REVIEW AND SYNTHESIS

Statistical inference for stochastic simulation models – theory and application

Florian Hartig,[1*] Justin M. Calabrese,[1,2] Björn Reineking,[3] Thorsten Wiegand[1] and Andreas Huth[1]

Abstract

Statistical models are the traditional choice to test scientific theories when observations, processes or boundary conditions are subject to stochasticity. Many important systems in ecology and biology, however, are difficult to capture with statistical models. Stochastic simulation models offer an alternative, but they were hitherto associated with a major disadvantage: their likelihood functions can usually not be calculated explicitly, and thus it is difficult to couple them to well-established statistical theory such as maximum likelihood and Bayesian statistics. A number of new methods, among them Approximate Bayesian Computing and Pattern-Oriented Modelling, bypass this limitation. These methods share three main principles: aggregation of simulated and observed data via summary statistics, likelihood approximation based on the summary statistics, and efficient sampling. We discuss principles as well as advantages and caveats of these methods, and demonstrate their potential for integrating stochastic simulation models into a unified framework for statistical modelling.

Keywords

Bayesian statistics, indirect inference, intractable likelihood, inverse modelling, likelihood approximation, likelihood-free inference, maximum likelihood, model selection, parameter estimation, stochastic simulation.

Ecology Letters (2011) **14**: 816–827

INTRODUCTION AND BACKGROUND

As ecologists and biologists, we try to find the laws that govern the functioning and the interactions among nature's living organisms. Nature, however, seldom presents itself to us as a deterministic system. Demographic stochasticity, movement and dispersal, variability of environmental factors, genetic variation and limits on observation accuracy are only some of the reasons. We have therefore learnt to accept stochasticity as an inherent part of ecological and biological systems and, as a discipline, we have acquired an impressive arsenal of statistical inference methods. These methods allow us to decide which among several competing hypotheses receives the most support from the data (model selection), to quantify the relative support within a range of possible parameter values (parameter estimation) and to calculate the resulting uncertainty in parameter estimates and model predictions (uncertainty estimation).

A limitation of most current statistical inference methodology is that it works only for models $M(\phi)$ with a particular property: given that we have observations D_{obs}, it must be possible to calculate $p(D_{obs}|\phi)$, the probability of obtaining the observed data, for each possible model parameterization ϕ. We will use the term *likelihood* synonymously with $p(D_{obs}|\phi)$ (see Box 1). On the basis of this probability, one can derive statistical methods for parameter estimation, model selection and uncertainty analysis (see Box 1).

For simple stochastic processes, the probability $p(D_{obs}|\phi)$ can be calculated directly. One refers to this property by saying that the process has a *tractable likelihood*. Practically all statistical models that are

used in ecology and biology make assumptions that result in tractable likelihoods. Regression models, for example, typically assume that the data were observed with independent observation errors that follow a
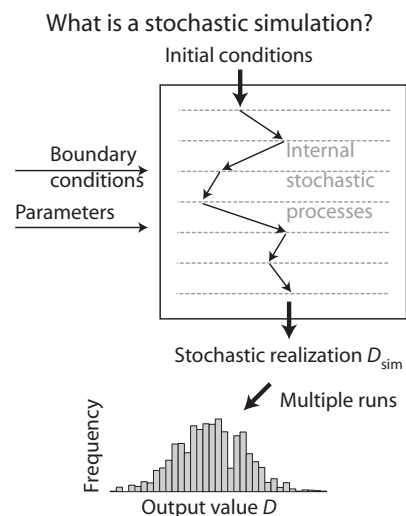

Figure 1 The term stochastic simulation refers to the method of drawing samples from a potentially complex stochastic process by explicitly sampling along the hierarchy of its dependent subprocesses. Thus, the output of a stochastic simulation emerges as the result of one realized random trajectory through the potential internal model states. Repeatedly simulating yields a frequency distribution for the possible outcomes.

[1]UFZ – Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany

[2]Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Rd., Front Royal, VA 22630, USA

[3]University of Bayreuth, Universitätsstrasse 30, 95440 Bayreuth, Germany

*Correspondence: E-mail: florian.hartig@ufz.de

fixed, specified distribution. As the errors are independent, $p(D_{obs}|\phi)$ simply separates into the probabilities of obtaining the individual data points, which greatly simplify the calculation. During this review, we will therefore use the term *statistical model* as a synonym for a stochastic model with a tractable likelihood.

In many relevant ecological or biological systems, however, multiple sources of heterogeneity interact and only parts of the system can be observed. Despite the progress that has been made in computational statistics to make likelihoods tractable for such interacting stochastic processes, for example, by means of data augmentation (Dempster *et al.* 1977), state-space models (Patterson *et al.* 2008), hierarchical Bayesian models (Wikle 2003; Clark & Gelfand 2006) or diffusion approximations (Holmes 2004), our ability to calculate likelihoods for complex stochastic systems is still severely constrained by mathematical difficulties. For this reason, *stochastic simulation models* (Fig. 1) are widely used in ecology and biology (Grimm & Railsback 2005; Wilkinson 2009).

A stochastic simulation is an algorithm that creates samples from a potentially complex stochastic process by explicitly sampling from all its sub-processes (Figs 1 and 2). This sampling allows researchers to model stochastic ecological processes exactly as they are known or

conjectured without having to concentrate on the mathematical tractability of the conditional probabilities that would need to be calculated to keep the probability $p(D_{obs}|\phi)$ tractable. Stochastic simulation models are therefore especially useful for describing processes where many entities develop and interact stochastically, for example, for population and community dynamics, including individual-based and agent-based models (Huth & Ditzer 2000; Grimm & Railsback 2005; Ruokolainen *et al.* 2009, Bridle *et al.* 2010), diversity patterns, neutral theory and evolution (Chave *et al.* 2002; Alonso *et al.* 2008; Arita & Vázquez-Domínguez 2008; de Aguiar *et al.* 2009), movement and dispersal models of animals and plants (Nathan *et al.* 2001; Couzin *et al.* 2005; Berkley *et al.* 2010), or for the simulations of cellular reactions in biological systems (Wilkinson 2009).

Hence, the crucial difference between a typical statistical model and a stochastic simulation model is not the model structure as such. Both are representations of a stochastic process. However, while typical statistical models allow the calculation of $p(D_{obs}|\phi)$ directly (tractable likelihood), stochastic simulation models produce random draws D_{sim} from the stochastic process by means of simulation (Fig. 1). This does not mean that the likelihood $p(D_{obs}|\phi)$ does not exist for a stochastic

---

Box 1 Maximum likelihood and Bayes in a nutshell

The key idea underlying both Bayesian inference and maximum likelihood estimation is that the support given to a parameter ϕ by the data D is proportional to $p(D|\phi)$, the probability that D would be observed given $M(\phi)$. In the words of Fisher (1922): 'the likelihood that any parameter (or set of parameters) should have any assigned value (or set of values) is proportional to the probability that if this were so, the totality of observations should be that observed'. The word proportional is crucial – both for Bayesian and likelihood-based inference, the value of $p(D|\phi)$ carries no absolute information about the support for a particular parameter, but is only used to compare parameters by their probabilities of producing the observed data given the model M.

Maximum likelihood estimation

The function that is obtained by viewing $p(D|\phi)$ as a function of the parameter ϕ is called the *likelihood function*.

$$\mathscr{L}(\phi) \propto p(D|\phi). \tag{1}$$

The method of *maximum likelihood estimation* is to search $\mathscr{L}(\phi)$ for its maximum and interpret this as the most likely value of the parameter ϕ. Usually, the maximum is determined by numerical optimization. A number of techniques exist to subsequently calculate confidence intervals from the curvature of $\mathscr{L}(\phi)$ around its maximum, or to test whether the likelihood value at the maximum likelihood estimate of a parameter is significantly different from a null hypothesis ϕ_0 where this parameter is kept at a fixed value (*likelihood ratio test*).

Bayesian statistics

From the definition of the likelihood function, it is only a small step to Bayesian Statistics. Bayes' formula states that the probability density $\mathscr{P}(\phi)$ that is associated with any parameter ϕ conditional on the data D is given by

$$\mathscr{P}(\phi) = p(\phi|D) = \frac{p(D|\phi) \cdot p(\phi)}{\int p(D|\phi) \cdot p(\phi)d\phi}. \tag{2}$$

$\mathscr{P}(\phi)$ is called the posterior distribution. It depends on the likelihood term $p(D|\phi)$, and additionally on a new term $p(\phi)$. $p(\phi)$ is called the prior, because it is interpreted as our prior belief about the parameter values, e.g. from previous measurements, before confronting the model with the data D. If there is no prior knowledge about the relative probability of the models or parameterizations to be compared, one may try to specify 'non-informative' priors that express this ignorance. There is a rich literature on how to choose non-informative (reference) priors, and we suggest Kass & Wasserman (1996) and Irony & Singpurwalla (1997) for further reading. It is worth noting one particularly important conclusion of this literature: despite widespread use in Bayesian applications, non-informative priors are by no means required to be flat (uniformly distributed).

