



The value of a datum – how little data do we need for a quantitative risk analysis?

Brian Leung^{1,2*} and Russell J. Steele³

¹Department of Biology, McGill University, Montreal, Quebec, H3A 1B1, Canada,

²School of Environment, McGill University, Montreal, Quebec, H3A 2A7, Canada,

³Department of Mathematics and Statistics, McGill University, Montreal, Quebec, H3A 0B9, Canada

ABSTRACT

Aim Conservation managers are typically faced with limited resources, time and information. The philosophy underlying risk assessment should be robust to these limitations. While there is a broad support for the concept of risk assessments, there is a tendency to rely on expert opinion and exclude formal data analysis, possibly because available information is often scarce. When data analyses are conducted, often much simplified models are advocated, even though this means excluding processes believed by experts to be important. In this manuscript, we ask: should statistical analyses be conducted and decisions modified based on a single datum? How many data points are needed before predictions are meaningful? Given limited data, how complex should models be?

Location World-wide.

Methods We use simulation approaches with known ‘true’ values to assess which inferences are possible, given different amounts of information. We use two metrics of performance: the magnitude of uncertainty (using posterior mean squared error) and bias (using P–P plots). We assess six models of relevance to conservation ecologists.

Results We show that the greatest reduction in uncertainty occurred at the smallest sample sizes for models examined, and much of parameter space could be excluded. Thus, analyses based on even a single datum potentially can be useful. Further, with only a few observations, the predicted distribution of outcomes matched the probabilities of actual occurrences, even for relatively complex state-space models with multiple sources of stochasticity.

Main conclusions We highlight the utility of quantitative analyses even with severely limited data, given existing practices and arguments in the conservation literature. The purpose of our manuscript is in part a philosophical discourse, as modifications are needed to how conservation ecologists are often trained to think about problems and data, and in part a demonstration via simulation analysis.

Keywords

Bayesian, hierarchical, mixed model, population viability analysis, simulation, uncertainty.

*Correspondence: Brian Leung, Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada.
E-mail: brian.leung2@mcgill.ca

INTRODUCTION

For conservation and environmental management decisions, there are consistent challenges of limited time, resources and information. Specifically, it is not possible to obtain perfect information for many environmental issues. Indeed, the

information available is often quite sparse, with few measurements available to make inferences on a variable of interest (e.g. population growth rate).

In these scenarios, there are three possible (non-mutually exclusive) options. First, one can do nothing and wait while more information is collected. However, to remain policy

relevant, it is often not feasible to wait (potentially years) while such information is obtained (i.e. ‘paralysis by analysis’, Shatkin, 2008). A second option is to use expert opinion. Work has proceeded on innovative ways to synthesize such opinions (Hosack *et al.*, 2008; Martin *et al.*, 2012; Estévez *et al.*, 2013). Nevertheless, biases and systematic underestimation of uncertainties are known to occur (Camerer & Johnson, 1997). Moreover, given the paucity of empirical data, quantitative statistical analyses often are dismissed entirely (Burgman, 2005). The third option is to use the available empirical evidence, even when the data are severely limited. While the third option may seem obvious to statisticians, our experience is that empirical data are often not incorporated in practice (e.g. Leung *et al.*, 2012). Thus, it is worthwhile to examine the inferences that are valid, given different amounts of information.

It is useful to structure our thinking on why one should incorporate limited data into risk assessments. First, the purpose of risk assessment is not to minimize risk per se, but rather to forecast outcomes and assess the magnitude of uncertainty, given available information. Second, many models (or parameter values) may be plausible given the limited data available, but certain models are more probable than others, that is, there is a probability distribution of potential models/outcomes. These distributions can be estimated using approaches such as Bayesian statistics (Clark *et al.*, 2005), although we recognize that other likelihood-based approaches are applicable too (e.g. Lele & Dennis, 2009). Nevertheless, we chose Bayesian methods and will henceforth refer

to these probability distributions of model parameters and outcomes as the posterior distribution and posterior predictive distribution, respectively. Third, from a decision theory perspective, high uncertainty is not a technical problem; one can use the output from a quantitative risk assessment regardless of the uncertainty. For instance, one can integrate across the posterior distribution, weighting by the probability that each model is true (Moschini & Hennessey, 2001), or apply a risk-aversion function (or dominance-based rankings) to the posterior to model risk-averse behaviour (Finnoff *et al.*, 2007; Yemshanov *et al.*, 2013) or determine the probability of exceeding a threshold of acceptable risk (Keller *et al.*, 2007). Fourth, given that data are often excluded in current risk assessments because they are deemed too limited to permit valid analyses, we should examine how few observations are required for the posterior (predictive) distribution to correctly reflect the probability that a model (outcome) is true.

