# *Calibration and evaluation of individual-based models using Approximate Bayesian Computation*

Article

Published Version

van der Vaart, E., Beaumont, M. A., Johnston, A. S. A. and Sibly, R. M. ORCID: https://orcid.org/0000-0001-6828-3543 (2015) Calibration and evaluation of individual-based models using Approximate Bayesian Computation. Ecological Modelling, 312. pp. 182-190. ISSN 0304-3800 doi: 10.1016/j.ecolmodel.2015.05.020 Available at https://centaur.reading.ac.uk/40526/

It is advisable to refer to the publisher's version if you intend to cite from the work.  See [Guidance on citing](#).

# www.reading.ac.uk/centaur

# CentAUR

Central Archive at the University of Reading

Reading's research outputs online
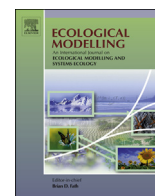
# Calibration and evaluation of individual-based models using Approximate Bayesian Computation

Elske van der Vaart [a,*], Mark A. Beaumont [b,c], Alice S.A. Johnston [a], Richard M. Sibly [a]

[a] School of Biological Sciences, University of Reading, Harborne Building, University of Reading, Whiteknights, Reading, Berkshire , RG6 6AS, United Kingdom
[b] School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom
[c] School of Biological Sciences, University of Bristol, Woodland Road, Bristol, BS8 1UG, United Kingdom

## ABSTRACT

This paper investigates the feasibility of using Approximate Bayesian Computation (ABC) to calibrate and evaluate complex individual-based models (IBMs). As ABC evolves, various versions are emerging, but here we only explore the most accessible version, rejection-ABC. Rejection-ABC involves running models a large number of times, with parameters drawn randomly from their prior distributions, and then retaining the simulations closest to the observations. Although well-established in some fields, whether ABC will work with ecological IBMs is still uncertain.

Rejection-ABC was applied to an existing 14-parameter earthworm energy budget IBM for which the available data consist of body mass growth and cocoon production in four experiments. ABC was able to narrow the posterior distributions of seven parameters, estimating credible intervals for each. ABC's accepted values produced slightly better fits than literature values do. The accuracy of the analysis was assessed using cross-validation and coverage, currently the best-available tests. Of the seven unnarrowed parameters, ABC revealed that three were correlated with other parameters, while the remaining four were found to be not estimable given the data available.

It is often desirable to compare models to see whether all component modules are necessary. Here, we used ABC model selection to compare the full model with a simplified version which removed the earthworm's movement and much of the energy budget. We are able to show that inclusion of the energy budget is necessary for a good fit to the data. We show how our methodology can inform future modelling cycles, and briefly discuss how more advanced versions of ABC may be applicable to IBMs. We conclude that ABC has the potential to represent uncertainty in model structure, parameters and predictions, and to embed the often complex process of optimising an IBM's structure and parameters within an established statistical framework, thereby making the process more transparent and objective.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Animal populations consist of autonomous, adaptive individuals, all figuring out their own ways of achieving their goals. From these activities of individuals emerge population consequences, such as spatial distributions, social structures and population dynamics. For many questions, both theoretical and applied, scientific knowledge exists at the level of the individuals or the population, but not both; in these cases, individual or agent-based models, here referred to as IBMs, can bridge the gap (DeAngelis and Mooij, 2005). In IBMs, the interactions among individuals and between them and their surroundings are explicitly simulated, and all individuals have their own characteristics and make their decisions accordingly. This makes IBMs uniquely suited to exploring the effects of individual decisions on collective behaviour and to predicting how populations will change across time and space (Grimm and Railsback, 2005).

Examples of theoretical questions studied with IBMs include how ants build their nests (Buhl et al., 2005), how starlings coordinate their flocks (Hildenbrandt et al., 2010) and how macaques establish their relationships (Hemelrijk and Puga-Gonzalez, 2012). In practical applications, IBMs have been used to estimate the impact of development on coastal birds (Stillman and Goss-Custard, 2010), to aid in the management of fish stocks (Hartman and Kitchell, 2008), and to assess the effects of pesticides on non-target organisms (Schmolke et al., 2010). In all these cases, model

processes are fit to some data. However, due to the inherent complexity of IBMs, this process is often complicated, and the resulting outcome is often difficult to evaluate (Augusiak et al., 2014).

Currently, most IBMs are, implicitly or explicitly, built and evaluated using 'pattern-oriented modelling' (POM). This approach is essentially a protocol to be followed (Grimm and Railsback, 2012; Grimm et al., 2005). It specifies how multiple patterns observed in the real world should be used to iteratively design, select and parameterise IBMs, with each pattern serving as a 'filter' that rejects unsuitable model versions or parameterisations. Although the method has worked well in practice, a future goal is ideally to embed IBMs within mainstream statistical modelling and prediction. The Bayesian framework offers a comprehensive and well-trodden approach to this, but standard Monte Carlo methods for Bayesian inference are computationally intractable for all but the most carefully structured models (Keith and Spring, 2013). Conversely, 'Approximate Bayesian Computation', or ABC, enables approximate Bayesian inference for models of almost arbitrary complexity (Beaumont, 2010; Csilléry et al., 2010).

One key advantage of ABC, compared with other Bayesian methods, is that it is not necessary to analytically express how the likelihood of the data depends on the model parameters. Instead, ABC approximates these likelihoods by running models a large number of times, with parameters drawn randomly from their prior distributions, and then retaining the simulations closest to the observations. In this way, ABC provides a systematic way of assessing the support that different model versions and parameterisations receive from the available data, given some prior beliefs about how likely they are. Thus, for individual-based modellers, ABC has the potential to complement POM by making its rejections of unsuitable model versions and parameterisations more transparent and statistically rigorous. In addition, because ABC approximates full posterior parameter distributions, it provides a concise overview of the uncertainty in a model's parameter values, which can then be propagated into a model's predictions. Especially for IBMs that are then used in practical ecological decision making, this is an important feature.

Although the potential benefit of using ABC with ecological IBMs has been noted (Sibly et al., 2013; Thiele et al., 2014; Topping et al., 2012), whether it will work in practice is still uncertain. The original development of ABC was within population genetics (Beaumont et al., 2002; Pritchard et al., 1999; Tavaré et al., 1997), and in its basic form, rejection-ABC, it has yet to be applied to an IBM with more than two parameters (Sottoriva and Tavaré, 2010). In this paper, we apply rejection-ABC to a fairly typical ecological example, a 14-parameter IBM fitted to four existing experiments. In this IBM, Johnston et al. (2014) simulated the dynamic energy budgets of individual earthworms as they forage, grow and reproduce. We use rejection-ABC both to parameterise the model and to compare it to a simpler, possibly better model. In this way, we aim to introduce ABC to a wider audience, and to show that even simple implementations of ABC can provide surprising insights. In particular, we demonstrate how standard elements of an ABC analysis can be used to inform future modelling cycles. We consider the potential of more advanced versions of ABC in the context of the specific challenges posed by IBMs in the Discussion. A gentle introduction to the use of ABC with ecological IBMs together with a primer on building energy budget models from first principles will be available in van der Vaart et al. (submitted).