Informally, one may think of a Bayesian analysis simply as a normalized likelihood function that was additionally multiplied by the prior. A fundamental difference to the likelihood approach, however, is that the posterior value $\mathscr{P}(\phi)$ is interpreted as a probability density function. Debates about this, also in comparison with likelihood approaches, have a long and noteworthy tradition in statistics (e.g Fisher 1922; Ellison 2004). However, discussing these arguments is beyond the scope of this article.
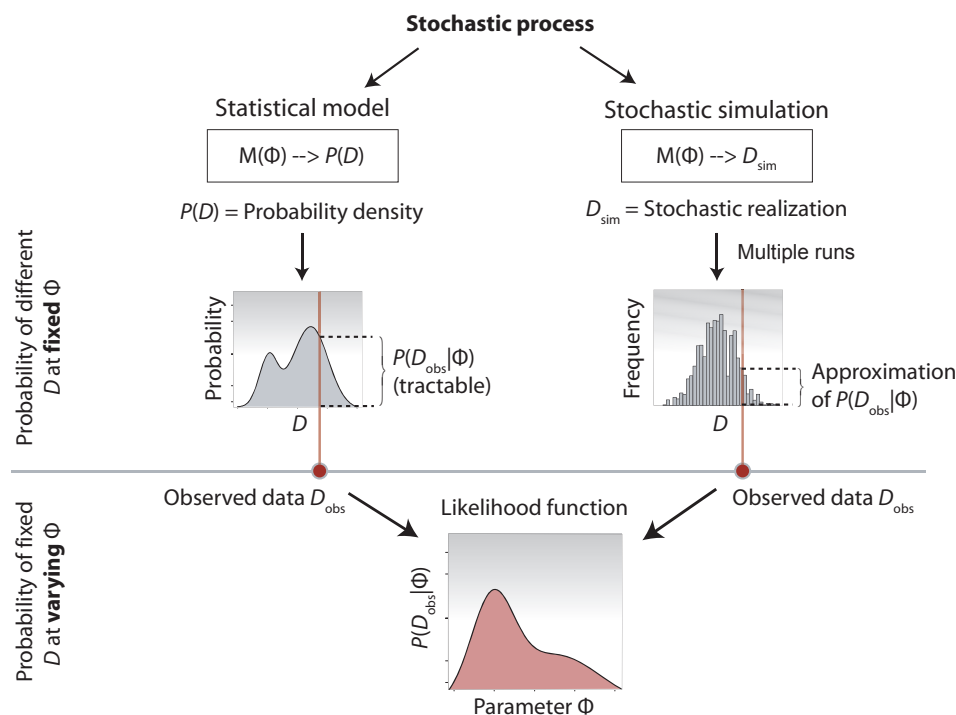
Stochastic process

Statistical model	Stochastic simulation
$M(\Phi) \dashrightarrow P(D)$	$M(\Phi) \dashrightarrow D_{sim}$

$P(D)$ = Probability density D_{sim} = Stochastic realization

Multiple runs

Probability of different *D* at **fixed** Φ

Probability of fixed *D* at **varying** Φ

$P(D_{obs}|\Phi)$ (tractable)

Approximation of $P(D_{obs}|\Phi)$

Observed data D_{obs} Likelihood function   Observed data D_{obs}

Parameter Φ

Figure 2 Likelihood-based inference for statistical vs. stochastic simulation models: The underlying natural phenomenon is a stochastic process. A statistical model (left) expresses model predictions as a function of a parameter ϕ in terms of probability density functions. Therefore, one may, for any ϕ, calculate $p(D_{obs}|\phi)$, the probability that is predicted by the model for observing the data D_{obs} (tractable likelihood). A stochastic simulation model (right) produces per model run only a single draw D_{sim} from an underlying probability density function that depends on the parameter ϕ, but cannot be explicitly calculated (intractable likelihood). In principle, however, the probability of obtaining D_{obs} at any fixed parameter ϕ may be approximated from the histogram of outcomes that is generated by repeated simulations. The result of this approximation may be used for subsequent inference in much the same way as tractable likelihoods are used in standard statistical models.

simulation model. As illustrated by Fig. 1, the histogram of many simulated outcomes D_{sim} will eventually converge to a fixed probability density function as the number of samples increases. For this reason, stochastic simulation models have also been termed 'implicit statistical models' (Diggle & Gratton 1984). In principle, it is therefore possible to estimate $p(D_{obs}|\phi)$ by drawing stochastic realizations from M until a sufficient certainty about the probability of obtaining D_{obs} is reached (see Fig. 2). Yet, while this is asymptotically exact, it is for most practical cases hopelessly inefficient for two reasons: (1) The predicted and observed data of most practically relevant models are high-dimensional (e.g. spatial data, phylogenetic trees, time series), and the individual dimensions are correlated, which means their likelihoods cannot be estimated independently. (2) For continuous variables, the probability of observing exactly the same outcome is infinitesimally small. Consequently, covering the output space of a stochastic simulation model with sufficient resolution to obtain reliable estimates of the likelihood is infeasible with such a straightforward approach. As a result, statistical parameter estimation and model selection techniques could hitherto not generally be applied to stochastic simulation models.

In recent years, however, a number of different strategies have been developed to address the problem of making stochastic simulation models usable for likelihood-based inference. Those include methods that explicitly approximate $p(D|\phi)$ such as Approximate Bayesian Computing (ABC) (Beaumont 2010; Csilléry *et al.* 2010), simulated (synthetic) pseudo-likelihoods (Hyrien *et al.* 2005; Wood 2010) or indirect inference (Gourieroux *et al.* 1993), and also other methods

that allow parameterizations without explicitly approximating $p(D|\phi)$, for example, informal likelihoods (Beven 2006) and Pattern-Oriented Modelling (POM; Wiegand *et al.* 2003, 2004b, Grimm *et al.* 2005). Despite different origins and little apparent overlap, most of these methods use the same three essential steps:

(1) The dimensionality of the data is reduced by calculating summary statistics of observed and simulated data.
(2) Based on these summary statistics, $p(D_{obs}|\phi)$, the likelihood of obtaining the observed data D_{obs} from the model M with parameters ϕ, is approximated.
(3) For the computationally intensive task of estimating the shape of the approximated likelihood as a function of the model parameters, state-of-the-art sampling and optimization techniques are applied.

These steps allow the linkage of stochastic simulation models to well-established statistical theory and therefore provide a general framework for parameter estimation, model selection and uncertainty estimation by comparison of model output and data (inverse modelling). In what follows, we structure and compare different strategies for finding summary statistics, approximating or constructing the likelihood, and exploring the shape of this likelihood to obtain parameter and uncertainty estimates. We hope that this collection of methods may not only serve as a toolbox from which different approaches can be selected and combined, but that it will also stimulate the exchange of ideas and methods across the communities that have developed different traditions of inverse modelling.
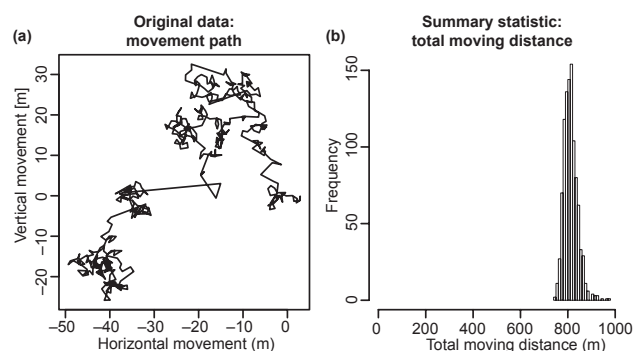
Figure 3 (a) An example of a summary statistic: a movement path of an individual (generated by 500 steps of a Lévy walk model). (b) One possible summary statistic: the total movement distance after 500 steps. The frequency distribution shows the outcome of 1000 independent simulation runs with the same parameter values.