We suggest two metrics of performance, using simulations to examine analyses based on limited data. First, we estimate the magnitude of uncertainty, using the posterior mean squared error (MSE). Second, we examine whether bias (in location, dispersion or shape) occurs in the posterior distribution, using probability–probability (P–P) plots. P–P plots are often used to determine whether two data sets agree or whether empirical cumulative distribution function (cdf) agrees with a theoretical one. Analogously, here we examine whether the posterior distribution agrees probabilistically with the true generating values (Fig. 1a). Specifically, if they

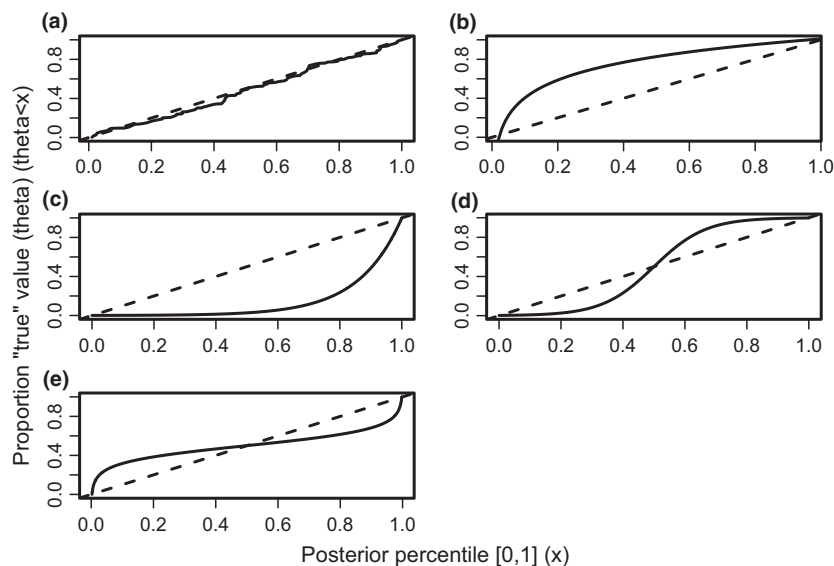


Figure 1 Possible patterns using P–P plots on simulated output. (a) The posterior (and posterior predictive) distribution is unbiased. The true generating values fall below the posterior percentile the correct proportion of the time, generating a 1:1 line in the P–P plot. That is, the true value falls below 1% of the cdf only 1% of the time, 2% of the cdf only 2% of the time, etc. (b) The posterior (predictive) distribution is biased and overestimates parameter values (outcomes). (c) The posterior (predictive) distribution is biased and underestimates parameter values (outcomes). (d) The posterior (predictive) distribution is biased, and the posterior variance overestimates uncertainty (the true values fall near the centre of the posterior distribution more often than expected). (e) The posterior (predictive) distribution is biased, and the posterior variance underestimates uncertainty (the true values fall outside the posterior distribution more often than expected).

agree, the true value should fall below 1% of the posterior cdf 1% of the time, 2% of the cdf 2% of the time, and so forth, generating a linear P–P plot with a 1:1 slope (Fig. 1a). P–P plots differ from other measures of performance, such as the magnitude of uncertainty, in that one can have high MSE (uncertainty) but good P–P plots (and vice versa). If the posterior distribution agrees with the true values, then decision theory operations should also be valid. In contrast, where the posterior distribution is biased (Fig. 1b–d), decision theory conclusions based on the posterior also should be biased. Other approaches also exist to assess model performance, such as posterior predictive checks (PPC), comparing model output to the actual data (Gelman *et al.*, 1996) or comparison of posterior mean or mode to the true parameter value, providing a point estimate of location bias. In contrast, P–P plots examine bias across the entire posterior distribution, including whether the posterior variance (uncertainty) is overestimated or underestimated. Additionally, theoretical P–P plots are complementary to empirical PPC, which are applied to statistical analysis on a single data set. Note that each PPC could be viewed as providing a single estimate contributing to a theoretical P–P plot (thousands of such estimates would underlie a P–P plot). To illustrate, while PPC may show that the probability of the model generating the data was only 0.01, suggesting possible model misspecification, P–P plots examine whether that result in fact occurs 1% of the time, as predicted. While P–P plots would be difficult to generate for real ecological systems, we can at least assess the theoretical behaviour of a given model using simulations.

In this manuscript, we examine the minimum number of observations needed for six common models. We start by examining two commonly used probability density functions, the normal and beta distributions. Probability distributions are the foundation of virtually all stochastic models, from which quantitative risk estimates can be derived. Next, we examine two general statistical approaches – multiple linear regression and hierarchical (mixed-model) linear regression. These are standard approaches used in ecology and therefore are good comparison models. Finally, we examine two relatively simple models used in biological conservation. The first estimates the effectiveness of a management strategy: ballast water exchange to reduce the numbers of exotic organisms transported (Bailey *et al.*, 2011). The original article (Bailey *et al.*, 2011) presented data on four ships and three estimates of density per ship before and after exchange. In this article, we examine the theoretical expectations for the approach presented in Bailey *et al.* (2011). Finally, we examine population viability analysis (PVA), which has been argued to require extensive data (Coulson *et al.*, 2001). Here, we explore how much data are requirement for PVA.