## 2. Material and methods

The simplest version of ABC, rejection-ABC, originally described by Pritchard et al. (1999), can be summarised as follows (Csilléry et al., 2010): First, for each of the parameters of each model, a

reasonable prior distribution is chosen. Statistically, for simplicity, we assume that each parameter has an independent prior distribution. Then, parameter values are sampled from these prior distributions, a large number of times, and the model is run for each of these samples, yielding some output. This output is compared with the empirical data. Some number of the runs that give the output closest to the empirical data are then 'accepted' as being 'close enough'. The accepted simulations now provide a sample of the posterior distributions of the model's parameters given the data. The same simulations can also be used for model comparison; in this case, the ratio in which different models are retained gives the relative probability that each model is correct. In the rest of this section, we first describe the empirical data available and the earthworm IBM; then, we give a detailed description of the ABC procedures used. All simulation results, the earthworm IBM and the ABC code have been deposited in a figshare repository (van der Vaart et al., 2015a,b), along with a brief guide to their use.

### 2.1. The empirical data

For our ABC analyses, we used the same empirical data that Johnston et al. (2014) originally used to assess the model's fit. This empirical data consists of the growth and reproduction data for *Eisenia fetida* earthworms in different laboratory setups (Gunadi et al., 2002; Gunadi and Edwards, 2003; Reinecke and Viljoen, 1990). In each case, five to ten earthworms were placed in small containers and supplied with cattle manure for food, under various feeding schedules (see Fig. 4 and Table S1). The earthworms were weighed and all cocoons were removed and counted at regular intervals. These procedures were replicated in our simulations; we assumed that the weighing of earthworms entailed a randomisation of their position, but no homogenisation of the substrate. The mean individual body masses and total cocoon numbers so obtained are referred to as the summary statistics. In total, 160 summary statistics were used.

### 2.2. The individual-based model

For parameter estimation, we used Johnston et al.'s (2014) IBM of the earthworm *Eisenia fetida*. For model selection, this IBM was compared with a simplified version of itself, which is described in Section 3.2. The model is implemented in NetLogo, a programming platform designed specifically for IBMs (Wilensky, 1999). Using the IBM to simulate all the available empirical data takes approximately half a second on a 3.4 GHz i7 iMac. Each time step, the earthworms in the IBM forage independently and allocate the acquired energy to maintenance, growth and cocoon production in a fixed order of priority (Fig. 1a). These priorities are represented by algorithms and equations derived from fundamental physiological ecology (Sibly et al., 2013). Each time step corresponds to a day. The model's parameters are in Table 1.

*Eisenia fetida* is predominantly surface-dwelling, and its environment becomes patchy when foraging earthworms deplete the food in their vicinity. To approximate this situation, we mapped the environment in two dimensions and divided it into four rectilinear patches, with food being homogeneously distributed within each patch and variation between patches. The total area simulated in this manner is 0.0144 m (see the Supplementary Information). Whenever the experimental setup calls for food to be added or removed, the amount is distributed equally across all four patches.

Each earthworm starts the simulation as a juvenile, with mass $M_b$, until it reaches mass $M_p$, when it becomes an adult. Every daily time step, it moves around randomly, with speed $s$, and ingests food. This movement is the only stochastic aspect of the model. An earthworm acquires energy according to Eq. (1), where $X$ is the food density and $M$ is the earthworm's current mass. This is
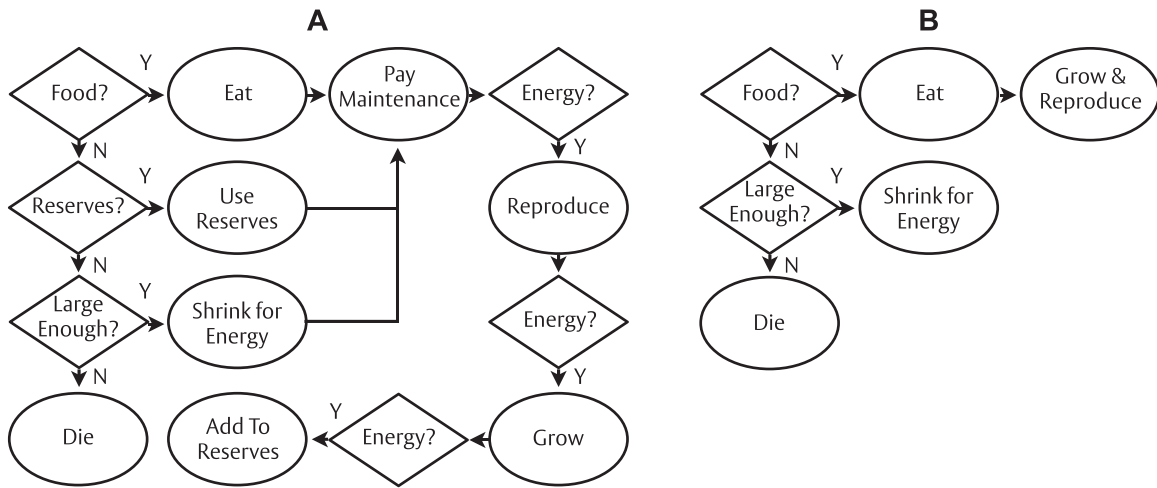
**Fig. 1.** Energy flow diagram for adult earthworms. Based on Fig. 2 in Johnston et al. (2014). (a) Full model (Section 2.2); (b) Simplified model (Section 3.2).

multiplied by the exponential Arrhenius function, which captures the temperature dependence of reactions. Here, $\kappa$ is Boltzmann's constant ($8.62 \times 10^{-5}$ eV K$^{-1}$), $T$ is the temperature in kelvins, and $T_{ref}$ is a reference temperature of 298.15 K.

$$\text{energy acquired} = IG_m \frac{X}{X+h} M^{2/3} E_f \, e^{-(E/k)(1/T-1/T_{ref})} \quad (1)$$

Each daily time step, the energy acquired by an earthworm is first used to pay basal metabolic costs, as given by Eq. (2). Then, in juveniles, energy is allocated to growth up to the maximum given by Eq. (3). Each gram of mass requires $E_c + E_s$ kilojoules of energy to synthesise. If that is not available, growth is reduced accordingly. In adults, energy is allocated to reproduction before growth, up to the maximum given by Eq. (4). A cocoon is produced as soon as $M_c$ * ($E_c + E_s$) energy has been allocated in this way.

$$\text{energy spent on maintenance} = B_0 M^{3/4} E_f \, e^{E/k} \quad (2)$$

$$\text{maximum mass growth} = r_B \, e^{(-E/k)(1/T-1/T_{ref})}(M_m^{1/3} M^{2/3} - M) \quad (3)$$

$$\text{maximum energy to reproduction} = r_m \, e^{(-E/k)(1/T-1/T_{ref})} M \quad (4)$$

If any energy remains, it is stored as glycogen in the individual's reserves. A gram of glycogen costs $E_s$ kilojoules to store, and an earthworm's maximum storage capacity is $1/2 \, M \, E_c$. If an earthworm has insufficient energy available to pay basal metabolic costs, stored energy is used to do so. If its reserves are at least $1/4 \, M \, E_c$, stored energy is also allocated to reproduction (but not growth). If an earthworm's reserves fall even lower, it is starving, and its mass declines. If an earthworm falls below its birth weight $M_b$, it dies.

