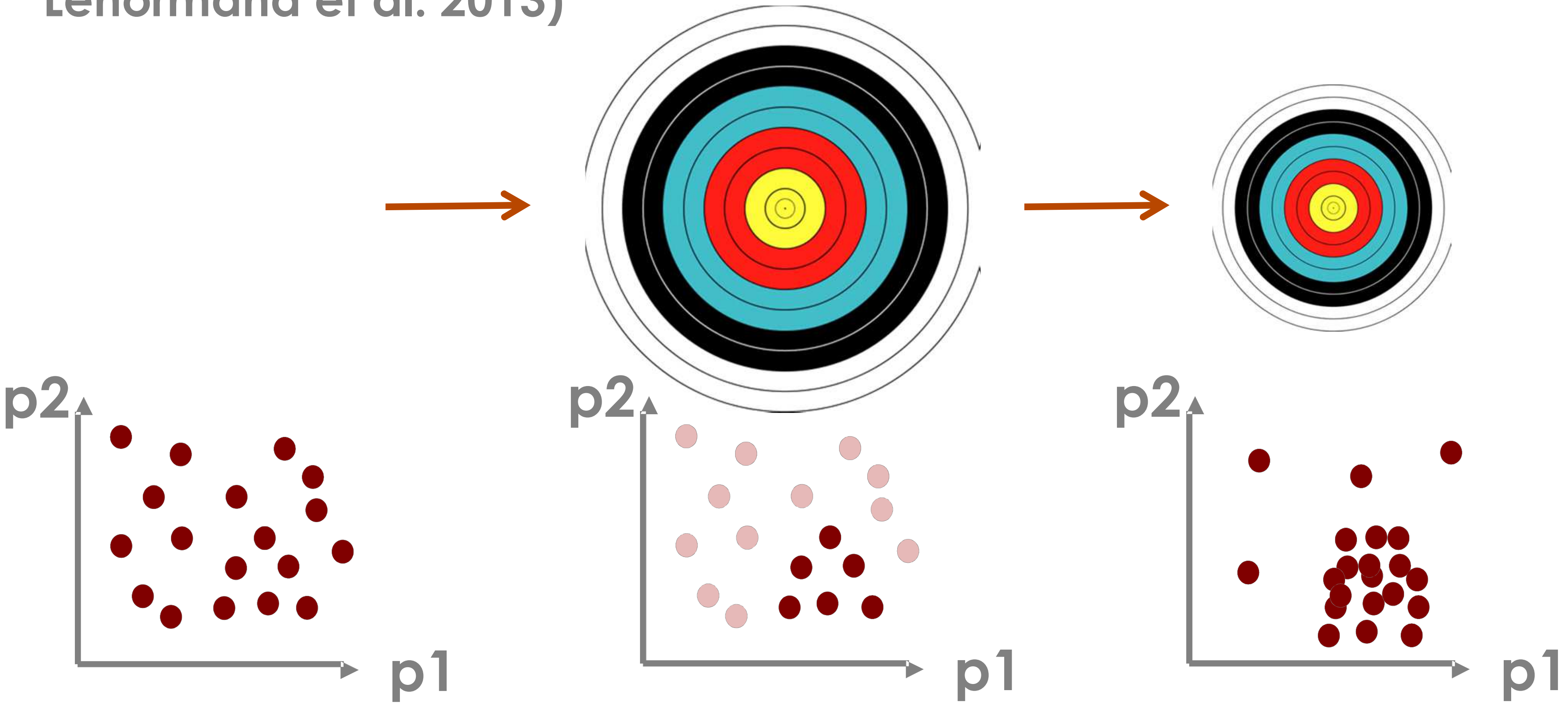
Exploration of the parameter space

package EasyABC, Jabot et al. *Meth. Ecol. Evol.* 2013

Algorithms to explore the parameter space efficiently

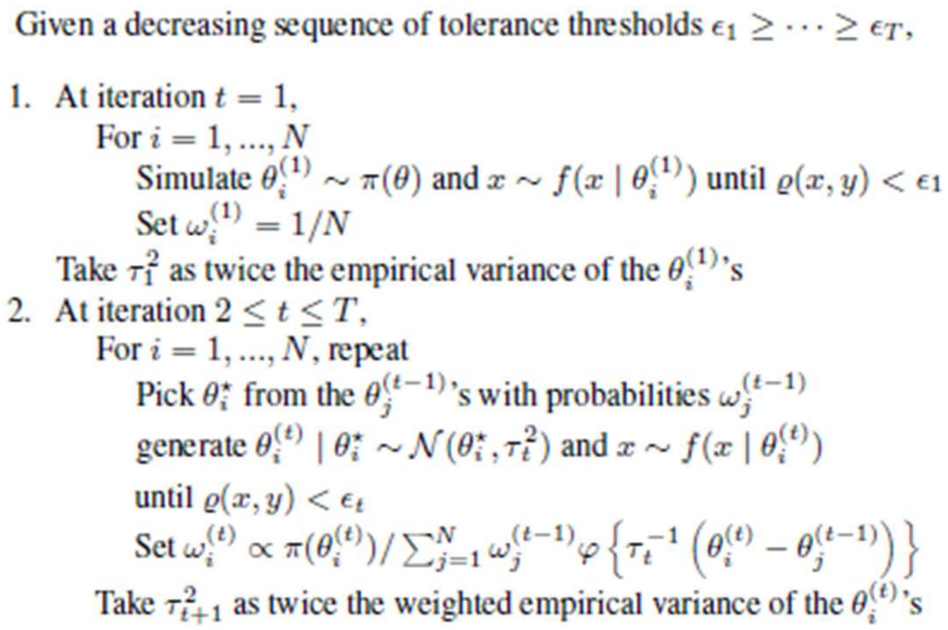
- ABC-MCMC (Marjoram et al. 2003, Wegmann et al. 2009) (replace likelihood by $1(d<\varepsilon)$)