METHODS

We used a Bayesian framework for our analyses (see Lee, 2004 for the basics of Bayesian statistics and Hilborn &

Mangel, 1997 for discussion in an ecological context). We used vague priors (diffuse and not based on previous data or subjective elicitation), as we were particularly interested in the conditions of limited information. However, our findings are extendable to informative priors. We also recognize that there are deep philosophical debates regarding the choice of vague priors in Bayesian analyses (Gelman, 2006), and that these choices can change outcomes (Van Dongen, 2006). However, our interest here was from a practical point of view of risk assessment; thus, we used two metrics of performance—posterior MSE and P–P plots as indicators of agreement between probabilistic predictions and true values, given the data and choice of priors. Such practical considerations are particularly important as models become more complex and specialized (as typically the case in ecology), and the consequences of the priors less studied. We examined three vague, improper priors: uniform, where each parameter (x) value a priori is equally likely [$p(x) \propto 1$], uniform on a log scale [i.e. $p(x) \propto 1/x$], where values between 1–10 and 10–100 and so forth are equally likely and an intermediate prior [$p(x) \propto 1/x^{0.5}$], which is uniform on the square root of the parameter value. The second and third priors are only valid for positive values, which may be particularly relevant for scale parameters (e.g. variance), which are also valid for the same range.

We simulated parameter values in three ways: from a (1) fixed point, (2) bounded uniform distribution and (3) normal distribution (truncated at invalid values, for example, negative values for variance). Method one reflected the scenario of repeatedly applying risk assessment to different realizations of the same system (i.e. if ‘reality’ was repeatedly generated from identical underlying processes). Methods two and three would be analogous to conducting the risk assessment procedure across numerous different systems (such as applying a risk assessment protocol to many different invasive species). Conclusions did not change for the methods of simulating parameter values, and thus, we only presented method three – generating parameter values from a normal distribution. Importantly, the parameter values were generated using distributions that were *different* than the prior distributions used in the Bayesian analyses, thus simulating the condition where prior knowledge is limited and ensuring that required sample sizes were not an artefact of the prior resembling the generating distribution.

For each of our models (discussed below), at different sample sizes (1–30), we ran 1000 simulations. For each simulation, we generated parameter values, which in turn were used to generate dynamics of the system and realisations of data. The data were then used in the Bayesian analysis. Specifically, for each simulated ‘reality’, we conducted a Markov Chain Monte Carlo (MCMC) procedure (using Metropolis–Hastings algorithm, Gilks *et al.*, 1996), with 30,000 iterations (and a 2000 iteration burn in) to numerically estimate the posterior distribution. We tested for convergence using Gelman–Rubin diagnostic, using $R < 1.05$ as a threshold for convergence (Gelman & Rubin, 1992). If chains did not converge,

we ran an additional 30,000 iterations per chain. If chains did not converge after nine attempts, we discarded that simulated ‘reality’ and tried again with new parameter values (this occurred < 3% of the time, for the most complex model, which was a function of the automation process).

To examine the metrics of performance, we thinned the MCMC chain by recording every tenth interval. To estimate the magnitude of uncertainty, we examined the posterior MSE (i.e. average squared deviation between estimate and true parameter value). To estimate bias, we compared the posterior distribution to true values using P–P plots. The process was as follows: the MCMC algorithm generated a set of parameter values (chosen from a proposal distribution), and either accepted or rejected the proposed parameters, based on the relative probability that the proposal was the set of ‘true’ generating parameter values, given the model. This relative probability was proportional to the likelihood multiplied by the prior probability of the parameter set (Gilks *et al.*, 1996). Thus, each MCMC iteration yielded a set of estimated parameter values. Taken across MCMC iterations, the relative frequencies of occurrences of each set of parameter values (divided by the total number of iterations) approximated the posterior probability distribution. Thus, we could simply use the fraction of MCMC iterations below parameter value θ as an estimate of the posterior cumulative distribution function (cdf) up to θ . Because we knew the true generating values (given that these were simulations), we could compute the posterior percentile of the true value using the MCMC estimated cdf. Repeated over 1000 simulated realities, if the posterior distribution agreed probabilistically with the true values, the true value should fall below 1% of the cdf only 1% of the time, 2% of the cdf only 2% of the time, and so forth, generating a linear P–P plot with a 1:1 slope (Fig. 1a).

We analysed the metrics of performance directly in terms of parameter values (posterior distribution, procedure described previously), as well as predicted outcomes (posterior predictive distribution). Analyses of predicted outcomes proceeded as follows: for each ‘reality’, a new outcome was generated using the true generating parameters. An outcome was then also generated using the parameter set from each iteration of the MCMC run. We then followed an analogous procedure to the one above, comparing the ‘real’ outcome against the cdf generated from the posterior predictive distribution. Predicted outcomes are directly of interest (e.g. the magnitude of habitat loss). Parameter values are also important because we may be interested in a variety of outcomes, which are ultimately dependent on the parameters underlying the model.