**Table 1**
Model parameters. Literature values were derived in Johnston et al. (2014).

| Symbol | Definition | Literature value | Unit |
|---|---|---|---|
| $B_o$ | Taxon-specific normalisation constant | 967 | kJ/day |
| $E$ | Activation energy | 0.25 | eV |
| $E_c$ | Energy cost of tissue | 3.6 | kJ/g |
| $E_f$ | Energy from food | 10.6 | kJ/g |
| $E_s$ | Energy cost of synthesis | 3.6 | kJ/g |
| $h$ | Half saturation coefficient | 3.5 | g/0.01 g |
| $IG_m$ | Maximum ingestion rate | 0.70 | g/day/g |
| $M_b$ | Mass at birth | 0.011 | g |
| $M_c$ | Mass of cocoon | 0.015 | g |
| $M_m$ | Maximum asymptotic mass | 0.5 | g |
| $M_p$ | Mass at sexual maturity | 0.25 | g |
| $r_B$ | Growth constant | 0.177 | day$^{-1}$ |
| $r_m$ | Maximum energy to reproduction | 0.182 | kJ/g/day |
| $s$ | Movement speed | 0.004 | m/day |

### 2.3. The ABC analyses

For parameter estimation, we simulated the earthworm IBM one million times. We drew parameter values randomly from lognormal priors with means equal to Johnston et al.'s (2014) literature values (Table 1) and standard deviations equal to 0.3536. This produces samples where 95% of the values lie between half and twice the literature values on the unlogged scale. The literature values themselves were estimated from various sources; it should be noted that a few were derived directly from Gunadi et al. (2002), our Experiment 1.

Simulations were run in parallel on ARCHER, the UK's national supercomputing service. Parallelisation was achieved using a $C$ programme that called the earthworm IBM through NetLogo's controlling interface. All further analysis was done in $R$ (R Core Team, 2014). The $R$ package RNetLogo was used to control NetLogo from $R$ (Thiele et al., 2012).

The steps of our ABC analysis are given in Fig. 2; overall, our code was inspired by Csillery et al.'s (2012) $R$ package *abc*. In Step 2, the data points are the 160 measurements of mean body mass and summed cocoons that are shown in Fig. 4. This means that we are fitting the earthworm IBM to all four experiments simultaneously.

In Step 3, the distance $\rho$ between the model output of run $i$ and the empirical data points $D$ was computed according to Eq. (5), where $m_{i,j}$ is run $i$'s output for data point $j$, $D_j$ is the empirical data for data point $j$, and sd ($m_j$) is the standard deviation of the model outputs for data point $j$ in all model runs. Here, sd ($m_j$) is a scaling factor used to normalise the scales of the various data points; for instance, our mean body masses are in the hundredths, while our summed cocoons are in the dozens. If the differences between the model outputs and the empirical data were not appropriately scaled, the distance calculations would be dominated by the cocoon totals, because of the choice of units used to measure them.

$$\rho(m_i, D) = \sqrt{\sum_j \left( \frac{m_{i,j} - D_j}{\text{sd}(m_j)} \right)^2} \quad (5)$$

In the *abc* package, the scaling factor is not sd ($m_j$) but mad ($m_j$), the median absolute deviation from the median of summary statistic $j$ in all model runs. Csillery et al. (2012) chose mad ($m_j$) as a scaling factor because it is very robust to outliers. Unfortunately, for discrete data, such as cocoon totals, it is possible for the mad to be zero, leading to an undefined distance, as occurs in our case. To circumvent this problem, we used sd ($m_j$) following Beaumont et al., (2002).

1. Do $10^6$ simulations, using random samples from the model's priors.

2. Scale each data point in both the model outputs and the empirical data by dividing by the standard deviation of that data point in all model outputs.

3. Compute the distance $\rho$ between the scaled model outputs and the scaled empirical data according to Equation 5.

4. Accept the 100 runs which minimize the distance $\rho$.

**Fig. 2.** Steps of the ABC parameter estimation.

In Step 4, we accepted the 100 model runs with the lowest distance to the empirical data as 'close enough', equivalent to an acceptance rate of 0.0001. We chose 100 as a compromise between the need to include only good runs that match the data well and the need to have sufficient accepted runs to produce posterior distributions. However, our posterior estimates do not change much if we use an acceptance rate of 0.001 instead; see the Supplementary Material. We report marginal posterior distributions for each parameter; that is, the distribution of each parameter irrespective of the values of other parameters.

After performing parameter estimation, we were motivated to design a simplified version of the earthworm IBM, which is described in Section 3.2. We then compared this simplified version with the original design using ABC. To do this, we simulated the simplified model one million times also, resulting in two million runs total. We again used an acceptance rate of 0.0001, accepting 200 runs across the two model versions. All prior distributions were kept the same as for the full model, with the exception of $IG_m$, which was now centred around 0.15, as this produced a better fit. To perform Bayesian model selection, we computed approximate Bayes factors for the models being compared. A Bayes factor $B_{x,y}$ expresses the degree to which the available empirical data favours model $x$ over model $y$. In the context of ABC, assuming that each model has identical prior probability, it is calculated as the ratio of acceptances of model $x$ to model $y$ (Eq. (6)).

$$B_{x,y} = \frac{\text{acceptances of model } x}{\text{acceptances of model } y} \qquad (6)$$

To assess a simulation's fit of the empirical data, we calculated $R^2$, i.e. the proportion of variance explained, for each experiment; see Eq. (7). Here, $\bar{D}$ is the mean of the empirical data in that experiment.

$$R^2 = 1 - \frac{\sum_j (m_j - D_j)^2}{\sum_j (D_j - \bar{D})^2} \qquad (7)$$

## 3. Results

### 3.1. Parameter estimation

Posterior distributions for the model's parameters as estimated by ABC are plotted in Fig. 3. Given that there is Monte-Carlo sampling error in these estimates, we were motivated to use a statistical test as a criterion for deciding whether there was a difference between the posteriors and the priors. The marginal posteriors for seven parameters were significantly narrower than their priors (Levene's test, $p < 0.01$ after correcting for multiple testing using Holm's method); the rest were not. Most strongly constrained were $E$, the activation energy, $M_m$, the maximum mass, $M_p$ and $r_B$, the growth constant. All narrowed parameters had posteriors centred close to their original literature values, but $IG_m$, the maximum ingestion rate, was estimated to be slightly higher, while $r_B$ and $r_m$, the maximum energy to reproduction, were estimated to be slightly lower.