- sequential techniques (Beaumont et al. 2008, Del Moral et al. 2012, Lenormand et al. 2013)



Algorithms to explore the parameter space efficiently

Beaumont et al. *Biometrika* 2009



```
Given a decreasing sequence of tolerance thresholds $\epsilon_1 \geq \cdots \geq \epsilon_T$,

1. At iteration $t = 1$,
      For $i = 1, \ldots, N$
         Simulate $\theta_i^{(1)} \sim \pi(\theta)$ and $x \sim f(x \mid \theta_i^{(1)})$ until $\varrho(x, y) < \epsilon_1$
         Set $\omega_i^{(1)} = 1/N$
   Take $\tau_1^2$ as twice the empirical variance of the $\theta_i^{(1)}$'s
2. At iteration $2 \leq t \leq T$,
      For $i = 1, \ldots, N$, repeat
         Pick $\theta_i^\star$ from the $\theta_j^{(t-1)}$'s with probabilities $\omega_j^{(t-1)}$
         generate $\theta_i^{(t)} \mid \theta_i^\star \sim \mathcal{N}(\theta_i^\star, \tau_t^2)$ and $x \sim f(x \mid \theta_i^{(t)})$
         until $\varrho(x, y) < \epsilon_t$
         Set $\omega_i^{(t)} \propto \pi(\theta_i^{(t)}) / \sum_{j=1}^{N} \omega_j^{(t-1)} \varphi \left\{ \tau_t^{-1} \left( \theta_i^{(t)} - \theta_j^{(t-1)} \right) \right\}$
      Take $\tau_{t+1}^2$ as twice the weighted empirical variance of the $\theta_i^{(t)}$'s
```

Algorithms to explore the parameter space efficiently

Lenormand et al. *Comput. Stat*. 2013

Idea 1: Automatic generation of the sequence of tolerance levels ε_i
(Drovandi & Pettit 2011, Del Moral et al. 2012)

-> as a quantile (median) of particle distances.

Idea 2: Criteria to stop the algorithm

-> proportion (0.05) of newly accepted particles.