Models examined

We examined six models based on simple distributions, general statistical tests and ecological models (see Table 1 for model parameters and values used). For each model, we described the generating model, which produced data points,

and the statistical estimation model. All models were written in C++, using the GNU scientific library (GSL).

General distributions

We examined the Normal and Beta distributions, which are commonly used in ecology and have substantially different properties. These form the basis for the error functions and likelihoods of many complex models and thus were obvious starting points for analyses.

Model 1: Normal distribution.

Generating model.

$$Y_i \sim N(\mu, \sigma) \tag{1}$$

Fitting model.

$$\pi(\mu, \sigma | \vec{Y}) \propto pr(\mu)pr(\sigma) \prod_{i=1}^n f(Y_i | \mu, \sigma) \tag{2}$$

where \vec{Y} was a vector of data points (Y_i was individual point i), generated from a normal (N) distribution with mean μ , standard deviation σ and number of data points n . The posterior probability distribution $\pi(\cdot)$ estimated the probability of the parameter values, given the observed data and priors, and $f(\cdot)$ denoted the relevant probability density function.

Model 2: Beta distribution.

Generating model.

$$Y_i \sim \text{Beta}(A, B) \tag{3}$$

Fitting model.

$$\pi(A, B | \vec{Y}) \propto pr(A)pr(B) \prod_{i=1}^n f(Y_i | A, B) \tag{4}$$

where A and B were parameters defining the shape of the beta distribution. The beta distribution, which ranges from zero to one, is useful for modelling probabilities or proportions. Other notation was as above.

General statistical approaches

Model 3: Multiple regression.

Generating model.

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_m X_{m,i} + \epsilon_i \tag{5}$$

$$\epsilon_i \sim N(0, \sigma) \tag{6}$$

Fitting model.

$$\pi(\sigma, \vec{\beta} | \vec{Y}, \mathbf{X}) \propto pr(\vec{\beta})pr(\sigma) \prod_{i=1}^n f(Y_i | \vec{X}_i, \vec{\beta}, \sigma) \tag{7}$$

Table 1 List of models, parameters, ‘true’ parameter ranges used to generate ‘realities’ [mean (standard deviation)], vague priors used to generate Figs 2–4, and number of observations required for the posterior distribution to accurately reflect the probabilities of generating values.

Model	Parameter	Simulated values	Priors	Sample size required	Maximum deviation
Normal distribution	μ (mean)	100 (20)	$\text{pr}(\mu) \propto 1$	$n = 2$	0.022
	σ (standard deviation)	30 (10)	$\text{pr}(\sigma) \propto 1/\sigma$		0.025
	Outcome				0.030
Beta distribution	A (shape)	5 (2)	$\text{pr}(a) \propto 1/a$	$n = 2$	0.078
	B (shape)	50 (10)	$\text{pr}(b) \propto 1/b$		0.073
	Outcome				0.028
Multiple regression (three independent variables)	β_0 (intercept)	10 (3)	$\text{pr}(\beta_0) \propto 1$	$n = 5$	0.024
	$\beta_1 \dots \beta_m$ (slopes)	10 (3)	$\text{pr}(\beta) \propto 1$		0.028
	σ (residual SD)	30 (10)	$\text{pr}(\sigma) \propto 1/\sigma$		0.030
	Outcome				0.042
Hierarchical regression	β (common slope)	10 (3)	$\text{pr}(\beta) \propto 1$	$n_1 = 3, n_2 = 2$	0.053
	σ_γ (standard deviation of slopes)	3 (1)	$\text{pr}(\sigma_\gamma) \propto 1/\sigma_\gamma^{0.5}$		0.078
	σ_ϵ (residual SD)	10 (3)	$\text{pr}(\sigma_\epsilon) \propto 1/\sigma_\epsilon$		0.026
	Outcome				0.081
	β			OR	0.025
	σ_γ			$n_1 = 10, n_2 = 1$	0.054
	σ_ϵ				0.067
Ballast water management effectiveness	A (shape from beta distribution)	4 (1)	$\text{pr}(a) \propto 1/a$	$n_1 = 3, n_2 = 2$	0.085
	B (shape from beta distribution)	20 (5)	$\text{pr}(b) \propto 1/b$		0.085
	$\lambda_{1,\cdot}$ (initial log-normal density)	100 (15)	$\text{pr}(\lambda_{1,\cdot}) \propto 1$		0.058
	Outcome				0.039
	A			OR	0.064
	B			$n_1 = 20, n_2 = 1$	0.088
	$\lambda_{1,\cdot}$				0.097
Population viability analysis	Outcome				0.044
	r (intrinsic growth rate)	1.6 (0.2)	$\text{pr}(r) \propto 1$	$T = 7, n_1 = 2, n_2 = 2$	0.043
	k (carrying capacity; scaled to r)	$0.7(0.1) * N_{0r}^T$	$\text{pr}(k) \propto 1/k$		0.14
	σ_γ (environmental stochasticity)	0.3 (0.1)	$\text{pr}(\sigma_\gamma) \propto 1/\sigma_\gamma^{0.5}$		0.066
	σ_ϵ (demographic stochasticity)	3 (1)	$\text{pr}(\sigma_\epsilon) \propto 1/\sigma_\epsilon^{0.5}$		0.092
	Outcome				0.029
	r			OR	0.074
	k			$T = 30, n_1 = 1, n_2 = 1$	0.13
σ_γ				0.13	
σ_ϵ				0.16	
Outcome				0.050	