SUMMARY STATISTICS – REDUCING THE STATE SPACE

The first step for comparing stochastic simulation models with observations is to reduce the dimensionality of simulated and observed data. Doing so is not a strict necessity, but in most cases, a practical requirement: imagine, for example, we had bird telemetry data (Fig. 3), and we had a stochastic model that describes the movement of these birds. As the possible spatial movement paths of such a model are virtually infinite, rerunning the model to calculate the probability of observing any particular movement path is practically impossible. Fortunately, however, it is often possible to compare model and data on a more aggregated level without losing information with respect to the inference. For example, if the unknown parameters of the movement model affect only the movement distance, we may probably aggregate model output and data by focusing only on patterns such as the total movement distance, which may greatly simplify the analysis (Fig. 3). Some other commonly used examples of aggregations are size class distributions of plants or animals (see, e.g. Dislich *et al.* 2009, for tree size distributions), movement patterns (Sims *et al.* 2008, for marine predator search behavior), spatial patterns that aggregate the relative positions of individuals or events (Zinck & Grimm 2008, for fire size distributions) or swarming patterns (Huth & Wissel 1992; Couzin *et al.* 2005).

Within this article, we use the term *summary statistic* for such an aggregation of model output and observed data. Other terms that are used in the literature are 'statistic', 'output variable', 'aggregate variable', 'intermediate statistic' (Jiang & Turnbull 2004), 'auxiliary parameter' (Gourieroux *et al.* 1993) or in the context of POM also 'pattern' (Grimm *et al.* 2005).

Sufficiency and the choice of summary statistics

The idea that data may often be reduced without losing information for the purpose of statistical inference is known as *sufficiency*: a summary statistic (or a set of summary statistics) is sufficient if it produces an aggregation of the data that contains the same information as the original data for the purpose of parameter estimation or model selection of a model or a set of models. A sufficient summary statistic that cannot be further simplified is called *minimally sufficient* (Pawitan 2001).

While sufficiency is fundamental to ensure the correctness of the inference, minimal sufficiency of the summary statistics is generally not.
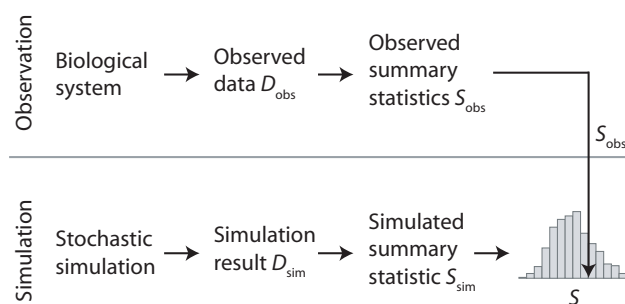

Figure 4 Illustration of the concept of comparing model and data through summary statistics.

Yet, many of the methods discussed in the next sections will work better and be more robust if the information in the summary statistics shows no unnecessary redundancies and correlations. Thus, our general aim is to find sufficient statistics that are as close to minimal sufficiency as possible. For standard statistical models, minimal sufficient statistics are often known. A classic example is the sample mean, which contains all information necessary to determine the mean of the normal model. Also, when symmetries are present (e.g. time translation invariance in Markov models or spatial isotropy), it may be obvious that a certain statistic can be applied without loss of information. Apart from these straightforward simplifications, a number of strategies for choosing summary statistics have been suggested.

One possibility is to compare the statistical moments (mean, variance, etc.) of observed and simulated data (the method of simulated moments; McFadden 1989). Similarly, but more flexible, the method of indirect inference uses as summary statistics the parameter estimates of a so called 'auxiliary', 'intermediate' or 'indirect' statistical model that is fit to simulated and observed data (Gourieroux *et al.* 1993; Heggland & Frigessi 2004; Jiang & Turnbull 2004; Drovandi *et al.* 2011). Wood (2010) gives some general hints for choosing summary statistics with a focus on separating the stationary from the dynamic aspects of temporal data. Wegmann *et al.* (2009) extract the most important components of a larger set of summary statistics by partial least square transformations. Joyce & Marjoram (2008) and Fearnhead & Prangle (2010) weight statistics by their importance for the inference. Wiegand *et al.* (2003, 2004b) and Grimm *et al.* (2005) stress the importance of combining summary statistics (patterns) that operate at different scales and hierarchical levels of the system as a good strategy to reach sufficiency.

In general, however, we will have to test (usually with artificially created data, see, e.g. Jabot & Chave 2009; Zurell *et al.* 2009) whether a statistic is sufficient with respect to a particular inferential task, and whether it can be further simplified. Particularly for more complex models, we may also decide to use summary statistics that are only close to sufficient, in return for a simpler description of the data. To a certain extent, it therefore depends on the experience and intuition of the scientist to find summary statistics that are as close to sufficiency as possible and at the same time simple enough to allow for efficient fits.

LIKELIHOOD APPROXIMATIONS FOR A SINGLE PARAMETER VALUE – THE GOODNESS-OF-FIT

In the previous section, we have discussed how to derive summary statistics that aggregate model output and observed data. Aggregation, however, does not mean that the simulated summary statistics do not
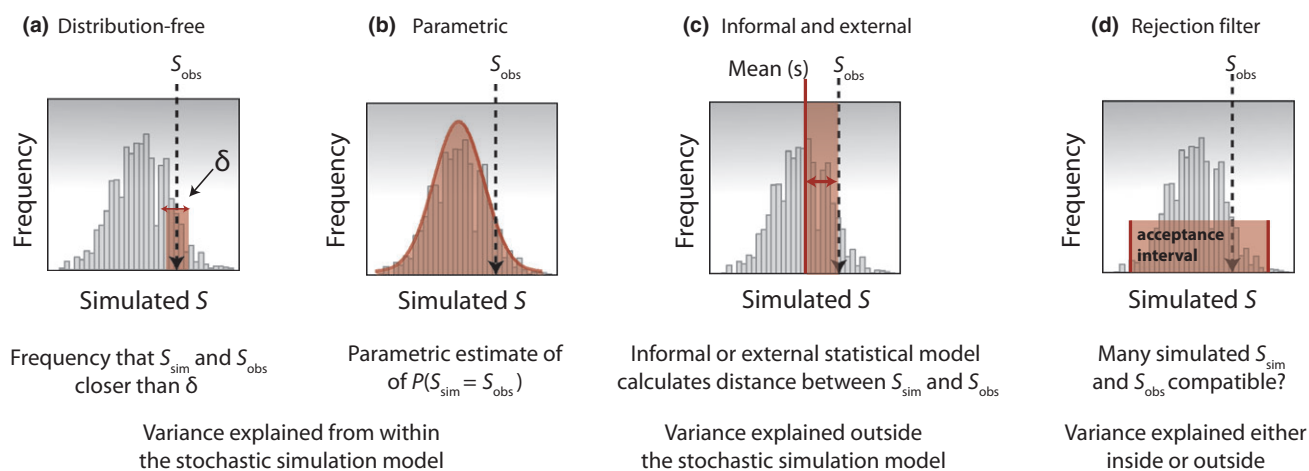
Figure 5 Point-wise likelihood approximation: assume that we simulate from a stochastic simulation model with fixed parameters ϕ. These simulations yield a set of simulated summary statistics, represented by the grey frequency distributions in the figure. Our aim is to estimate the probability for obtaining S_{obs}, the summary statistic of the observed data, from the simulation. (a) Nonparametric approximations locally estimate the probability density of many simulated S_{sim} at the value of the observed summary statistics S_{obs}. (b) Parametric approximations use a parametric model to approximate the distribution of many simulated S_{sim}, and then estimate the probability of obtaining S_{obs} from this model. (c) Instead of approximating $p(S_{obs}|\phi)$ directly, one may also use an external error model or an informal likelihood to compare the simulation output with the data. (d) Rejection filter use the simulated S_{sim} to determine whether S_{obs} is likely to have been generated by the tested model or parameterization.

vary at all (see, e.g. Fig. 3). A sufficient statistic, by definition, only averages out that part of the simulation variability that is irrelevant for the inference. Instead of estimating the likelihood $p(D_{obs}|\phi)$ of obtaining the observed data from the model, one can therefore work with $p(S_{obs}|\phi)$, the probability of simulating the same summary statistics as observed (Fig. 4). In what follows, we will discuss different methods to conduct statistical inference based on comparing S_{obs} with a number of simulated summary statistics (Fig. 4). As we noted before, the use of summary statistics is usually a computational necessity, but may not be essential: all the methods that we discuss in what follows could, in principle, also be applied to compare the original data with simulation outputs under a given model.

Nonparametric likelihood approximations

A brute force approach to compare S_{obs} with the model at a fixed parameter combination ϕ would be simply to create more and more simulation results until sufficient certainty about the probability $p(S_{obs}|\phi)$ of obtaining exactly S_{obs} is reached. Due to computational limitations, however, we have to find means to speed up the estimation of this value. A possible modification of this 'brute force' approach is to replace the probability of obtaining exactly S_{obs} by the probability of obtaining nearly S_{obs}. More precisely, *nonparametric* or *distribution-free* approximations are based on the idea of approximating the probability density of the simulation output at S_{obs} based on those samples within many simulated S_{sim} that are close to S_{obs} (Fig. 5a). A traditional method to do this is kernel density estimation (Tian *et al.* 2007, see also Alg. 1 in Appendix S1). Recently, however, it was realized that a simpler nonparametric approximation can be combined very efficiently with the sampling techniques that are discussed in the next section (Tavare *et al.* 1997; Marjoram *et al.* 2003; Sisson *et al.* 2007). A detailed discussion of these methods, known collectively as ABC, is given in Box 2.