Algorithms to explore the parameter space efficiently

Lenormand et al. *Comput. Stat.* 2013

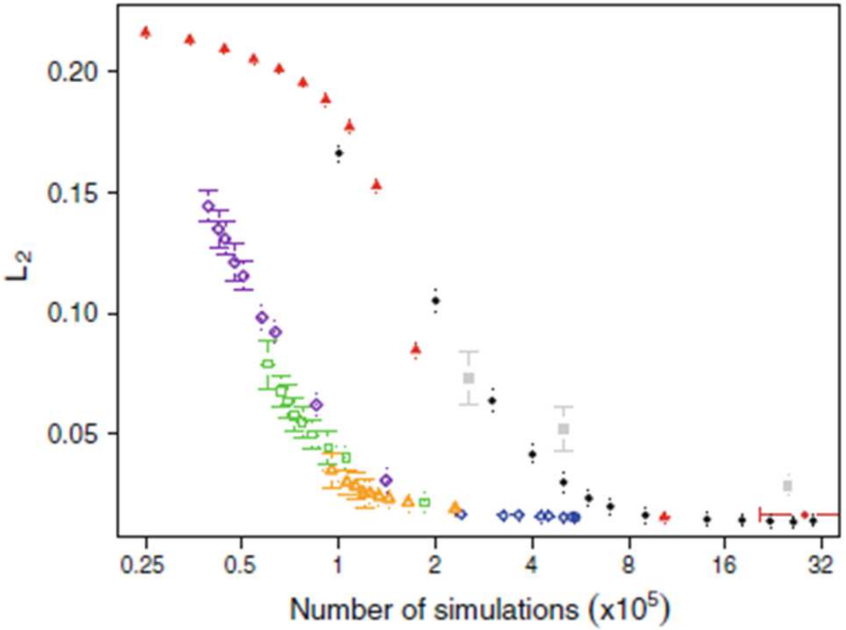
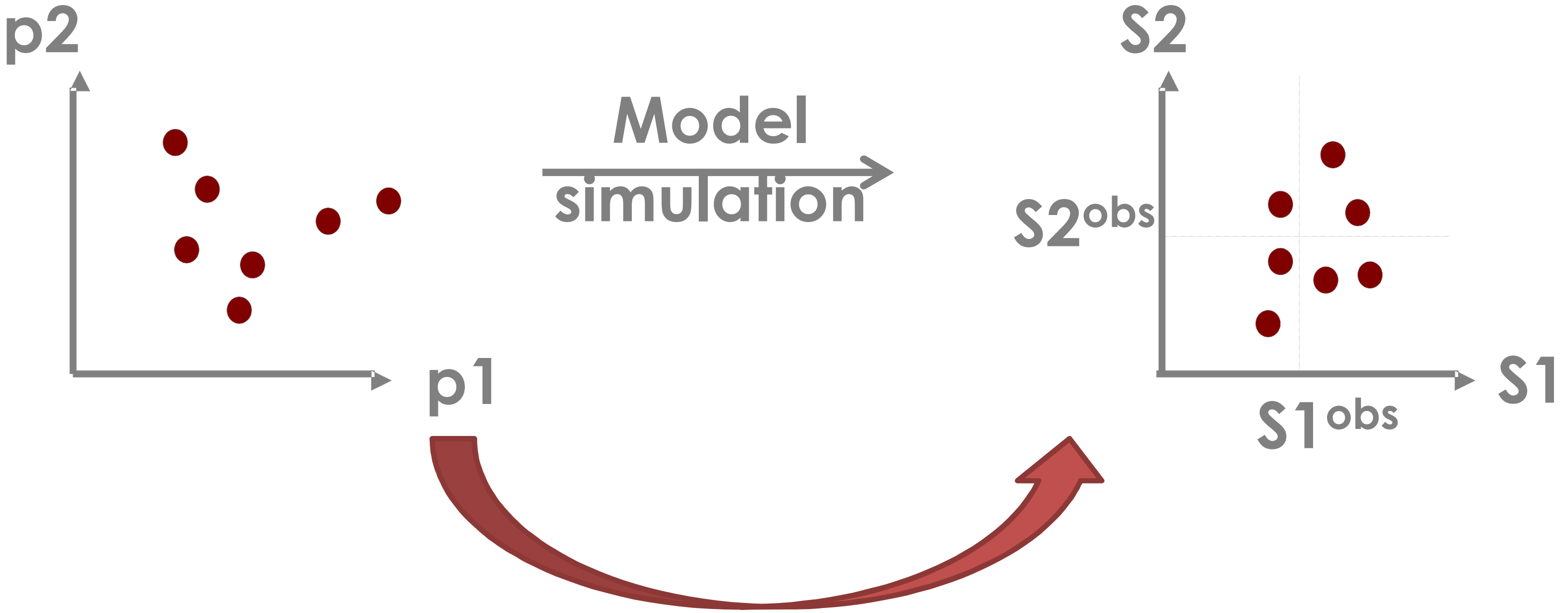


Fig. 2 Posterior quality ($\mathbb{L}_2$) versus computing cost (number of simulations) averaged over 50 replicates. *Vertical* and *horizontal bars* represent the standard deviations among replicates. Algorithm parameters used for APMC: α in $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and $p_{acc_{min}}$ in $\{0.01, 0.05, 0.1, 0.2\}$. *Blue circles* are used for $p_{acc_{min}} = 0.01$, *orange triangles* for $p_{acc_{min}} = 0.05$, *green squares* for $p_{acc_{min}} = 0.1$, and *purple diamonds* for $p_{acc_{min}} = 0.2$. PMC: *red plain triangles* for a sequence of tolerance levels from $\varepsilon_1 = 2$ down to $\varepsilon_{11} = 0.01$. SMC: *grey plain square* for α in $\{0.9, 0.95, 0.99\}$ (from *left* to *right*), $M = 1$ and a ε target equal to 0.01. RSMC: *brown plain diamond* for $\alpha = 0.5$ and a ε target equal to 0.01. Results obtained with a standard rejection-based ABC algorithm are depicted with *black plain circles* (colour figure online)

Algorithms to explore the parameter space efficiently

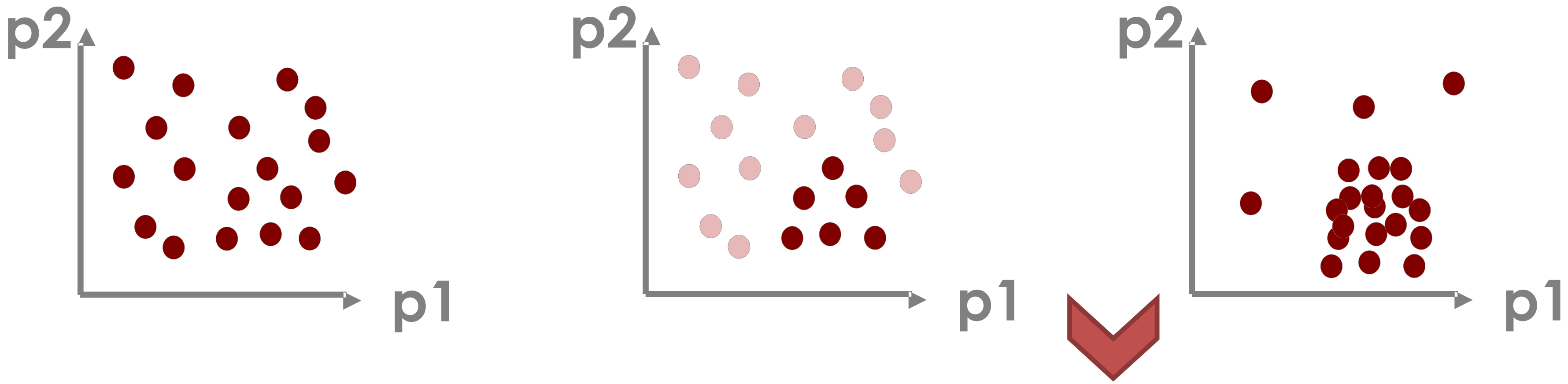
Emulation-based sequential ABC (Jabot et al. Arxiv 2014)



Fit of a meta-model (emulator):
eg. local regressions.