Fig. 4 shows how well the IBM fitted the empirical data when parameterised either with the literature values or with the parameter values accepted by ABC. Such visual checks are part of 'posterior predictive checking' (Gelman et al., 2003). No parameterisation of the model closely reproduced the empirical data from all experiments simultaneously. Comparing ABC's mean $R^2$ values to the mean $R^2$ values produced by running the IBM at its literature values 100 times, ABC's values produced better fits for Experiment 2 (Panel D), the cocoon totals of Experiment 3 (Panel E), and the mean masses of Experiment 4 (Panel C); conversely, the literature values did better for the remaining measurements.

Averaged over all six data sets in Fig. 4, the mean $R^2$ values of the 100 runs with ABC's best values (0.34) were higher than those with the literature values (0.22). Looking more qualitatively, the two parameterisations produced similar fits, although the literature values caused the cocoon totals in Experiment 4 (Panel F) to track the food availability more closely. Also, some patterns were not adequately captured by either set of parameters, most notably in Experiment 3, where the variability in earthworm mass (Panel B) and the timing of the final cocoon increase (Panel E) were never
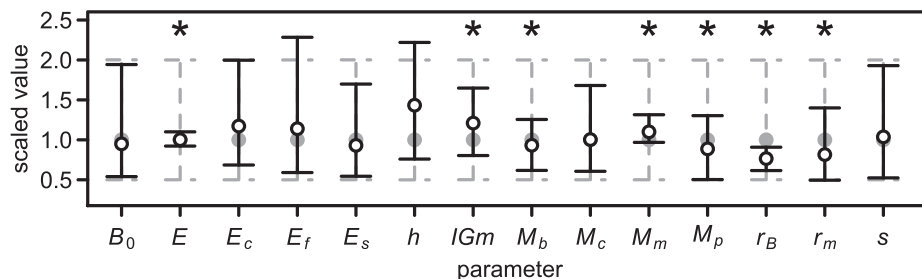
**Fig. 3.** Distributions of parameter values. Grey lines show the priors; black lines the posteriors. Circles represent medians, whiskers 95% credible intervals. Asterisks mark significant narrowing. All parameter values were scaled by dividing by the corresponding literature values, so that a value of 1 equals Johnston et al. (2014).
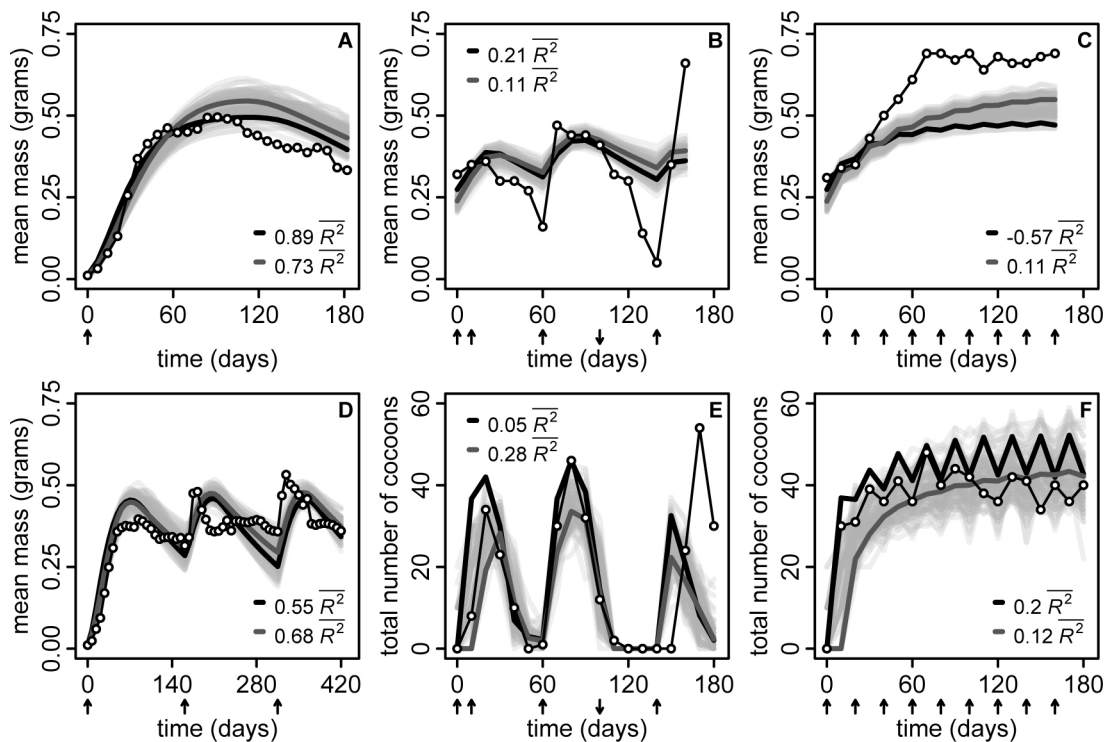
**Fig. 4.** Fits of the model to the empirical data. The thick black and grey lines are the result of averaging 100 runs each, using Johnston et al. (2014) literature values and ABC's best values, respectively, where "ABC's best values" refers to the parameters belonging to the best-fitting run. The open circles are the empirical data, and the semi-transparent grey lines are the 'posterior predictive check', i.e. the output of 100 new simulations using random samples from the accepted runs. Arrows mark the days when food was either added (↑) or removed (↓). (a) Experiment 1 (Gunadi et al., 2002); (d) Experiment 2 (Gunadi and Edwards, 2003); (b, e) Experiment 3 (Reinecke and Viljoen, 1990, 'variable condition'); (c, f) Experiment 4 (Reinecke and Viljoen, 1990, 'constant condition'). $R^2$ is the proportion of variance explained, a measure of goodness of fit (Eq. (7)).

fit well; the same holds for the maximum size achieved by the earthworms in Experiment 4 (Panel C).

To check the accuracy of ABC's estimates, we performed two kinds of quality control: Cross-validation (Csillery et al., 2012) and coverage (Prangle et al., 2013). Both methods use some subset of model outputs as "pseudo-data" and then use the remaining runs to do ABC. Because the parameter values that produced the "pseudo-data" are known, this makes it possible to check whether ABC is accurately estimating them. There are two main differences between cross-validation and coverage: Firstly; cross-validation uses randomly selected model outputs as "pseudo-data" while following Prangle et al. (2013) coverage uses the 100 'best' runs; i.e. those with the smallest error according to Eq. (5); secondly, cross-validation summarises ABC's posteriors by a point estimate–in our case, the median accepted value–while coverage looks at the accuracy of the posteriors as a whole.