Parametric likelihood approximations

The estimation of $p(S_{obs}|\phi)$ by distribution-free methods makes, as the name suggests, no assumptions about the distribution that would

asymptotically be generated by the stochastic simulation model (recall Fig. 1). However, when the summary statistic consists, for example, of a sum of many independent variables, the central limit theorem suggests that the distribution of the simulated summary statistics should be approximately normal (see, e.g. our example in Fig. 3). In this case, it seems obvious to approximate the outcomes of several simulated S_{sim} by a normal model.

The advantage of such a *parametric approximation* (Fig. 5b) as opposed to a distribution-free approximation is that imposing additional information about the distribution of the simulation output can help generate better estimates from a limited number of simulation runs. On the other hand, those estimates may be biased if the assumed distribution $g(S)$ does not conform to the true shape of the model output. We therefore view $p(S_{obs}|g(S))$ as a pseudo-likelihood (Besag 1974).

There are a number of authors and methodologies that explicitly or implicitly use parametric approximations. An instructive example, matching closely the illustration in Fig. 5b, is Wood (2010), who summarizes the variability of the simulated summary statistics by their mean values and their covariance matrix, and uses these together with a multivariate normal model to generate what he calls a 'synthetic likelihood' (see also Alg. 5 in Appendix S1). Very similar approaches are simulated pseudo-maximum likelihood estimation (Laroque & Salanié 1993; Concordet & Nunez 2002; Hyrien *et al.* 2005) and the 'simulated goodness-of-fit' (Riley *et al.* 2003). A further related method is Bayesian emulation (Henderson *et al.* 2009). In a wider sense, we also view Monte Carlo within Metropolis Approximation and grouped independence Metropolis-Hastings (O'Neill *et al.* 2000; Beaumont 2003; Andrieu & Roberts 2009) as parametric approximations, although we see their prime concern not in the approximation of $p(D|\phi)$ as such, but rather in the connection of a point-wise approximation with the sampling algorithms discussed in the next section.

External error models and informal likelihoods

The nonparametric and parametric likelihood approximations that we have discussed in the two previous subsections try to estimate the

output variability that is predicted by the stochastic simulation model, and use this information for the inference (Fig. 5). It is possible, however, that the simulated summary statistics show much less variability than the data. For example, despite being highly stochastic on the individual level, an individual-based population model with a large number of individuals may produce population size predictions that are practically deterministic. When such a model is compared with field data, it may turn out that it is highly unlikely that the variability in the field data could originate from the assumed stochastic processes only. In such a case, there must be either a fundamental model error (in the sense that the mean model predictions do not fit to the data), or additional stochastic processes have acted on the data that are not included in the model, for example, additional observation uncertainty or unobserved environmental covariates. To be able to apply the approximations discussed in the previous subsections, one would need to include processes that explain this variability within the stochastic simulation model (see, e.g. Zurell *et al.* 2009). However, particularly when those processes are not of interest for the scientific question asked, it is simpler and more parsimonious to express this unexplained variability outside the stochastic simulation (Fig. 5c).

One way to do this is adding an *external error model* with a tractable likelihood on top of the results of the stochastic simulation. This error model can be based on known observation uncertainties. An alternative, particularly when working with summary statistics, is estimating the error model from the variability of the observed data. For example, if our summary statistic was the mean of the data, we can use standard parametric or nonparametric methods to estimate the error of the mean (or rather its asymptotic distribution) from the observations. Most studies that use this approach then explain all the variability by the external error model and treat the stochastic model as deterministic on the level of the simulated summary statistics, potentially by calculating the mean of multiple simulated outcomes. Martínez *et al.* (2011), for example, compare the mean predictions of a stochastic individual-based model of trees with observed alpine tree line data under an external statistical model that is generated from the data. In this example, the stochasticity within the simulation may still be important to generate the correct mean model predictions, but all deviance between model and data is explained by the empirical variability within the observed data. In principle, however, it would also be possible to combine the likelihood approximations discussed in the previous sections with an external error model (see, e.g. Wilkinson 2008).

If it is difficult to specify an explicit statistical error model from the data, *informal likelihoods* offer an alternative. By informal likelihoods, we understand any metric that quantifies the distance between the predictions of a stochastic simulation model and the observed data, but is not immediately interpretable as originating from an underlying stochastic process (see Beven 2006; Smith *et al.* 2008, for a discussion of informal likelihoods in the context of the Generalized Likelihood Uncertainty Estimation method). Other terms that are often used synonymously are 'objective function' (Refsgaard *et al.* 2007) or 'cost function'. A common example is the sum of the squared distances between S_{obs} and the mean of S_{sim} (Refsgaard *et al.* 2007; Winkler & Heinken 2007), but many other measures are possible (Smith *et al.* 2008; Zinck & Grimm 2008; Duboz *et al.* 2010, see also Schröder & Seppelt 2006, for objective functions used in landscape ecology).

Structurally, there may be no difference between informal likelihoods and external error models – the sum of squared distances between mean model predictions and data could be interpreted as an informal likelihood as well as an external observation error, depending on whether it was chosen *ad hoc*, or with the knowledge that the deviation from the data is well described by an independent and identically distributed normal error. There is, however, a fundamental difference in the interpretation of the two. Only if the distance between model and data is calculated from an external statistical model that is in agreement with our knowledge about the system and the data, it makes sense to use confidence intervals and posterior distributions with their usual statistical interpretation. In principle, it is therefore always advisable either to approximate the likelihood directly (see previous subsections), or to construct an external error model. If there is reason to think, however, that the dominant part of the discrepancy between model and data does not originate from stochastic variation, but from a systematic or structural error, informal likelihoods offer an alternative for parameter and uncertainty estimation (Beven 2006).

Rejection filters

A fourth group of methods that is frequently used is what we call *rejection filters* (Fig. 5d). Rejection filters do not aim to provide a direct approximation of the likelihood, but rather divide models or parameterizations into two classes – likely and unlikely. For this purpose, they use (multiple) filter criteria to choose those models or parameter combinations that seem to be reasonably likely to reproduce the data, and reject the rest (e.g. Alg. 6 in Appendix S1). One may view them as analogous to classical rejection tests. Wiegand *et al.* (2004a) or Rossmanith *et al.* (2007), for example, use filter criteria that explicitly use the variability that is created by the simulation model (see Fig. 6). Other authors use filter criteria that correspond more to a filter-based version of an external error model or an informal likelihood, in the sense that acceptance intervals are not based on the variability of the simulation outputs, but on other criteria such as the estimated measurement uncertainty of the data. Examples of the latter are Kramer-Schadt *et al.* (2004), Rossmanith *et al.* (2007), Swanack *et al.* (2009), Topping *et al.* (2010) within the POM approach (Wiegand *et al.* 2003; Grimm *et al.* 2005), or Liu *et al.* (2009) and Blazkova & Beven (2009) who call the filter criteria the 'limits of acceptability' (Beven 2006).

The advantage of using multiple independent filters as opposed to combining all information into one informal likelihood approximation is that filters require fewer *ad hoc* assumptions, may ideally be grounded on statistical rejection tests, and are more robust to correlations between summary statistics. The cost, on the other hand, is that many of the optimization and sampling methods discussed in the next section cannot be applied because they rely on calculating the likelihood ratio between two sets of parameters. As a mixture between multiple rejection filters and the informal likelihood approximation, one may also apply pareto-optimization of multiple informal objectives (Komuro *et al.* 2006), which may potentially ease problems of correlations between summary statistics within informal likelihoods, while still allowing for systematic optimization.