Algorithms to explore the parameter space efficiently

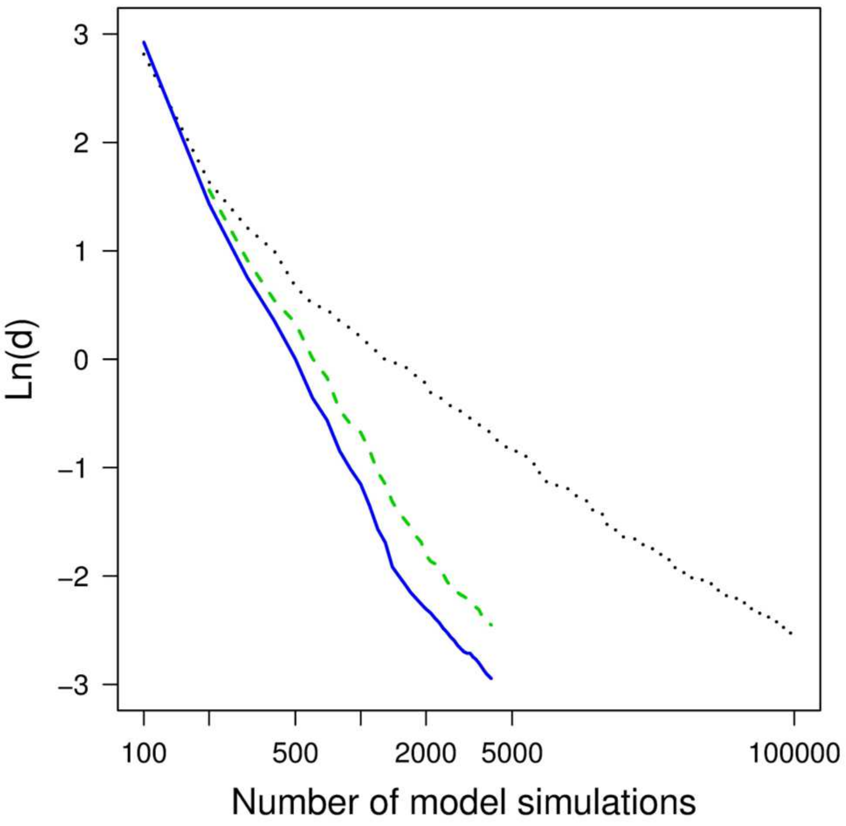
Emulation-based sequential ABC (Jabot et al. Arxiv 2014)



sequential ABC using the emulator

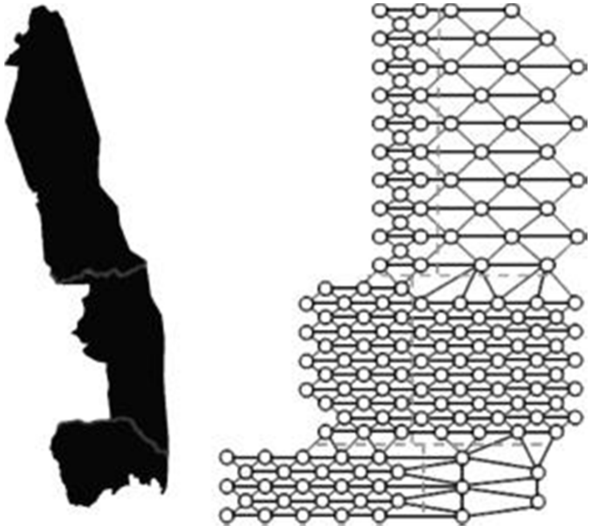
Algorithms to explore the parameter space efficiently

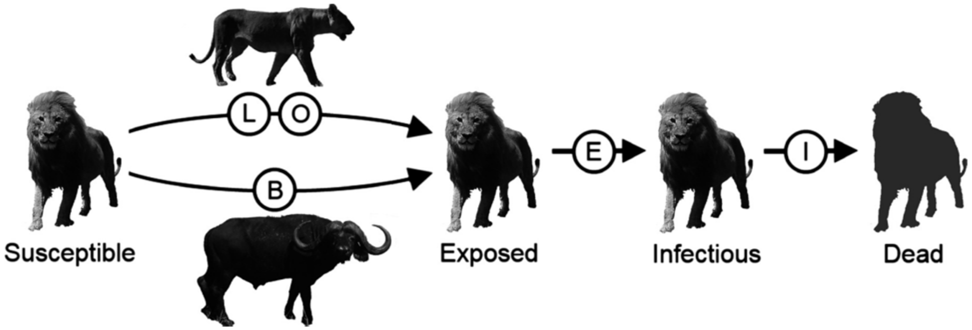
Emulation-based sequential ABC (Jabot et al. Arxiv 2014)