Maximum absolute deviation between posterior percentile and cdf of true values was also presented [i.e. $\text{abs}(x-y)$ on P–P plot, Figs 1–3]. ‘Outcome’ denotes comparison between the posterior predictive distribution and new data points generated using the ‘true’ models. For hierarchical models, two sets of analyses were presented with different replications at different levels of the hierarchy.

$$\begin{aligned} \mu_i &= \vec{X}_i^T * \vec{\beta} \\ f(Y_i | \vec{X}_i, \vec{\beta}, \sigma) &= f(Y_i | \mu_i, \sigma) \end{aligned} \tag{8}$$

where $\vec{\beta}$ was the vector of parameters (slopes and intercept), $\beta_0 \dots \beta_m$, \mathbf{X} was the matrix of predictor values, \vec{X}_i was the vector of predictors for individual i , $1, X_{1,i}, \dots, X_{m,i}$, ϵ_i was the unexplained variation, modelled as a normal random deviate with standard deviation σ (i.e. same error distribution as equation 2). Other notation was as above. For our analysis, we simulated three independent variables ($m = 3$).

*Model 4: Hierarchical (mixed-model) regression.
Generating model.*

$$Y_{i,j} = (\beta + \gamma_i)X_{i,j} + \epsilon_{i,j} \tag{9}$$

$$\gamma_i \sim N(0, \sigma_\gamma) \tag{10}$$

$$\epsilon_{i,j} \sim N(0, \sigma_\epsilon) \tag{11}$$

Fitting model.

$$\begin{aligned} \pi(\sigma_\epsilon, \sigma_\gamma, \beta | \mathbf{Y}, \mathbf{X}) &\propto \text{pr}(\beta) \text{pr}(\sigma_\epsilon) \text{pr}(\sigma_\gamma) \\ &\prod_{i=1}^{n_1} \left(f(\gamma_i | \sigma_\gamma) \prod_{j=1}^{n_2} f(Y_{i,j} | X_{i,j}, \beta, \gamma_i, \sigma_\epsilon, \sigma_\gamma) \right) \end{aligned} \tag{12}$$

$$\begin{aligned} \mu_{ij} &= \beta X_{ij} + \gamma_i X_{ij} \\ f(Y_{ij}|X_{ij}, \beta, \gamma_i, \sigma_\varepsilon, \sigma_\gamma) &= f(Y_{ij}|\mu_{ij}, \sigma_\varepsilon) \end{aligned} \tag{13}$$

where \mathbf{X} and \mathbf{Y} were the matrices of predictor and response variables, respectively, β was the common slope, γ_i was the deviation from the common slope allowing each group i to have its own slope, ε was the unexplained within group variation, j was an individual of the i th group, n_1 was the number of groups, n_2 was the number of individuals in each group, σ_γ was the standard deviation of slopes between groups and σ_ε was the standard deviation of errors ε within each group. For simplicity, we did not include the intercept, either in the generating or fitting models. Note that this model contains two sources of error, one for slopes and one for variation around the slope within each group. γ scaled with X , whereas ε did not.

Ecological models

Model 5: management effectiveness – ballast water exchange.

Generating model. We modelled the effectiveness of ballast water exchange on reducing the density of organisms. Ballast water exchange is one of the primary methods to reduce the introduction of exotic species (Locke *et al.*, 1993; Briski *et al.*, 2010). Logically, different ships (i) may have different levels of effectiveness (E_i – the proportion of organisms that survive treatment), which we modelled using the beta distribution. Our objective was to estimate effectiveness across all ships, given our observations – the numbers of individuals counted in a sample of water before and after exchange in n_1 ships (Bailey *et al.*, 2011). We assumed that observation error occurred, and for simplicity, that organisms were randomly distributed; we used a Poisson distribution to generate observation errors.

$$\begin{aligned} E_i &\sim \text{Beta}(A, B) \\ \lambda_{2,i} &= \lambda_{1,i} * E_i \\ E_i &= \lambda_{2,i} / \lambda_{1,i} \\ Y_{1,i,j} &\sim \text{Pois}(\lambda_{1,i}) \\ Y_{2,i,j} &\sim \text{Pois}(\lambda_{2,i}) \end{aligned} \tag{14}$$

Fitting model.