LIKELIHOOD ACROSS THE PARAMETER SPACE – EFFICIENT SAMPLING

In the previous section, we have discussed different possibilities of approximating $p(D_{obs}|\phi)$ for a fixed model parameterization ϕ. For most practical applications, what we are really interested in is to see
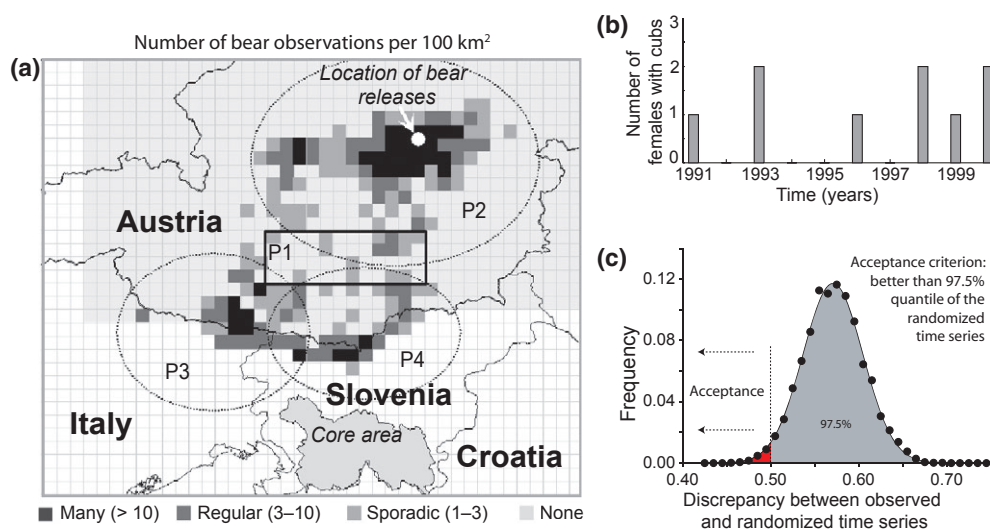
Figure 6 An example of Pattern-Oriented Modelling. The objective of the study was to reconstruct and understand the demographics of brown bears (*Ursus arctos*) after the reintroduction into the Eastern Alps. To this end, a spatially explicit stochastic simulation model was constructed. Details are provided in Wiegand *et al.* (2004a,b) and in Algorithm 6 in Appendix S1. The model fit was based on two patterns (summary statistics): the frequency of bear observations at 10×10 km resolution (P1–P4 in Panel A), and a time series of observed female cubs (Panel B). For the time series, the simulation was rejected if the difference between model predictions and data (based on the average of the sum-of-squares of the 2-year running mean of the simulation) was larger than the 97.5% quantile of 5000 randomized generated from the observed time series by permutation (Panel C). A similar test was created for the spatial bear observations. Finally, rejection sampling (see Alg. 6 in Appendix S1) was applied to filter those parameters that could not be rejected by both criteria.

how this estimate varies over a larger parameter space, that is, to find the maximum or the shape of the likelihood $\mathscr{L}(\phi) \propto p(D_{obs}|\phi)$ or the posterior density $\mathscr{P}(\phi)$ (see Box 1) as a function of the parameters ϕ.

Recall that, although we may approximate $\mathscr{L}(\phi)$ for each ϕ, we can in general not express $\mathscr{L}(\phi)$ as an analytical function. For a low-dimensional problem, for example, a model with only one parameter, this poses no problem because we may simply calculate $\mathscr{L}(\phi)$ for a number of points and use them to interpolate maximum and shape of $\mathscr{L}(\phi)$. With a growing number of parameters, however, it becomes increasingly difficult to cover the parameter space densely. Therefore, we need a second approximation step to generate estimates for maximum and shape of $\mathscr{L}(\phi)$ from the point-wise likelihood approximations that have been discussed in the previous section.

Two classes of algorithms are relevant in this context: *optimization algorithms* for finding the parameter combination with the highest likelihood or posterior value, and *sampling algorithms* such as *Markov Chain Monte Carlo* (MCMC) or *particle filters* that explore the shape of the likelihood or posterior distribution in high-dimensional parameter spaces. Optimization functions such as simplex search methods, simulated annealing or genetic algorithms are generally well supported by all major computational environments such as Matlab, Mathematica, Octave, R and Python (Scipy). In the following, we therefore concentrate on sampling algorithms that aim at creating samples from a function of ϕ (usually called the target distribution) that is unknown analytically, but can be evaluated point-wise for each ϕ. These algorithms are typically applied in Bayesian statistics, where the target distribution is the posterior density $\mathscr{P}(\phi)$ (see Box 1). To avoid new notation, however, we use for the following examples $\mathscr{L}(\phi)$ as the target distribution, assuming that the integral of $\mathscr{L}(\phi)$ is finite (integrability of the target distribution is a requirement for the sampling algorithms). Moreover, note that all methods discussed in this section are suited for models with likelihoods (or posterior

densities) that are point-wise approximated by simulation, but may be applied to models with tractable likelihoods alike. A few particularities that arise from the fact that $\mathscr{L}(\phi)$ itself is an estimate that varies with each simulation are discussed at the end of this section.

Rejection sampling

The simplest possibility of generating a distribution that approximates $\mathscr{L}(\phi)$ is to sample random parameters ϕ and accept those proportionally to their (point-wise approximated) value of $\mathscr{L}(\phi)$ (Fig. 7, left). This approach can be slightly improved by importance sampling or stratified sampling methods such as the Latin hypercube design, but rejection approaches encounter computational limitations when the dimensionality of the parameter space becomes larger than typically 10–15 parameters. Examples for rejection sampling are Thornton & Andolfatto (2006) in a population genetic study of *Drosophila melanogaster* and Jabot & Chave (2009) who combined a neutral model with phylogenetic data (both using ABC, see Box 2 and Alg. 2 in Appendix S1) or Kramer-Schadt *et al.* (2004), Swanack *et al.* (2009) and Topping *et al.* (2010) who used POM (see Alg. 6 in Appendix S1) to parameterize population models for lynx, amphibians, and grey partridges, respectively.

Markov chain Monte Carlo

A more sophisticated class of algorithms comprises MCMC. These algorithms construct a Markov chain of parameter values $(\phi_1, \ldots, \phi_n)$, where the next parameter combination ϕ_{i+1} is chosen by proposing a random move conditional on the last parameter combination ϕ_i, and accepting conditional on the ratio of $\mathscr{L}(\phi_{i+1})/\mathscr{L}(\phi_i)$ (Fig. 7, middle). Given that certain conditions are met (see, e.g. Andrieu *et al.* 2003), the Markov chain of parameter values will eventually converge to the target distribution $\mathscr{L}(\phi)$. The advantage of an MCMC is that
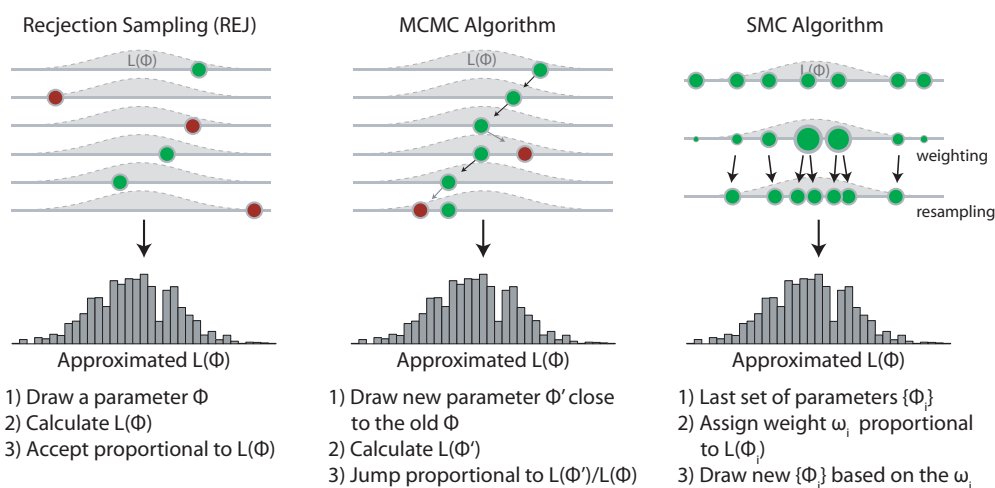
Figure 7 Illustration of algorithms for approximating the distribution that results from calculating the point-wise likelihood approximation across the parameter space of ϕ. The light grey shapes depict this distribution, which is, however, usually not known before applying the algorithm. Circles depict parameter combinations within the algorithm. Rejection sampling (left) draws random parameters and accepts (green) or rejects (red) according to the calculated point-wise likelihood approximation. A Metropolis-Hastings MCMC sampler (middle) proposes a new value conditional on the last, and accepts (green) or rejects (red) according to the ratio of the point-wise likelihood approximations. Sequential Monte Carlo samplers (right) start with an ensemble of parameter values, weight them according to their approximated point-wise likelihood values, and potentially draw new values from the last ensemble according to those weights.

the time needed to obtain acceptable convergence is typically much shorter than for rejection sampling, because the sampling effort is concentrated in the areas of high likelihood or posterior density. We recommend Andrieu *et al.* (2003) as a more thorough introduction to MCMC algorithms, and Van Oijen *et al.* (2005) as a good ecological example. MCMCs are used widely, for example, within ABC (Marjoram *et al.* 2003, see Alg. 3 in Appendix S1), and also for sampling informal likelihoods.