To carry out cross-validation, we set aside 100 random outputs of the model, and then performed ABC using the remaining runs (see also Fig. S4). The medians of the accepted runs resulting from this procedure were mostly accurate, as shown in Fig. 5. In particular, all parameters with narrowed posteriors (most strongly $M_b$, $M_m$, and $r_B$, but also $E$, $IG_m$, $M_p$, and $r_m$; Fig. 3) were also adequately estimated in this cross-validation. Conversely, four of the estimates of the parameters with posteriors that were not narrowed did not correlate with their true values at all ($B_0$, $E_f$, $E_s$ and $s$). This means that ABC could not identify their values even when it was using 'pseudo-empirical data' generated by the model itself. That leaves three parameters ($E_c$, $h$ and $M_c$) which were estimated reasonably well from the simulated data but not from the empirical data.

These three parameters, $E_c$, $h$ and $M_c$, were found to correlate significantly with other parameters in ABC's accepted runs (Fig. 6). $E_c$ and $r_m$ correlated with each other, while $r_m$ also correlated with

$M_c$; these three parameters all regulate reproduction. Also, $h$, the saturation coefficient, correlated with $IG_m$, the maximum ingestion rate; these parameters both affect energy intake. Thus, although the posteriors of $E_c$, $h$ and $M_c$ were not narrowed when considered independently, they were narrowed in relationship to each other.

To estimate coverage, following Prangle et al. (2013), we took the model outputs of the 100 best runs as "pseudo-data" and then used the remaining runs to generate posterior parameter distributions using ABC. For each posterior parameter distribution so obtained, we calculated the relative frequency, $p$, of accepted parameter values that were less than the value actually used in that run (see also Fig. S5). The histograms of these $p$-values are shown in Fig. 7. In most cases ($B_0$, $E$, $E_c$, $E_f$, $E_s$, $M_c$, $M_p$, $r_m$, and $s$) the distributions are uniform, indicating that the corresponding posteriors are accurately estimated; therefore, 'coverage' is said to hold. However, this does not mean that the empirical data was necessarily informative with respect to estimating these parameters; when the empirical data is uninformative, ABC will return the prior distributions, and these will produce uniform coverage plots. In our analyses $B_0$, $E_f$, $E_s$ and $s$ had uniform coverage (Fig. 6), were not narrowed from their priors (Fig. 2), and could not be estimated from the data (Fig. 4). We conclude that the data are uninformative for these four parameters.

In the coverage plots of $h$ and $IG_m$, there is an excess of $p$-values on the right-hand side in Fig. 7, indicating that ABC generally underestimates these parameters. The remaining non-uniform coverage plots ($M_b$, $M_m$, and $r_B$) show deficits of $p$-values in the tails of the histograms, indicating that the estimated posterior distributions are broader than they should be. In other words, testing based on these estimated posteriors will be conservative. Overall, we conclude that $h$ and $IG_m$ were underestimated, that the credible intervals of the other eight parameters for which ABC provided information were either accurately or conservatively estimated, and that the
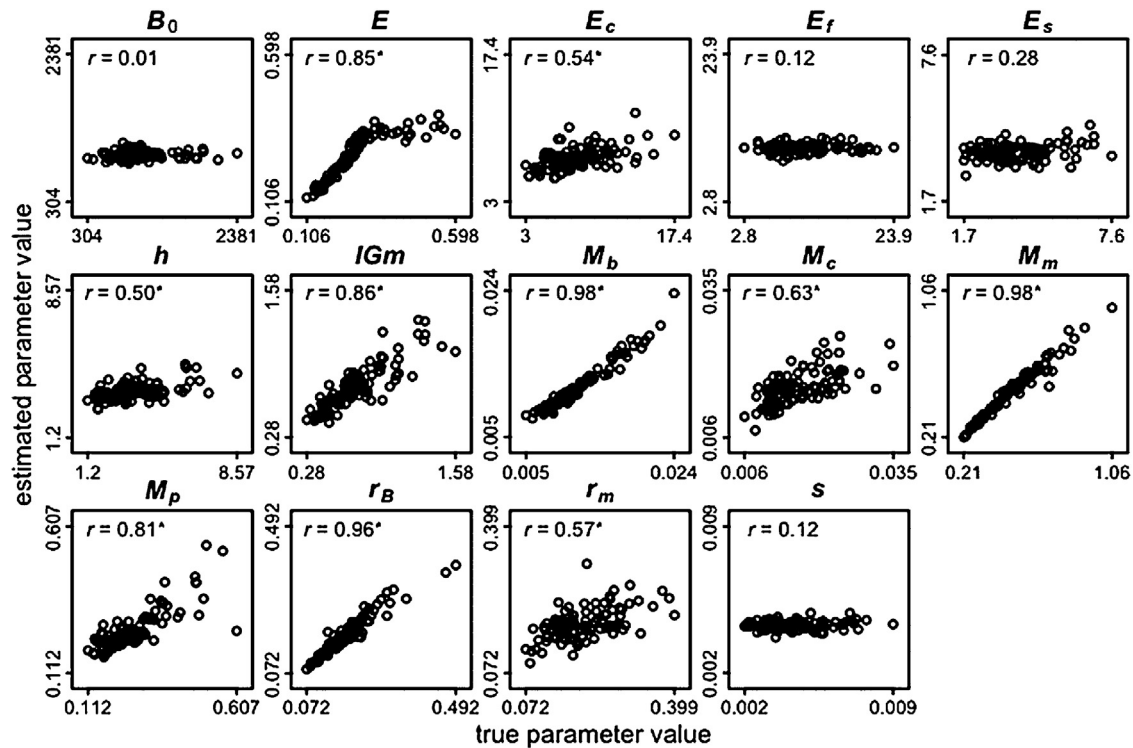
**Fig. 5.** Cross-validation for parameter estimation. Parameter values estimated by ABC in relation to true values for each of the model's 14 parameters. Correlation coefficients $r$ are shown at the top of each panel, asterisks denote significance ($p < 0.01$, corrected for multiple testing using Holm's method).
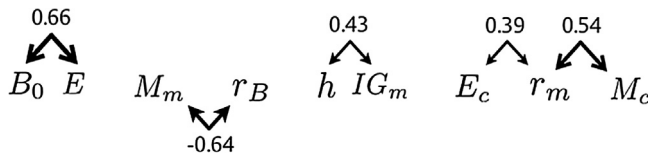


**Fig. 6.** Correlations between accepted parameter values. These were the only five correlations that were significant ($p < 0.01$, corrected for multiple testing using Holm's method). Arrow thickness is proportional to correlation strength.

remaining four ($B_0$, $E_f$, $E_s$ and $s$) could not be estimated using the available data.

### 3.2. Model selection

Of the earthworm IBM's 14 parameters, only seven were significantly narrowed by ABC. This raises the possibility that a structurally simpler model may fit the empirical data equally well. We investigated this possibility by constructing a simplified version of the earthworm IBM and then using ABC model selection to compare it with the original.