$$\begin{aligned} \pi(A, B, \lambda|\mathbf{Y}) &\propto pr(A)pr(B)pr(\lambda_{1,i}) \\ \prod_{i=1}^{n_1} f(\lambda_{2,i}|\lambda_{1,i}, E_i) f(E_i|A, B) &\prod_{j=1}^{n_2} f(Y_{2,i,j}|\lambda_{2,i}) f(Y_{1,i,j}|\lambda_{1,i}) \end{aligned} \tag{15}$$

where A and B were parameters that describe the beta distribution, E_i was the proportional reduction in living organisms on ship i (i.e. the ballast water exchange effectiveness), \mathbf{Y} were the matrix of observations across sampled ships (i) and samples within ships (j), before ($Y_{1,i,j}$) and after ($Y_{2,i,j}$) exchange, n_1 was the number of ships sampled, n_2 was the

number of samples per ship, $\lambda_{1,i}$ and $\lambda_{2,i}$ were the rate parameters (i.e. unobserved true number of organisms, per unit volume of water, residing in the ballast water before and after treatment, respectively).

From the generating model, it should be clear that $\lambda_{2,i} = E_i * \lambda_{1,i}$ and so we could disregard $f(\lambda_{2,i}|\lambda_{1,i}, E_i)$ (equation 15). If we also used a uniform prior for $pr(\lambda_{1,i})$, equation 15 could be simplified to equation 16, which has the same formulation as the beta distribution (equation 4) multiplied by the probability density functions due to observation error.

$$\begin{aligned} \pi(A, B, \lambda|\mathbf{Y}) &\propto pr(A)pr(B) \\ \prod_{i=1}^{n_1} f(E_i|A, B) &\prod_{j=1}^{n_2} f(Y_{2,i,j}|\lambda_{2,i}) f(Y_{1,i,j}|\lambda_{1,i}) \end{aligned} \tag{16}$$

Model 6: Population viability analysis.

Generating model. Population viability analysis can be used to predict both species extinction (Brook *et al.*, 2000) and species invasions (Anderson, 2005). We used a simple logistic growth model incorporating both environmental and demographic stochasticity (Ovaskainen & Meerson, 2010). Each has a different effect on the dynamics of the system (Dennis, 2002). We estimated population rates using time-series of simulated population density data, which were modelled with observation error. This was also a state-space model, given that we made inferences on unknown states based on observations, states were interlinked and dependent on one another, and states were modelled as parameters (Hinrichsen, 2009). In notation,

$$\begin{aligned} X_{t,i} &= X_{t-1,i}(1 + r(1 - X_{t-1,i}/k) + \gamma_t + \varepsilon_{t,i}) \\ \gamma_t &\sim N(0, \sigma_\gamma) \\ \varepsilon_{t,i} &\sim N(0, \sigma_\varepsilon / \sqrt{X_{t-1,i}}) \\ Y_{t,i} &\sim \text{Pois}(X_{t,i}) \end{aligned} \tag{17}$$

Fitting model.

$$\begin{aligned} \pi(r, k, \sigma_\gamma, \sigma_\varepsilon|\mathbf{Y}) &\propto pr(r)pr(k)pr(\sigma_\gamma)pr(\sigma_\varepsilon) \\ \prod_{t=1}^T \left(f(\gamma_t|\sigma_\gamma) \prod_{i=1}^{n_1} \left(f(X_{i,t}|X_{i,t-1}, r, k, \gamma_t, \sigma_\gamma, \sigma_\varepsilon) \prod_{j=1}^{n_2} f(Y_{t,i,j}|X_{t,i}) \right) \right) \end{aligned} \tag{18}$$

$$\begin{aligned} \mu_{t,i} &= X_{t-1,i} * (1 + r(1 - X_{t-1,i}/k) + \gamma_t) \\ \sigma_c &= \sigma_\varepsilon \sqrt{X_{t-1,i}} \\ f(X_{t,i}|X_{t-1,i}, r, k, \gamma_t, \sigma_\gamma, \sigma_\varepsilon) &= f(X_{t,i}|\mu_{t,i}, \sigma_c) \end{aligned} \tag{19}$$

where r was the intrinsic growth rate, k was the carrying capacity, $X_{t,i}$ was the true size of subpopulation i at time t , $Y_{t,i,j}$ was the sample estimate of population size, σ_γ and γ_t were the standard deviation and realization of environmental stochasticity, respectively, and $\varepsilon_{t,i}$ was the realization of demographic stochasticity. Note that both $X_{t,i}$ and γ_t were

treated as parameters as well as states. Further note the similarity between this model and the hierarchical regression model above (equation 12). γ_t and $\varepsilon_{t,i}$ were both sources of variation and differed to each other in their relation with X . In this case, the variance in $X_{t,i}$ caused by γ_t scaled with the squared term $(X_{t-1,i})^2$, whereas the variance caused by $\varepsilon_{t,i}$ scaled with $X_{t-1,i}$ because the standard deviation of $\varepsilon_{t,i}$ was $\sigma_\varepsilon/\sqrt{X_{t-1,i}}$. These differences in scaling were theoretically expected for environmental γ_t and demographic stochasticity $\varepsilon_{t,i}$, such that demographic stochasticity has a diminishing per capita effect as population size increases, whereas the per capita effect of γ_t does not decline with population size (Dennis, 2002). Second, note that we used a hierarchical model with multiple estimates $\varepsilon_{t,i}$ nested within every estimate of γ_t . In reality, such subpopulations may be rare (but see Hinrichsen, 2009), but was important for fitting the model as we show below.