Sequential Monte Carlo methods

Particle filters or sequential Monte Carlo methods (SMCs) also try to concentrate the sampling effort in the areas of high likelihood or posterior density based on previous samples. Unlike MCMCs, however, each step of the algorithm contains not a single ϕ, but N parameter combinations ϕ_i (particles), that are assigned weights ω_i proportional to their likelihood or posterior value $\mathscr{L}(\phi_i)$ (see Arulampalam *et al.* 2002). When starting with a random sample of parameters, many particles may be assigned close to zero weights, meaning that they carry little information for the inference (degeneracy). To avoid this, a resampling step is usually added where a new set of particles is created based on the current weight distribution (Gordon *et al.* 1993; Arulampalam *et al.* 2002; Fig. 7, right). The traditional motivation for a particle filter is to include new data in each filter step, but the filter may also be used to work on a fixed dataset or to subsequently add independent subsets of the data. Particularly for the ABC approximation (Box 2), SMC algorithms may exhibit advantages over MCMCs, because they are less prone to get stuck in areas of low likelihood (Sisson *et al.* 2007; Beaumont *et al.* 2009; Toni *et al.* 2009, see Alg. 4 in Appendix S1).

A remark on the approximation error

So far, we have described inference of likelihood or posterior distributions based on two approximations: first, we have estimated

$p(D_{\mathrm{obs}}|\phi)$ point-wise for fixed parameters ϕ, and secondly, we have estimated the shape of the distribution that is generated by these point-wise approximations as a function of ϕ. The properties of the sampling algorithms for deterministic target distributions are well known: if implemented correctly, their sampling distribution will converge exactly to the target distribution in the limit of infinitely many steps. The only pitfall is to determine whether a sampler has already converged sufficiently close to the exact solution after a fixed number of steps. For non-pathological cases, this can usually be assessed by convergence diagnostics (see, e.g. Cowles & Carlin 1996), although a rigorous proof of convergence is usually not possible.

The properties of sampling algorithms in combination with a stochastic target distribution (likelihood or posterior) that results from simulation-based approximations of $p(S_{\mathrm{obs}}|\phi)$ as discussed in the previous section, however, are less widely known. A basic requirement is that the expectation value of the point-wise approximation of $p(S_{\mathrm{obs}}|\phi)$ must be unbiased (if there is an approximation bias of $p(S_{\mathrm{obs}}|\phi)$, one may try to correct it, see e.g. O'Neill *et al.* 2000). It is easy to see that, if $p(S_{\mathrm{obs}}|\phi)$ is unbiased for all ϕ, rejection sampling algorithms will converge exactly. However, when new samples depend on previous samples and are therefore not fully independent, the situation is somewhat more complex: if a point-wise approximation of $p(S_{\mathrm{obs}}|\phi)$ is used several times in an MCMC (e.g. when the algorithm remains several times at the same parameter value), the estimate must not be recalculated, otherwise the resulting distribution may not be unbiased anymore, even if $p(S_{\mathrm{obs}}|\phi)$ is unbiased (Beaumont 2003; Andrieu & Roberts 2009). The good news is that the combined approximation still converges exactly as long as the previous requirements are met. The downside is that MCMCs convergence may be slowed down considerably when the variance of the point-estimate of $p(S_{\mathrm{obs}}|\phi)$ is large compared with typical likelihood differences between parameter values. If the latter is the case, MCMCs may get repeatedly stuck at likelihood estimates that are particularly favourable due to the stochasticity in the approximation. One way out of this dilemma would be to recalculate point-wise likelihood

---

Box 2 Approximate Bayesian Computing

Approximate Bayesian Computing is a class of sampling algorithms that have attracted a lot of interest in recent years (Beaumont 2010). The key innovation of ABC is to combine a nonparametric point-wise likelihood approximation (Fig. 5a) in one step with the efficient sampling methods (Fig. 7).

The likelihood approximation is achieved by defining conditions under which a sample from the stochastic simulation model is close enough to the observed data to be considered equal. More technically, ABC algorithms approximate the likelihood of sampling a S_{sim} that is *identical* to S_{obs} by the probability of sampling summary statistics S_{sim} that are closer than ε to S_{obs} under a metric $d(S,S')$:

$$p(S_{sim} = S_{obs}|\phi) \approx c \cdot p(d(S_{sim}, S_{obs}) < \epsilon|\phi). \tag{3}$$

where c is a proportionality constant. In Fig. 5a, we depict this idea graphically.

The second step for constructing an ABC algorithm is the realization that the sampling algorithms (Fig. 7) do not actually require the value of $p(S_{obs}|\phi)$ as such. What they need is an algorithm that returns an 'acceptance-decision' for new parameters that is proportional to their likelihood. Therefore, instead of approximating $p(S_{obs}|\phi)$ according to eqn 3 and then using this value to decide about the next step, one can generate such a draw directly by testing whether $d(M,S_{obs}) \leq \epsilon$ and accept according to the result. This step was first included in the rejection sampling algorithm (Fu & Li 1997; Tavare *et al.* 1997; Pritchard *et al.* 1999), then in a Metropolis MCMC algorithm (Marjoram *et al.* 2003), and finally into sequential Monte Carlo algorithms (Sisson *et al.* 2007; Beaumont *et al.* 2009; Toni *et al.* 2009) (see algorithms 2, 3 and 4 in Appendix S1). We suggest Beaumont (2010) as a more detailed reference to this development and current trends in ABC, as well as the reviews of Bertorelle *et al.* (2010), Csilléry *et al.* (2010) and Lopes & Boessenkool (2010). Some interesting examples of studies that use ABC are Ratmann *et al.* (2007), François *et al.* (2008), Jabot & Chave (2009), Jabot (2010) and Wilkinson *et al.* 2011).

The ABC approach is asymptotically exact, meaning that it will, for suitable distance measures and sufficient summary statistics, reproduce the true shape of the likelihood function in the limit of $\epsilon \rightarrow 0$ and $N \rightarrow \infty$, N being the sampling effort (Marjoram *et al.* 2003). For all practical applications, however, we will have to choose $\epsilon > 0$ to speed up the convergence. The larger ε, the more posterior distributions are biased towards the prior and therefore typically wider than the true posterior. The approximation error becomes particularly important because the approximation eqn 3 suffers from the curse of dimensionality: the higher the number of summary statistics used for the fit, the larger will ε typically be chosen to achieve reasonable convergence rates.

Fortunately, a few strategies may be applied to reduce the approximation error. It is advisable to scale the metric $d(S,S')$ used for eqn 3 to the variance of the summary statistic s', to have a comparable approximation error for each dimension of S (Beaumont *et al.* 2002; Bazin *et al.* 2010). Blum (2010) suggests, in the context of a particular regression adjustment, to rescale the summary statistics to achieve a homoscedastic response to the parameters of interest. Also, it may be useful to test whether the choice of the metric $d(S,S')$ influences the approximation error (Sousa *et al.* 2009). The remaining approximation error may be corrected at least partly by *post-sampling regression adjustment*. The idea behind this is to use the posterior sample together with the recorded distances under the summary statistics to fit a (weighted) regression model that relates the model parameters with the distance to the data (see Fig. 8). The result is used to correct the sampled parameter values (Beaumont *et al.* 2002; Wegmann *et al.* 2009; Blum & François 2010; Leuenberger & Wegmann 2010). Another very appealing idea was presented by Wilkinson (2008): the error in the acceptance criterion eqn 3 may also be interpreted as the exact fit to a different model with an additional statistical error model that is represented by the approximation eqn 3 on top of the stochastic simulation. Moreover, the acceptance rules of eqn 3 may be adjusted to represent practically any error model. Thus, for cases where there is a large observation error on top of the stochastic simulation model, this error may be encoded in eqn 3 and ABC yields posteriors that are exact for the combined model.
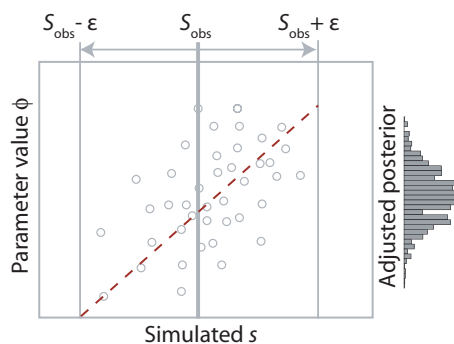

Figure 8 The principle of regression adjustment: circles denote the value of the summary statistics and the parameter value of all points that were accepted by the ABC algorithm. Note that all points have values of *s* that are within a distance of ε from the observed value S_{obs}. Ideally, we would have liked ε to be zero, but in this case, no value would have been accepted. We can, however, estimate what should have happened for $\epsilon = 0$ by using a linear (Beaumont *et al.* 2002) or nonlinear (Blum & François 2010) regression model to relate the parameter value ϕ with the distance between the simulated *s* and S_{obs}. Based on this model, the 'true' conditional where $\epsilon \rightarrow 0$ (bold grey line in the middle) can be estimated.

approximations every time they are used, but one has to be aware that convergence to the true posterior is then not guaranteed.