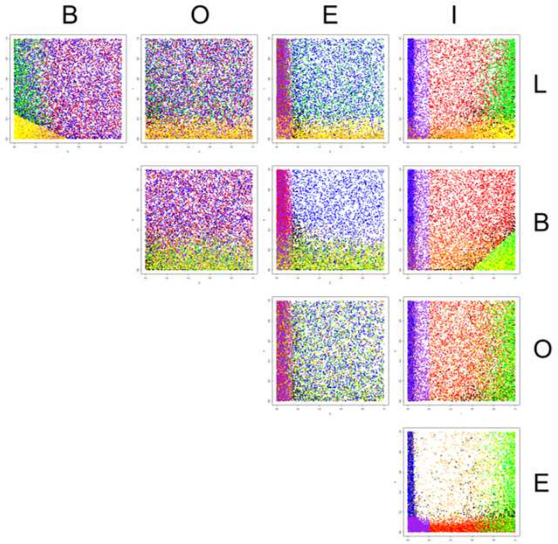


Algorithm available in the package EasyABC

SMC-ABC and multiple modes

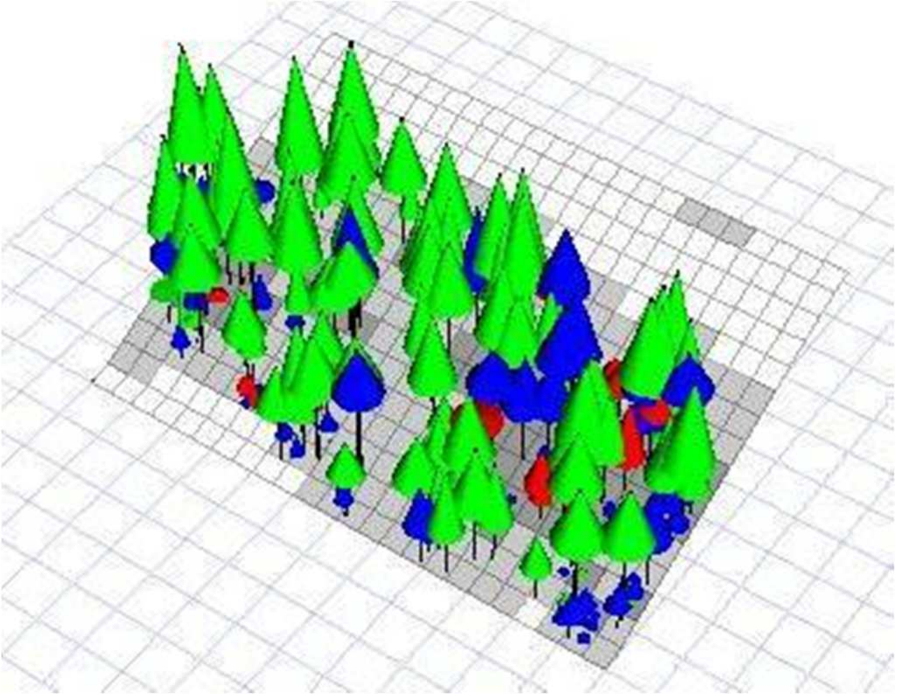






Kosmala et al., *Ecol. Appli.* 2016

ABC « re-calibration »



Lagarrigues et al., *Ecol. Model.* 2015

ABC « re-calibration »