RESULTS

Across all models, few observations were needed to achieve a 1:1 line on the P–P plot, with a maximum absolute difference < 0.1 (10%), between posterior percentile and cdf of true values ($\max(|x-y|)$ on P–P plot), except for carrying capacity in PVA (Table 1). Moreover, difference in sample size of even a single data point changed the P–P plot from a 1:1 line to one showing strong bias (compare Figs 2a–c & 4e–g). For non-hierarchical models, bias in the posterior distribution was low when the number of data points equalled the number of parameters estimated (Fig. 2g–j). Thus, for the normal and beta distributions, only two data points were needed, and for a multiple regression with three independent variables (and intercept and variance), five data points were needed. Note that using a different vague prior could require substantially larger sample sizes (Appendix S1 in Supporting Information).

For the hierarchical models, such simple relations between data requirements and number of parameters were no longer as straightforward, although the number of data points required remained reasonably small. For the simple hierarchical regression (model 4), one could obtain relatively unbiased posterior (and posterior predictive) distributions with only three groups and two data points per group (Fig. 2k–n, Table 1); here, six parameters were estimated ($\beta, \sigma_\gamma, \sigma_\varepsilon$, and an estimate of γ_i for each group, equation 12). Alternatively, it was possible to obtain an unbiased posterior distribution, even if only a single measurement was taken per group, if 10 groups were measured (Fig. 2o–r, Table 1). Here, 13 parameters were fit.

For the management model (model 5), only three ships and two samples per ship were needed for relatively unbiased estimation of densities of organisms within each ship and extrapolation to the distribution of ballast water exchange effectiveness across all ships in the population (Fig. 3a–e, Table 1). Eight parameters were estimated (A, B , and the true density before ($\lambda_{1,j}$) and after ($\lambda_{2,i}$) exchange, for each ship (equation 16). In contrast, approximately 20 ships (42 parameters estimated)

were required when only one sample was taken per ship (Fig. 3f–j, Table 1).

For the PVA model (model 6), the posterior distribution was relatively unbiased with 7 years, two subpopulations and two measures per subpopulation (Fig. 3k–o, Table 1), requiring 24 parameters to be estimated ($r, k, \sigma_\gamma, \sigma_\varepsilon$, and each state $X_{i,t}$ and the realization of environmental stochasticity γ_t for $T-1$ years, equation 20). However, carrying capacity k had higher bias (14%, Table 1), possibly because population size sometimes did not reach k . When only one subpopulation was measured and only one replicate measure taken per time interval, there was more difficulty resolving demographic from environmental stochasticity, and bias existed (13–16%), even with 30 years of data (63 parameters were estimated; Fig. 3p–t, Table 1). The ecological relevance of such bias would be case specific. However, we note that the posterior predictive distribution was more robust for PVA (Fig. 3t, Table 1).

If the data available are insufficient to generate an unbiased posterior distribution, it may still be worthwhile to conduct statistical analyses. To facilitate convergence for this analysis, we put diffuse bounds on the priors, equal to 100 times the standard deviation of the generating distribution of true parameter values (Table 1). For illustration, we used the normal distribution (compare Fig. 4a–c, e–g with Fig. 2a–c, analyses with other models yielded the same conclusions). Caution is certainly warranted below the sample sizes identified in Figs 2 and 3. However, there can be large benefits in terms of reducing the magnitude of uncertainty (posterior MSE) particularly at low sample sizes (Fig. 4i–k). Additionally, we could identify which combinations of parameter values were unlikely, using a single data point, and thereby exclude parameter space (Fig. 4d versus h).

DISCUSSION

Risk analyses for conservation issues typically are data poor, and yet decisions need to be made in a timely manner. It is in this context of limited time, information and resources that we structure four arguments. First, we argue that statistical analyses on even a few data points can be worthwhile. We recognize that rules of thumb on minimum sample sizes pervade conventional wisdom (e.g. at least 5–10 data points required per independent variable, Bartlett *et al.*, 2001). While these may be relevant from an academic perspective, where time exists for further research, in our view, they are not relevant from a risk analysis perspective, where the alternative is to make decisions in the absence of data. Moreover, our results demonstrated that the posterior (predictive) distribution probabilistically reflects the true model (outcomes) and can be valid for risk analysis, even with few data points. Of course, the exact degree of acceptable error will be case specific, depending on the cost of those errors.