The energy flow of the simplified model is given in Fig. 1b. In the simplified model, the earthworms no longer store energy, and they grow (Eq. (3)) and reproduce (Eq. (4)) maximally every daily time step that they can eat. Thus, there is no possibility for energy shortfall; any ingested food is assumed to be enough. Only if there is no food at all do the earthworms start to starve; then they must catabolise mass (i.e. shrink) to cover their basal metabolic costs (Eq. (2)). This scheme eliminates the $E_f$ parameter, which determines the energy value of food. In addition, in the simplified model, the earthworms always eat as much as is determined by $IG_m$, the maximum ingestion rate. Thus, there are no density-dependent effects on foraging, removing $h$, the half-saturation coefficient. Finally, the earthworms in the simplified model do not move, eliminating the need for $s$, the parameter that controls the earthworms' speed.

**Table 2**
Bayes factors. The table shows how often the model in each row $x$ was accepted relative to the model in each column $y$, giving the Bayes factor B $x,y$ (Eq. (6)).

|                  | Full model | Simplified model |
|------------------|------------|------------------|
| Full model       | –          | 99               |
| Simplified model | 0.01       | –                |

Of the 200 runs accepted by ABC, 198 were of the full model, while two were of the simplified model. This produces the Bayes factors of Table 2. Thus, the simplified model was strongly rejected in favour of the full model, although the overall fit of the simplified model was reasonable (mean $R^2 = 0.25$ over all six data sets; see Fig. S2).

To analyse ABC's ability to correctly distinguish the full and simplified versions of the model, we again carried out 'cross-validation' (Csillery et al., 2012). To do this, we used 100 random runs of the full and simple models and investigated how well ABC could identify the underlying model versions using the remaining simulations. We summarised the outcome of each analysis by the model version assigned the highest posterior probability. As Fig. 8 shows, ABC was well able to distinguish the two models, with the full model confused for the simple model only once, and the reverse happening just six times.

### 4. Discussion

We have provided the first comprehensive demonstration of rejection-ABC's ability to effectively parameterise and evaluate an IBM using real empirical data. ABC was able to narrow the posterior distributions of seven out of 14 model parameters, estimating credible intervals for each. The accuracy of these posterior distributions was assessed using cross-validation and coverage, currently the best available tests. Both diagnostics showed that all of our narrowed parameters were likely to be accurately or conservatively
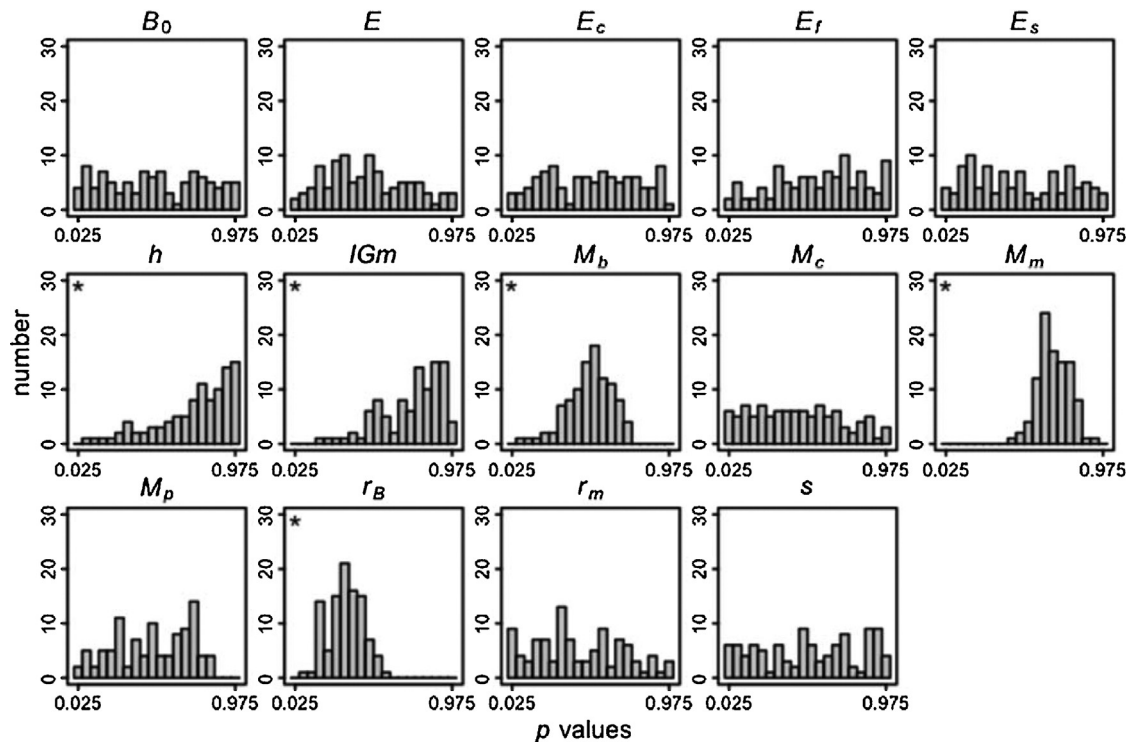
**Fig. 7.** Coverage for parameter estimation. Relative frequency $p$ of accepted parameter values that were less than the true value in ABC analyses using the 100 'best runs' as "pseudo-data". Asterisk mark significant departures from uniformity (Kolmogorov–Smirnov test, $p < 0.01$, corrected for multiple testing using Holm's method).
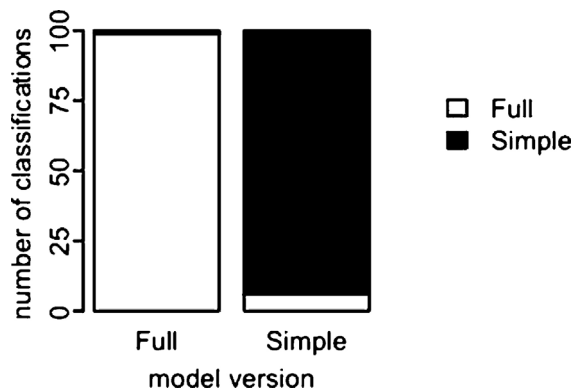


**Fig. 8.** Cross-validation for model selection. The $y$-axis shows how often each model version was classified by ABC as the full or simplified model, respectively.

estimated, with the exception of $IG_m$ and $h$, for which coverage suggested true values may be higher. As measured by the proportion of variance explained, ABC's accepted values provided slightly better fits than the literature values themselves. ABC also revealed that of the seven parameters that were not narrowed, three were correlated with other parameters, while the remaining four could not be estimated using the empirical data available.

The correlations between ABC's posterior parameter values suggest that some of the model's derived parameters can be narrowed if others that can be measured directly are measured more precisely. For example, within the accepted runs, the maximum energy to reproduction, $r_m$, correlates with the weight of cocoons, $M_c$, and the growth constant, $r_B$, correlates with the mass at sexual maturity, $M_p$. This is a useful insight, as cocoons and sexually mature earthworms are easy to weigh, but energy allocations and growth constants are derived from other measurements. This illustrates how ABC can guide empirical data collection; however, ABC can also assist in future model development.