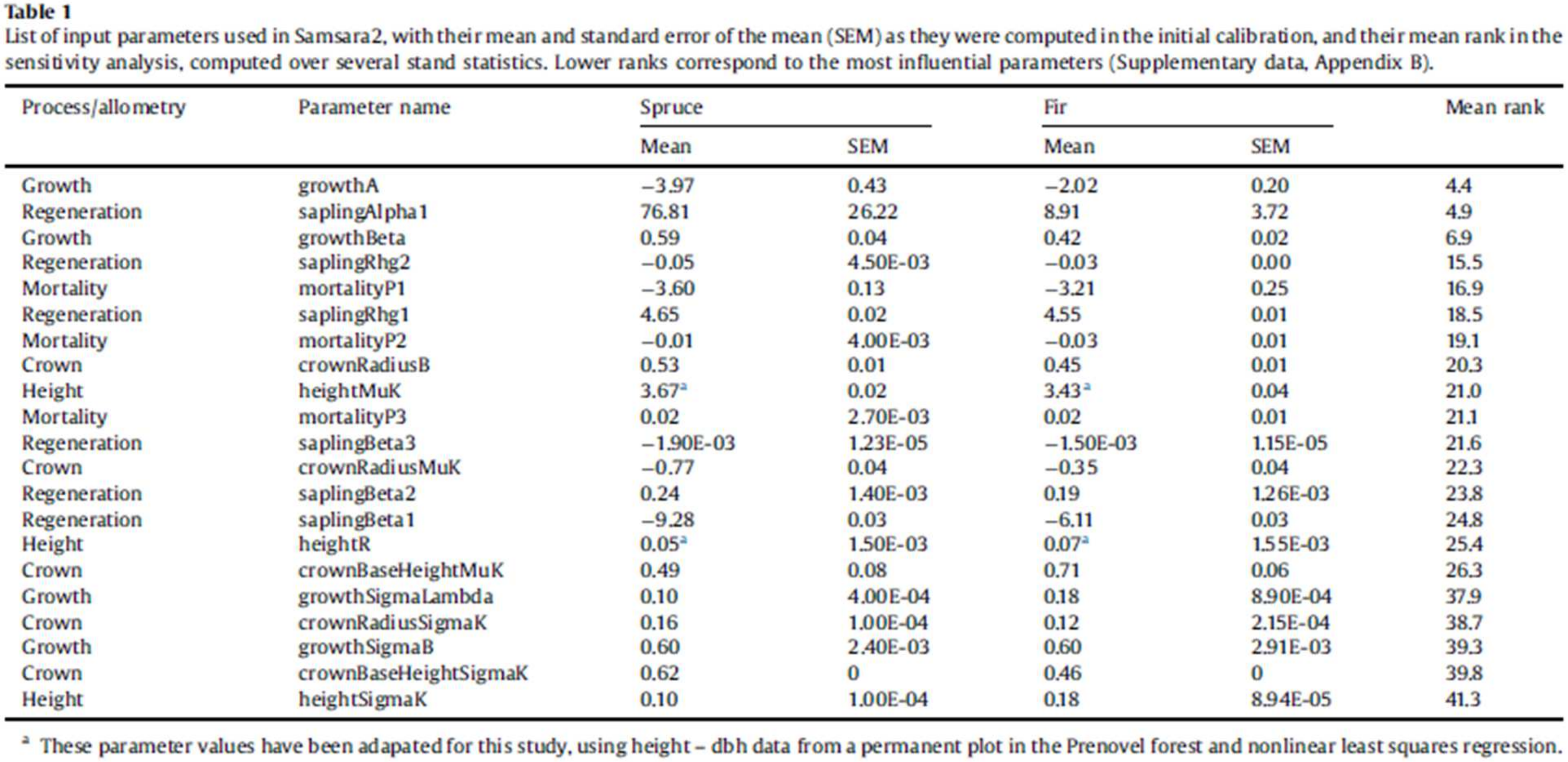


Table 1
List of input parameters used in Samsara2, with their mean and standard error of the mean (SEM) as they were computed in the initial calibration, and their mean rank in the sensitivity analysis, computed over several stand statistics. Lower ranks correspond to the most influential parameters (Supplementary data, Appendix B).

Process/allometry	Parameter name	Spruce		Fir		Mean rank
		Mean	SEM	Mean	SEM	
Growth	growthA	−3.97	0.43	−2.02	0.20	4.4
Regeneration	saplingAlpha1	76.81	26.22	8.91	3.72	4.9
Growth	growthBeta	0.59	0.04	0.42	0.02	6.9
Regeneration	saplingRhg2	−0.05	4.50E-03	−0.03	0.00	15.5
Mortality	mortalityP1	−3.60	0.13	−3.21	0.25	16.9
Regeneration	saplingRhg1	4.65	0.02	4.55	0.01	18.5
Mortality	mortalityP2	−0.01	4.00E-03	−0.03	0.01	19.1
Crown	crownRadiusB	0.53	0.01	0.45	0.01	20.3
Height	heightMuK	3.67[a]	0.02	3.43[a]	0.04	21.0
Mortality	mortalityP3	0.02	2.70E-03	0.02	0.01	21.1
Regeneration	saplingBeta3	−1.90E-03	1.23E-05	−1.50E-03	1.15E-05	21.6
Crown	crownRadiusMuK	−0.77	0.04	−0.35	0.04	22.3
Regeneration	saplingBeta2	0.24	1.40E-03	0.19	1.26E-03	23.8
Regeneration	saplingBeta1	−9.28	0.03	−6.11	0.03	24.8
Height	heightR	0.05[a]	1.50E-03	0.07[a]	1.55E-03	25.4
Crown	crownBaseHeightMuK	0.49	0.08	0.71	0.06	26.3
Growth	growthSigmaLambda	0.10	4.00E-04	0.18	8.90E-04	37.9
Crown	crownRadiusSigmaK	0.16	1.00E-04	0.12	2.15E-04	38.7
Growth	growthSigmaB	0.60	2.40E-03	0.60	2.91E-03	39.3
Crown	crownBaseHeightSigmaK	0.62	0	0.46	0	39.8
Height	heightSigmaK	0.10	1.00E-04	0.18	8.94E-05	41.3

[a] These parameter values have been adapated for this study, using height – dbh data from a permanent plot in the Prenovel forest and nonlinear least squares regression.