Optimization algorithms will generally be less robust to large point-wise approximation errors. One should therefore make sure that either the variance of the estimate of $p(S_{obs}|\phi)$ is low for all ϕ, which may be influenced by the number of simulation runs that are used for the approximation, or that the employed algorithm is robust with respect to stochasticity in the objective function.

CONCLUSIONS

Stochastic simulation models are of high relevance for biological and ecological research because they allow the simulation of complex stochastic processes without having to represent these processes in a traditional statistical model with a tractable likelihood function. To connect stochastic simulation models to data, however, it is necessary to construct likelihood approximations that make them usable for statistical inference. In this review, we have discussed methods to derive such likelihood approximations from samples that are drawn from stochastic simulation models. Although originating from different fields, all use three essential steps:

(1) Comparing observed and simulated data through summary statistics.
(2) Approximating the likelihood that the observed summary statistics are obtained from the simulation.
(3) Efficient sampling of the parameter space.

We have concentrated our discussion mainly on parameter estimation, but once appropriate likelihood approximations are established, model selection and uncertainty estimation can, in principle, be done in the same way as in other statistical applications (e.g. Beaumont 2010; Toni & Stumpf 2010). Yet, there is one particularity that has to be kept in mind regarding model selection with summary statistics: the fact that a summary statistic is sufficient for parameter estimation of a set of models does not yet imply that this statistic is also sufficient for model selection, that is, for a comparison between these models (Didelot *et al.* 2011; Robert *et al.* 2011). Didelot *et al.* (2011) point out a few cases where sufficiency can be guaranteed, but Robert *et al.* (2011) caution that it may be very difficult and costly to assure model selection sufficiency in general. Whether this problem can be satisfyingly solved will remain a question for further research.

By transferring the problem of inference for stochastic models to the problem of inference for statistical models, we have inherited some discussions that are held within statistical research, for example, the choice of appropriate model selection criteria (Johnson & Omland 2004), the effective number of parameters (Spiegelhalter *et al.* 2002; Plummer 2008) or the choice of non-informative priors (Kass & Wasserman 1996; Irony & Singpurwalla 1997) for (implicit) statistical models. In our opinion, however, being able to build on this experience is a clear advantage. An increasing use of statistical inference with stochastic simulation models may even provide valuable stimulation to these debates, as some classical statistical questions such as the effective number of parameters of a model become particularly important for complex simulation models.

The main issues, however, that still need to be addressed to make statistical inference for stochastic simulation models widely accessible are usability and standardization. Likelihood approximation of stochastic simulation models is an emerging field and for many problems there are no solutions that work out-of-the box.

With time, we will be able to build on more experience about which summary statistics are sufficient for which model types. Also, simulation models will be built or modified with the purpose of parameterization in mind: the efficiency of sampling algorithms, for example, may be increased dramatically when parameterizations are chosen as independent and linear as possible with respect to the model output. And finally, judging from the references reviewed and their terminology, there has been little discussion across the borders of different fields that have developed inferential methods for stochastic simulation models. We therefore hope that this review will not only draw attention to and provide practical guidance for applying these useful methods, but that it will also stimulate the exchange of ideas across existing likelihood approximation methods, and in general between the communities using statistical and stochastic simulation models.

REFERENCES

de Aguiar, M.A.M., Baranger, M., Baptestini, E.M., Kaufman, L. & Bar-Yam, Y. (2009). Global patterns of speciation and diversity. *Nature*, 460, 384–387.

Alonso, D., Ostling, A. & Etienne, R.S. (2008). The implicit assumption of symmetry and the species abundance distribution. *Ecol. Lett.*, 11, 93–105.

Andrieu, C. & Roberts, G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Stat.*, 37, 697–725.

Andrieu, C., de Freitas, N., Doucet, A. & Jordan, M.I. (2003). An introduction to MCMC for machine learning. *Mach. Learn.*, 50, 5–43.

Arita, H.T. & Vázquez-Domínguez, E. (2008). The tropics: cradle, museum or casino? A dynamic null model for latitudinal gradients of species diversity. *Ecol. Lett.*, 11, 653–663.

Arulampalam, M.S., Maskell, S., Gordon, N. & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal. Process.*, 50, 174–188.

Bazin, E., Dawson, K.J. & Beaumont, M.A. (2010). Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics*, 185, 587–602.

Beaumont, M.A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164, 1139–1160.

Beaumont, M.A. (2010). Approximate Bayesian computation in evolution and ecology. *Ann. 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.

Beaumont, M.A., Cornuet, J.M., Marin, J.M. & Robert, C.P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, 96, 983–990.

Berkley, H.A., Kendall, B.E., Mitarai, S. & Siegel, D.A. (2010). Turbulent dispersal promotes species coexistence. *Ecol. Lett.*, 13, 360–371.

Bertorelle, G., Benazzo, A. & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.*, 19, 2609–2625.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. B Methodol.*, 36, 192–236.

Beven, K. (2006). A manifesto for the equifinality thesis. *J. Hydrol.*, 320, 18–36.

Blazkova, S. & Beven, K. (2009). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resour. Res.*, 45, W00B16.

Blum, M.G.B. (2010). Approximate Bayesian computation: a nonparametric perspective. *J. Am. Stat. Assoc.*, 105, 1178–1187.

Blum, M.G.B. & François, O. (2010). Non-linear regression models for approximate bayesian computation. *Stat. Comput.*, 20, 63–73.

Bridle, J.R., Polechová, J., Kawata, M. & Butlin, R.K. (2010). Why is adaptation prevented at ecological margins? New insights from individual-based simulations. *Ecol. Lett.*, 13, 485–494.

Chave, J., Muller-Landau, H.C. & Levin, S.A. (2002). Comparing classical community models: theoretical consequences for patterns of diversity. *Am. Nat.*, 159, 1–23.

Clark, J.S. & Gelfand, A.E. (2006). A future for models and data in environmental science. *Trends Ecol. Evol.*, 21, 375–380.

Concordet, D. & Nunez, O. (2002). A simulated pseudo-maximum likelihood estimator for nonlinear mixed models. *Comput. Stat. Data Anal.*, 39, 187–201.

Couzin, I.D., Krause, J., Franks, N.R. & Levin, S.A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature*, 433, 513–516.

Cowles, M.K. & Carlin, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Stat. Assoc.*, 91, 883–904.

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

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B Methodol.*, 39, 1–38.

Didelot, X., Everitt, R.G., Johansen, A.M. & Lawson, D.J. (2011). Likelihood-free estimation of model evidence. *Bayesian Anal.*, 6, 1–28.

Diggle, P.J. & Gratton, R.J. (1984). Monte Carlo methods of inference for implicit statistical models. *J. R. Stat. Soc. B Methodol.*, 46, 193–227.

Dislich, C., Günter, S., Homeier, J., Schröder, B. & Huth, A. (2009). Simulating forest dynamics of a tropical montane forest in South Ecuador. *Erdkunde*, 63, 347–364.

Drovandi, C.C., Pettitt, A.N. & Faddy, M.J. (2011). Approximate Bayesian computation using indirect inference. *J. R. Stat. Soc. Ser. C. (Appl. Stat.)*, 60, 317–337.

Duboz, R., Versmisse, D., Travers, M., Ramat, E. & Shin, Y.J. (2010). Application of an evolutionary algorithm to the inverse parameter estimation of an individual-based model. *Ecol. Model.*, 221, 840–849.

Ellison, A.M. (2004). Bayesian inference in ecology. *Ecol. Lett.*, 7, 509–520.

Fearnhead, P. & Prangle, D. (2010). Semi-automatic Approximate Bayesian Computation. arXiv:1004.1112v1.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. A.*, 222, 309–368.

François, O., Blum, M.G.B., Jakobsson, M. & Rosenberg, N.A. (2008). Demographic history of European populations of Arabidopsis thaliana. *PLoS Genet.*, 4, e1000075.

Fu, Y.X. & Li, W.H. (1997). Estimating the age of the common ancestor of a sample of DNA sequences. *Mol. Biol. Evol.*, 14, 195–199.

Gordon, N.J., Salmond, D.J. & Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE. Proc. F*, 140, 107–113.

Gourieroux, C., Monfort, A. & Renault, E. (1993). Indirect inference. *J. Appl. Econ.*, 8, S85–S118.