As an example of how ABC can assist in model development we note that $E_f$, the energy value of food, was not narrowed in our study (Fig. 3) and could not be estimated even when using "pseudo-data" generated by the model itself (Fig. 6). This suggests that $E_f$ affects model outputs only when growth and reproduction are limited, but not halted, by energy shortfall, and this occurs in only 2% of daily time steps. Accordingly, we built a simplified model (Fig. 1b), where growth and reproduction occurred maximally or not at all, depending on whether the earthworms had any food to eat. However, ABC model selection robustly rejected this simplified model in favour of the full version (Table 2).

A particularly attractive feature of model selection using Bayes Factors is that it automatically accounts for differences in model complexity (Beaumont, 2010). This is because if the same number of simulations is done for all models, more parameters mean that a model's parameter space will be sampled less exhaustively. This results in fewer accepted simulations for more complex models unless the additional parameters add additional explanatory power. However, ABC model selection can be vulnerable to biases (Robert et al., 2011). This makes it important to check the accuracy of model selection; the cross-validation procedure (Fig. 8) shows how this can be done. Calculating coverage of model selection is possible in principle (Prangle et al., 2013); however, to produce meaningful insights, all models must be at least somewhat represented in the accepted runs, and that was not the case in our analysis.

Overall, we hope to have persuaded individual-based modellers that ABC (1) offers some advantages relative to current practice, and (2) is accessible enough for them to try it. We hope this will help bring inferential practices for IBMs in line with mainstream statistics. However, our study does have limitations, some of which may be solved by applying advanced methods already present in the literature (e.g. Hartig et al., 2011, for an overview), while others may require theoretical advancement specific to IBMs.

One limitation of our study is that we may not have sampled our priors sufficiently. With 14 parameters to calibrate, the odds

of obtaining a run where each parameter was sampled in the top quartile of its range are very small. Although our cross-validation and coverage analyses support the idea that our posterior distributions are sufficiently accurate to be useful, we cannot exclude the possibility that additional runs would have revealed additional narrowing. Two innovations already exist in the literature that may help alleviate this problem: Firstly, it is possible to sample a model's priors more efficiently. The core idea behind both MCMC-ABC (Markov Chain Monte Carlo, Marjoram et al., 2003) and SMC-ABC (Sequential Monte Carlo, Sisson et al., 2007; Toni et al., 2009) is that the output of individual runs can be used to gradually 'zoom in' on a model's posteriors, without wasting runs just sampling randomly. These methods have been used to successfully parameterise a number of IBMs (Martínez et al., 2011; Rasmussen and Hamilton, 2012) as well as related models (May et al., 2013; Scranton et al., 2014). However, it must be noted that both MCMC and SMC require simulations to be run sequentially, at least to some degree, and that this may make it difficult or impossible to parallellise them on large computing clusters. It also remains to be seen whether these methods will work well with IBMs that have strong dependencies between their parameters, as ours does; in such cases, it can be difficult to define how the 'zooming in' process should occur.

A second existing approach to minimising the problem of insufficient runs is 'post-sampling regression adjustment' (Beaumont et al., 2002; Blum and François, 2010), which is a family of techniques for more effectively using the runs already done. Such regression methods model the relationship between the accepted parameter values and the summary statistics, and then "correct" the accepted parameter values accordingly. Regression methods are powerful and may decrease uncertainty by a considerable margin, but they can give unreliable results if some observations are outside the range of values produced by the model, as happens here.

This lack of model fit – the fact that our IBM cannot reproduce the empirical data perfectly even when parameterised by ABC – is also a limitation of our study in its own right. This is a problem because the theory underlying ABC starts from the premise that likelihoods become asymptotically perfectly estimated as tolerance tends to zero, and by Bayes Theorem posteriors then become perfectly estimated too. One way forward in this situation is to amalgamate data into summary statistics that can in principle be perfectly fitted by the model. This would have the additional advantage of reducing the dimensionality of the data (Blum et al., 2013); with 160 data points (as shown in Fig. 4), the odds of obtaining a run that fits all 160 at once are very small. Indeed, the reason rejection-ABC works as well as it does may be that many of our data points are correlated. However, amalgamating data points just to improve model fit entails a loss of information and with others (e.g. Ratmann et al., 2009), we believe that scientifically more attractive ways forward should be sought. But this leaves us with a seemingly inevitable mismatch between model outputs and data. To cope with such situations, some theoreticians have proposed that mismatches between model and data be considered as errors, either observational measurement errors by the experimenter, or model errors arising from structural defects in the model. The suggestion is that such errors be modelled in such a way that the extended model can in principle fit the empirical data perfectly (Ratmann et al., 2009; Wilkinson, 2013). Bayes theorem can then be used to estimate the posteriors exactly. We are at present open-minded as to whether this procedure will significantly improve the estimation process for our IBM.

In conclusion, we believe that ABC facilitates model selection, parameterisation and uncertainty estimation, and can also provide input to future modelling cycles. In this way, it may be able to complement and advance POM, or 'pattern-oriented modelling', the current approach to designing, selecting and parameterising IBMs (Grimm and Railsback, 2012; Grimm et al., 2005; Jakoby et al., 2014). Importantly, ABC can also compare models of different complexity, as it automatically penalises models with many parameters (Beaumont, 2010). We believe that ABC has the potential to provide a statistical foundation for several of the steps of POM's verbal protocol, and make more transparent and objective the often complex process of optimising an IBM's structure and parameters.

## Acknowledgements

## Appendix A.  Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolmodel.2015.05.020

## References

Augusiak, J., van den Brink, P.J., Grimm, V., 2014. Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. Ecol. Model. 280, 117–128.

Beaumont, M.A., 2010. Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41, 379–406.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162, 2025–2035.

Blum, M.G.B., François, O., 2010. Non-linear regression models for approximate Bayesian computation. Stat. Comput. 20, 63–73.

Blum, M.G.B., Nunes, M.A., Prangle, D., Sisson, S.A., 2013. A comparative review of dimension reduction methods in approximate Bayesian computation. Stat. Sci. 28, 189–208.

Buhl, J., Deneubourg, J.-L., Grimal, A., Theraulaz, G., 2005. Self-organized digging activity in ant colonies. Behav. Ecol. Sociobiol. 58, 9–17.

Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., François, O., 2010. Approximate Bayesian computation (ABC) in practice. Trends Ecol. Evol. 25, 410–418.

Csillery, K., François, O., Blum, M.G.B., 2012. abc: an R package for approximate Bayesian computation (ABC). Meth. Ecol. Evol. 3, 475–479.

DeAngelis, D.L., Mooij, W.M., 2005. Individual-based modeling of ecological and evolutionary processes. Annu. Rev. Ecol. Evol. Syst. 36, 147–168.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2003. Bayesian Data Analysis. Chapman & Hall, London.

Grimm, V., Railsback, S.F., 2005. Individual-Based Modeling and Ecology. Princeton University Press, NJ.