Lagarrigues et al., *Ecol. Model.* 2015

ABC « re-calibration »

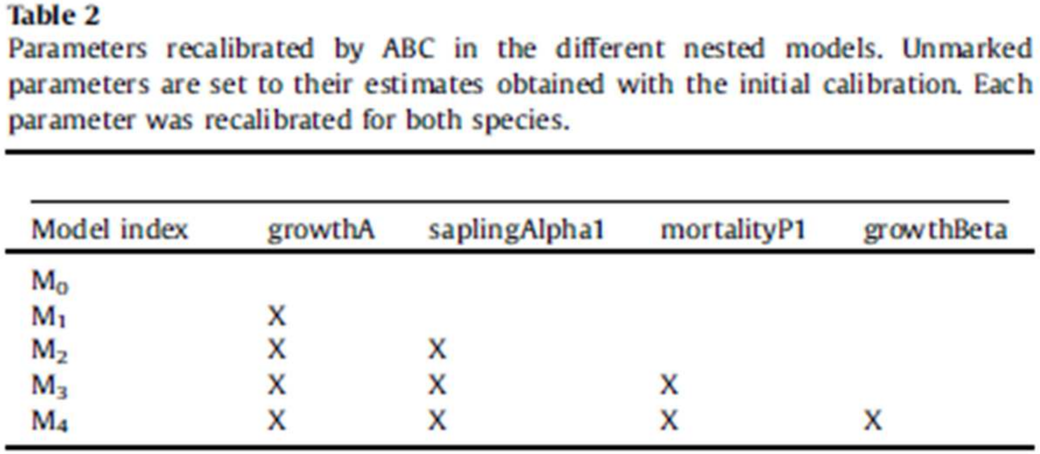
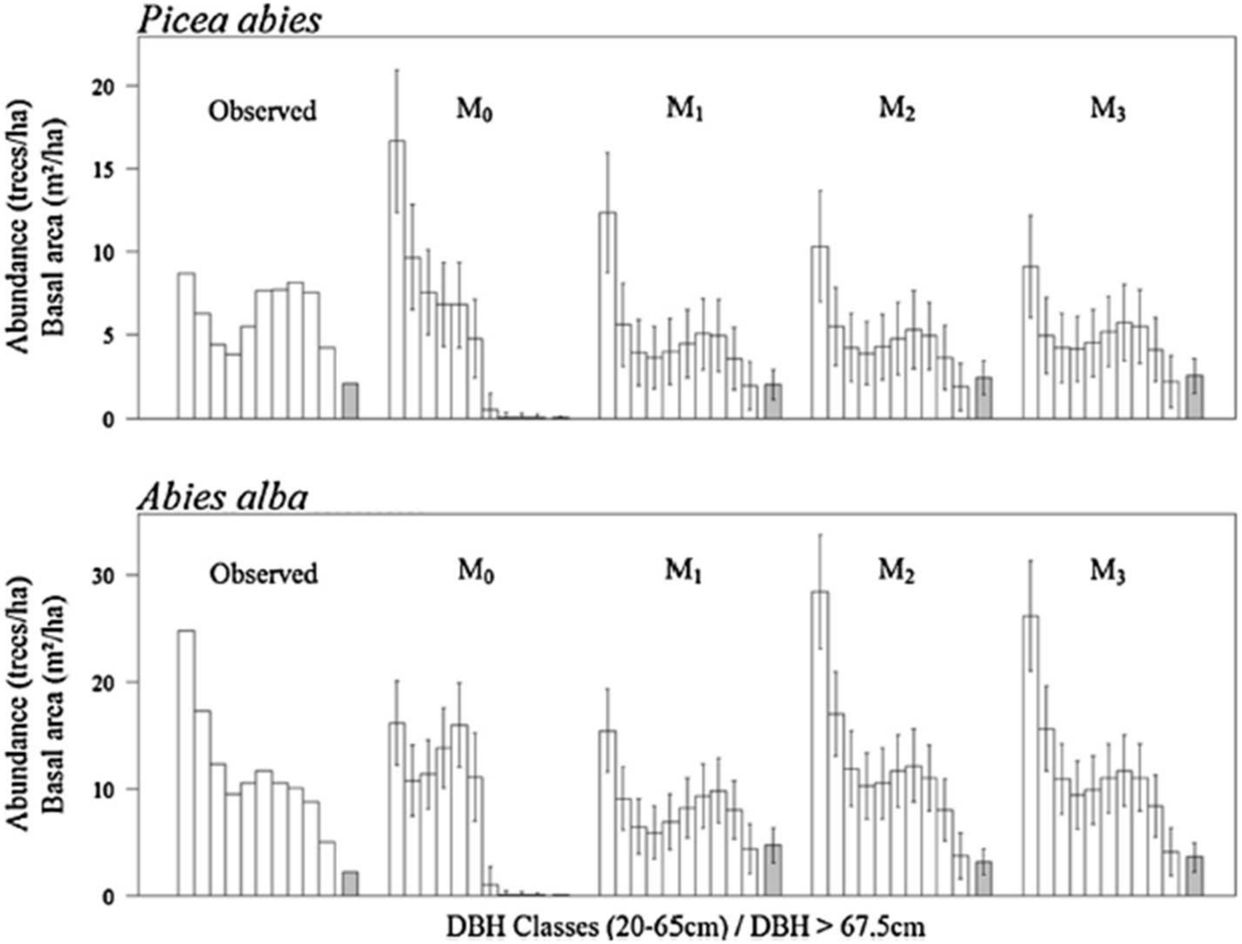


Table 2
Parameters recalibrated by ABC in the different nested models. Unmarked parameters are set to their estimates obtained with the initial calibration. Each parameter was recalibrated for both species.