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

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F. *et al.* (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*, 310, 987–991.

Heggland, K. & Frigessi, A. (2004). Estimating functions in indirect inference. *J. R. Stat. Soc. B Methodol.*, 66, 447–462.

Henderson, D.A., Boys, R.J., Krishnan, K.J., Lawless, C. & Wilkinson, D.J. (2009). Bayesian emulation and calibration of a stochastic computer model of mitochondrial DNA deletions in Substantia Nigra neurons. *J. Am. Stat. Assoc.*, 104, 76–87.

Holmes, E.E. (2004). Beyond theory to application and evaluation: diffusion approximations for population viability analysis. *Ecol. Appl.*, 14, 1272–1293.

Huth, A. & Ditzer, T. (2000). Simulation of the growth of a lowland Dipterocarp rain forest with FORMIX3. *Ecol. Model.*, 134, 1–25.

Huth, A. & Wissel, C. (1992). The simulation of the movement of fish schools. *J. Theor. Biol.*, 156, 365–385.

Hyrien, O., Mayer-Pröschel, M., Noble, M. & Yakovlev, A. (2005). A stochastic model to analyze clonal data on multi-type cell populations. *Biometrics*, 61, 199–207.

Irony, T.Z. & Singpurwalla, N.D. (1997). Non-informative priors do not exist – a dialogue with José M. Bernardo. *J. Stat. Plan. Infer.*, 65, 159–177.

Jabot, F. (2010). A stochastic dispersal-limited trait-based model of community dynamics. *J. Theor. Biol.*, 262, 650–661.

Jabot, F. & Chave, J. (2009). Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol. Lett.*, 12, 239–248.

Jiang, W. & Turnbull, B. (2004). The indirect method: inference based on intermediate statistics – a synthesis and examples. *Stat. Sci.*, 19, 239–263.

Johnson, J.B. & Omland, K.S. (2004). Model selection in ecology and evolution. *Trends Ecol. Evol.*, 19, 101–108.

Joyce, P. & Marjoram, P. (2008). Approximately sufficient statistics and Bayesian computation. *Stat. Appl. Genet. Mo. B.*, 7, 26.

Kass, R.E. & Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.*, 91, 1343–1370.

Komuro, R., Ford, E.D. & Reynolds, J.H. (2006). The use of multi-criteria assessment in developing a process model. *Ecol. Model.*, 197, 320–330.

Kramer-Schadt, S., Revilla, E., Wiegand, T. & Breitenmoser, U. (2004). Fragmented landscapes, road mortality and patch connectivity: modelling influences on the dispersal of Eurasian lynx. *J. Appl. Ecol.*, 41, 711–723.

Laroque, G. & Salanié, B. (1993). Simulation-based estimation of models with lagged latent variables. *J. Appl. Econ.*, 8, S119–S133.

Leuenberger, C. & Wegmann, D. (2010). Bayesian computation and model selection without likelihoods. *Genetics*, 184, 243–252.

Liu, Y., Freer, J., Beven, K. & Matgen, P. (2009). Towards a limits of acceptability approach to the calibration of hydrological models: extending observation error. *J. Hydrol.*, 367, 93–103.

Lopes, J. & Boessenkool, S. (2010). The use of approximate Bayesian computation in conservation genetics and its application in a case study on yellow-eyed penguins. *Conserv. Genet.*, 11, 421–433.

Marjoram, P., Molitor, J., Plagnol, V. & Tavare, S. (2003). Markov chain Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. U S A*, 100, 15324–15328.

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. *Am. Nat.*, 5, E136–E152.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 5, 995–1026.

Nathan, R., Safriel, U.N. & Noy-Meir, I. (2001). Field validation and sensitivity analysis of a mechanistic model for tree seed dispersal by wind. *Ecology*, 82, 374–388.

O'Neill, P.D., Balding, D.J., Becker, N.G., Eerola, M. & Mollison, D. (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. C. Appl.*, 49, 517–542.

Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends Ecol. Evol.*, 23, 87–94.

Pawitan, Y. (2001). *In all Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, Oxford, NY.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9, 523–539.

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.

Ratmann, O., Jørgensen, O., Hinkley, T., Stumpf, M., Richardson, S. & Wiuf, C. (2007). Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*. *PLoS Comp. Biol.*, 3, e230.

Refsgaard, J., der Sluijs, J., Højberg, A. & Vanrolleghem, P. (2007). Uncertainty in the environmental modelling process – a framework and guidance. *Environ. Model. Softw.*, 22, 1543–1556.

Riley, S., Donnelly, C.A. & Ferguson, N.M. (2003). Robust parameter estimation techniques for stochastic within-host macroparasite models. *J. Theor. Biol.*, 225, 419–430.

Robert, C.P., Cornuet, J.M., Marin, J.M. & Pillai, N. (2011). Lack of confidence in ABC model choice. arXiv:1102.4432v3.

Rossmanith, E., Blaum, N., Grimm, V. & Jeltsch, F. (2007). Pattern-oriented modelling for estimating unknown pre-breeding survival rates: the case of the Lesser Spotted Woodpecker (Picoides minor). *Biol. Conserv.*, 135, 555–564.

Ruokolainen, L., Ranta, E., Kaitala, V. & Fowler, M.S. (2009). When can we distinguish between neutral and non-neutral processes in community dynamics under ecological drift? *Ecol. Lett.*, 12, 909–919.

Schröder, B. & Seppelt, R. (2006). Analysis of pattern-process interactions based on landscape models–overview, general concepts, and methodological issues. *Ecol. Model.*, 199, 505–516.

Sims, D.W., Southall, E.J., Humphries, N.E., Hays, G.C., Bradshaw, C.J.A., Pitchford, J.W. *et al.* (2008). Scaling laws of marine predator search behaviour. *Nature*, 451, 1098–1102.

Sisson, S.A., Fan, Y. & Tanaka, M.M. (2007). Sequential Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. U S A*, 104, 1760–1765.

Smith, P., Beven, K.J. & Tawn, J.A. (2008). Informal likelihood measures in model assessment: theoretic development and investigation. *Adv. Water Resour.*, 31, 1087–1100.

Sousa, V.C., Fritz, M., Beaumont, M.A. & Chikhi, L. (2009). Approximate Bayesian computation without summary statistics: the case of admixture. *Genetics*, 181, 1507–1519.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. B*, 64, 583–639.

Swanack, T.M., Grant, W.E. & Forstner, M.R. (2009). Projecting population trends of endangered amphibian species in the face of uncertainty: a pattern-oriented approach. *Ecol. Model.*, 220, 148–159.

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

Thornton, K. & Andolfatto, P. (2006). Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster. *Genetics*, 172, 1607–1619.

Tian, T., Xu, S., Gao, J. & Burrage, K. (2007). Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics*, 23, 84–91.

Toni, T. & Stumpf, M.P.H. (2010). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26, 104–110.

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., Høye, T.T., Odderskær, P. & Aebischer, N.J. (2010). A pattern-oriented modelling approach to simulating populations of grey partridge. *Ecol. Model.*, 221, 729–737.

Van Oijen, M., Rougier, J. & Smith, R. (2005). Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiol.*, 25, 915–927.

Wegmann, D., Leuenberger, C. & Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, 182, 1207–1218.

Wiegand, T., Jeltsch, F., Hanski, I. & Grimm, V. (2003). Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. *Oikos*, 100, 209–222.

Wiegand, T., Knauer, F., Kaczensky, P. & Naves, J. (2004a). Expansion of brown bears (*Ursus arctos*) into the Eastern alps: a spatially explicit population model. *Biodivers. Conserv.*, 13, 79–114.

Wiegand, T., Revilla, E. & Knauer, F. (2004b). Dealing with uncertainty in spatially explicit population models. *Biodivers. Conserv.*, 13, 53–78.

Wikle, C.K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84, 1382–1394.

Wilkinson, R.D. (2008). Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. arXiv:0811.3355v1.

Wilkinson, D.J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.*, 10, 122–133.

Wilkinson, R.D., Steiper, M.E., Soligo, C., Martin, R.D., Yang, Z. & Tavaré, S. (2011). Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst. Biol.*, 60, 16–31.

Winkler, E. & Heinken, T. (2007). Spread of an ant-dispersed annual herb: an individual-based simulation study on population development of Melampyrum pratense L. *Ecol. Model.*, 203, 424–438.

Wood, S.N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466, 1102–1104.

Zinck, R.D. & Grimm, V. (2008). More realistic than anticipated: a classical forest-fire model from statistical physics captures real fire shapes. *Open Ecol. J.*, 6, 8–13.

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T. *et al.* (2009). The virtual ecologist approach: simulating data and observers. *Oikos*, 119, 622–635.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Algorithms.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.