Grimm, V., Railsback, S.F., 2012. Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. Phil. Trans. R. Soc. B 367, 298–310.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.-H., Weiner, J., Wiegand, T., DeAngelis, D.L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310, 987–991.

Gunadi, B., Blount, C., Edwards, C.A., 2002. The growth and fecundity of Eisenia fetida (Savigny) in cattle solids pre-composted for different periods. Pedobiologia 46, 15–23.

Gunadi, B., Edwards, C.A., 2003. The effects of multiple applications of different organic wastes on the growth, fecundity and survival of Eisenia fetida (Savigny) (Lumbricidae). Pedobiologia 47, 321–329.

Hartig, F., Calabrese, J.M., Reineking, B., Wiegand, T., Huth, A., 2011. Statistical inference for stochastic simulation models—theory and application. Ecol. Lett. 14, 816–827.

Hartman, K.J., Kitchell, J.F., 2008. Bioenergetics modeling: progress since the 1992 symposium. Trans. Am. Fish. Soc. 137, 216–223.

Hemelrijk, C.K., Puga-Gonzalez, I., 2012. An individual-oriented model on the emergence of support in fights, its reciprocation and exchange. PLoS One 7, e37271.

Hildenbrandt, H., Carere, C., Hemelrijk, C.K., 2010. Self-organized aerial displays of thousands of starlings: a model. Behav. Ecol. 21, 1349–1359.

Jakoby, O., Grimm, V., Frank, K., 2014. Pattern-oriented parametrization of general models for ecological application: towards realistic evaluations of management approaches. Ecol. Model. 275, 78–88.

Johnston, A.S.A., Hodson, M.E., Thorbek, P., Alvarez, T., Sibly, R.M., 2014. An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides. Ecol. Model. 280, 5–17.

Keith, J.M., Spring, D., 2013. Agent-based Bayesian approach to monitoring the progress of invasive species eradication programs. Proc. Natl. Acad. 110, 13428–13433.

Martínez, I., Wiegand, T., Camarero, J.J., Batllori, E., Gutiérrez, E., 2011. Disentangling the formation of contrasting tree-line physiognomies combining model selection and Bayesian parameterization for simulation models. The American Naturalist 177, E136–E152.

Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S., 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl Acad. Sci. USA 100, 15324-15238.

May, F., Giladi, I., Ristow, M., Ziv, Y., Jeltsch, F., 2013. Metacommunity, mainland-island system or island communities? Assessing the regional dynamics of plant communities in a fragmented landscape. Ecography 36, 1–12.

Prangle, D., Blum, M.G.B., Popovic, G., Sisson, S.A., 2013. Diagnostic tools for approximate Bayesian computation using the coverage property. Aust. NZ J. Stat. 56, 309–329.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., Feldman, M.W., 1999. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. Mol. Biol. Evol. 16, 1791–1798.

R Core Team, 2014. R: A Language and Environment for Statistical Computing, http://www.R-project.org, 3.1.1 ed. The R Foundation for Statistical Computing, Vienna, Austria.

Rasmussen, R., Hamilton, G., 2012. An approximate Bayesian computation approach for estimating parameters of complex environmental processes in a cellular automata. Environ. Model. Softw. 29, 1–10.

Ratmann, O., Andrieu, C., Wiuf, C., Richardson, D.M., 2009. Model criticism based on likelihood-free inference, with an application to protein network evolution. Proc. Natl. Acad. Sci. 106, 10576–10581.

Reinecke, A.J., Viljoen, S.A., 1990. The influence of feeding patterns on growth and reproduction of the vermicomposting earthworm *Eisenia fetida* (Oligochaeta). Biol. Fert. Soils 10, 184–187.

Robert, C.P., Cornuet, J.-M., Marin, J.-M., Pillai, N.S., 2011. Lack of confidence in approximate Bayesian computation model choice. Proc. Natl Acad. Sci. 108, 15112–15117.

Schmolke, A., Thorbek, P., Chapman, P., Grimm, V., 2010. Ecological models and pesticide risk assessment: current modeling practice. Environ. Toxicol. Chem. 29, 1006–1012.

Scranton, K., Knape, J., de Valpine, P., 2014. An approximate Bayesian computation approach to parameter estimation in a stochastic stage-structured population model. Ecology 9, 1418–1428.

Sibly, R.M., Grimm, V., Martin, B.T., Johnston, A.S.A., Kułakowska, K., Topping, C.J., Calow, P., Nabe-Nielsen, J., Thorbek, P., DeAngelis, D.L., 2013. Representing the acquisition and use of energy by individuals in agent-based models of animal populations. Methods Ecol. Evol. 4, 151–161.

Sisson, S.A., Fan, Y., Tanaka, M.A., 2007. Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America 104, 1760–1765.

Sottoriva, A., Tavaré, S., 2010. Integrating approximate Bayesian computation with complex agent-based models for cancer research. In: Saporta, G., Lechevallier, Y. (Eds.), COMPSTAT 2010–Proceedings in Computational Statistics. Physiva Verlag, Heidelberg, Germany, pp. 57–66.

Stillman, R.A., Goss-Custard, J.D., 2010. Individual-based ecology of coastal birds. Biol. Rev. 85, 413–434.

Tavaré, S., Balding, D.J., Griffiths, R.C., Donnelly, P., 1997. Inferring coalescence times from DNA sequence data. Genetics 145, 505–518.

Thiele, J.C., Kurth, W., Grimm, V., 2012. RNetLogo: an R package for running and exploring individual-based models implemented in NetLogo. Methods Ecol. Evol. 3, 480–483.

Thiele, J.C., Kurth, W., Grimm, V., 2014. Facilitating parameter estimation and sensitivity analysis of agent-based models: a cookbook using NetLogo and R. J. Artif. Soc. Soc. Sim. 17, 11.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.H., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface 6, 187–202.

Topping, C.J., Dalkvist, T., Grimm, V., 2012. Post-hoc pattern-oriented testing and tuning of an existing large model: lessons from the field vole. PLoS One 7, e45872.

van der Vaart, E., Beaumont, M.A., Johnston, A.S.A., Sibly, R.M. (2015a): ABC for IBMs - Earthworm Runs & Results of Checks. figshare. http://dx.doi.org/10.6084/m9.figshare.1412720

van der Vaart, E., Beaumont, M.A., Johnston, A.S.A., Sibly, R.M. (2015b): ABC for IBMs - Earthworm Models & ABC Code. figshare. http://dx.doi.org/10.6084/m9.figshare.1420526

van der Vaart, E., Johnston, A.S.A., Sibly, R.M., submitted. Predicting how many animals will be where: A recipe for building, calibrating and evaluating individual-based models. Ecological Modelling.

Wilensky, U., 1999. NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, http://ccl.northwestern.edu/netlogo/.

Wilkinson, R.D., 2013. Approximate Bayesian Computation (ABC) gives exact results under the assumption of model error. Stat. Appl. Genet. Mol. Biol. 12, 129–141.