Model index	growthA	saplingAlpha1	mortalityP1	growthBeta
M_0				
M_1	X			
M_2	X	X		
M_3	X	X	X	
M_4	X	X	X	X

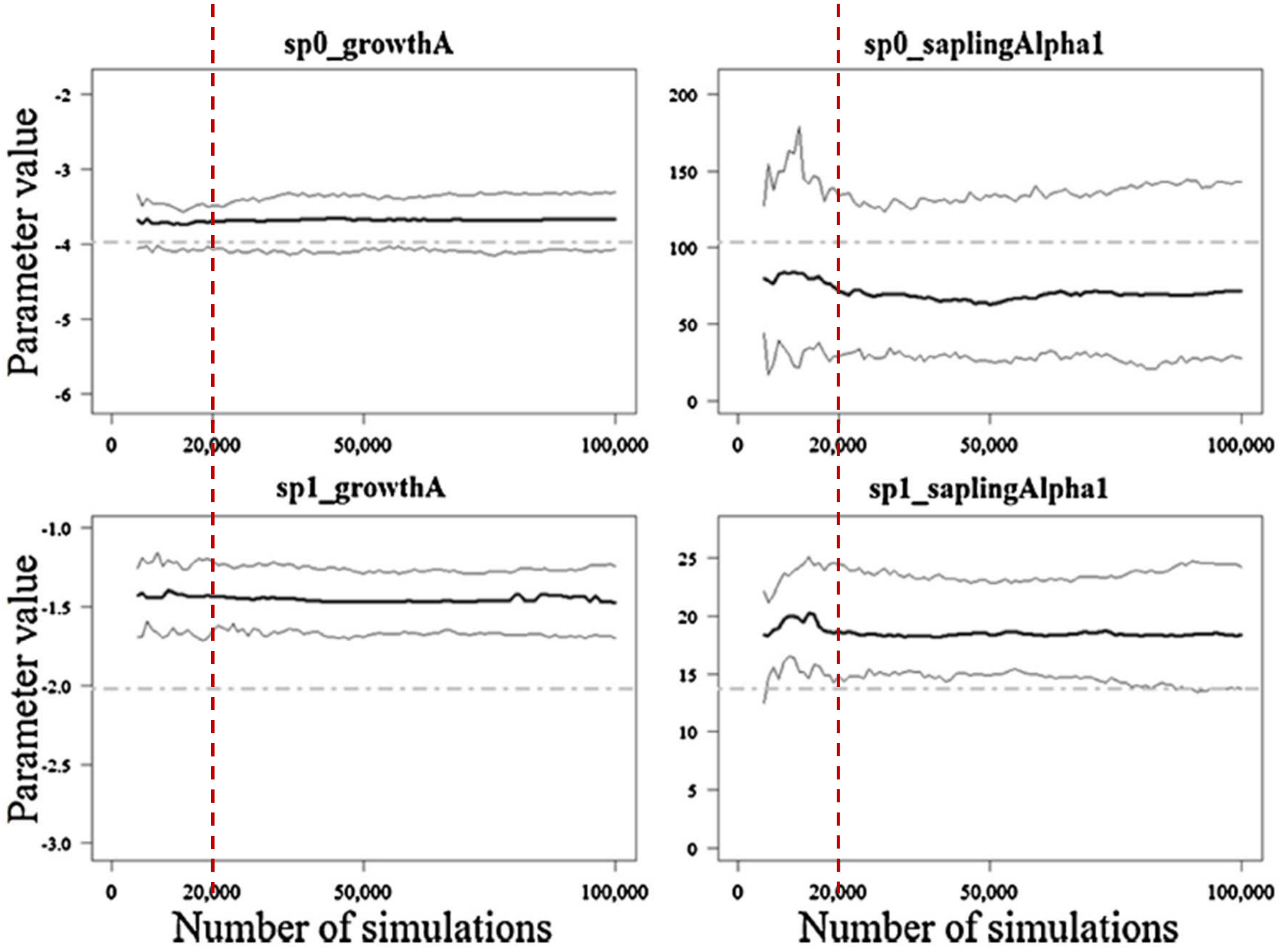
Lagarrigues et al., *Ecol. Model.* 2015

ABC « re-calibration »



Lagarrigues et al., *Ecol. Model.* 2015

ABC « re-calibration »



Lagarrigues et al., *Ecol. Model.* 2015

Collaborators :

LISC: Nicolas Dumoulin
Thierry Faure
Maxime Lenormand
Théophile Lohier

Jordi Bascompte, Univ. Zürich
Arnaud Bridier, Irstea Antony
Jérôme Chave, CNRS Toulouse
Benoît Courbaud, Irstea Grenoble
Yasuhiro Kubota, Univ. Ryukyus
Guillaume Lagarrigues, Irstea Grenoble
Xiujuan Qiao, Univ. Peking

irstea
AUVERGNE
Agence Nationale de la Recherche ANR
ADEME
Agence de l'Environnement et de la Maîtrise de l'Energie
Office National des Forêts