Given these rules of thumb and real-world limitations in time and data, it is not surprising that risk analyses are often conducted purely using expert opinion (e.g. Burgman, 2005). We are generally supportive of using expert opinion, but

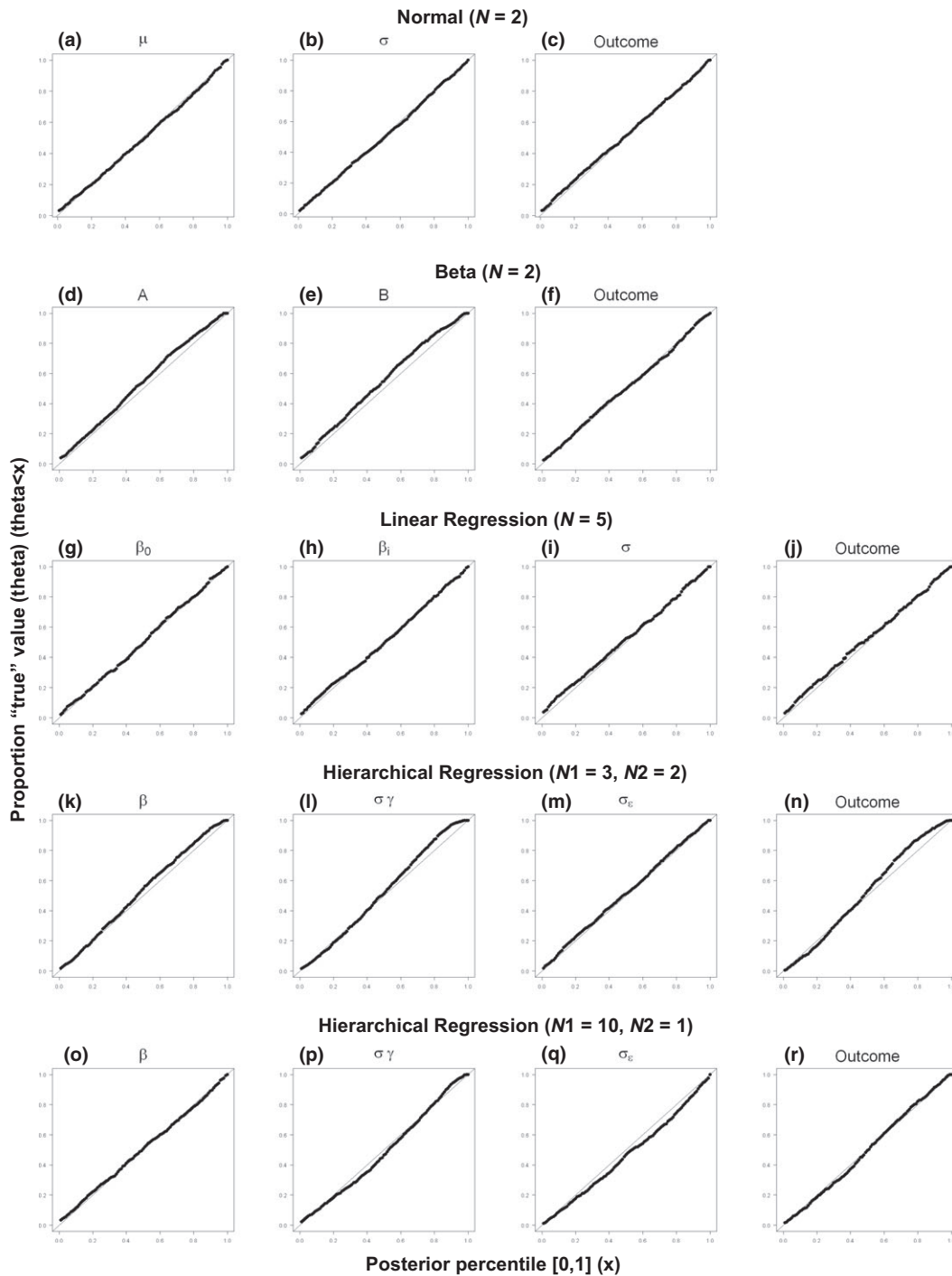


Figure 2 Simulated results of bias in the posterior distribution for four models: the normal distribution (panels a–c, equations 1 and 2), the beta distribution (panels d–f, equations 3 and 4), multiple regression with three predictor variables (panels g–j, equations 5–8) and hierarchical regression using two sample points per group (panels k–n, equations 9–13). And hierarchical regression using only one sample point per group (panels o–r). Sample sizes (n) required are shown for each panel in parentheses (see also Table 1). For the hierarchical model, n_1 denotes the number of groups, and n_2 denotes the sample size per group required. ‘Outcome’ denotes comparison between the posterior predictive distribution and new data points generated using the ‘true’ models. Compare with Fig. 1 for expected patterns given different biases, Fig. 4 for illustration of biases with lower sample sizes and Appendix S1 for illustration of the consequences of using different priors.

with cautious interpretation. The consequences might best be conceptualized through a thought experiment: if we were to conduct a risk assessment protocol on 1000 invasive species

and evaluated the outcomes across species, would the distribution of actual outcomes follow predictions? If not, predictions are biased. This consideration most closely

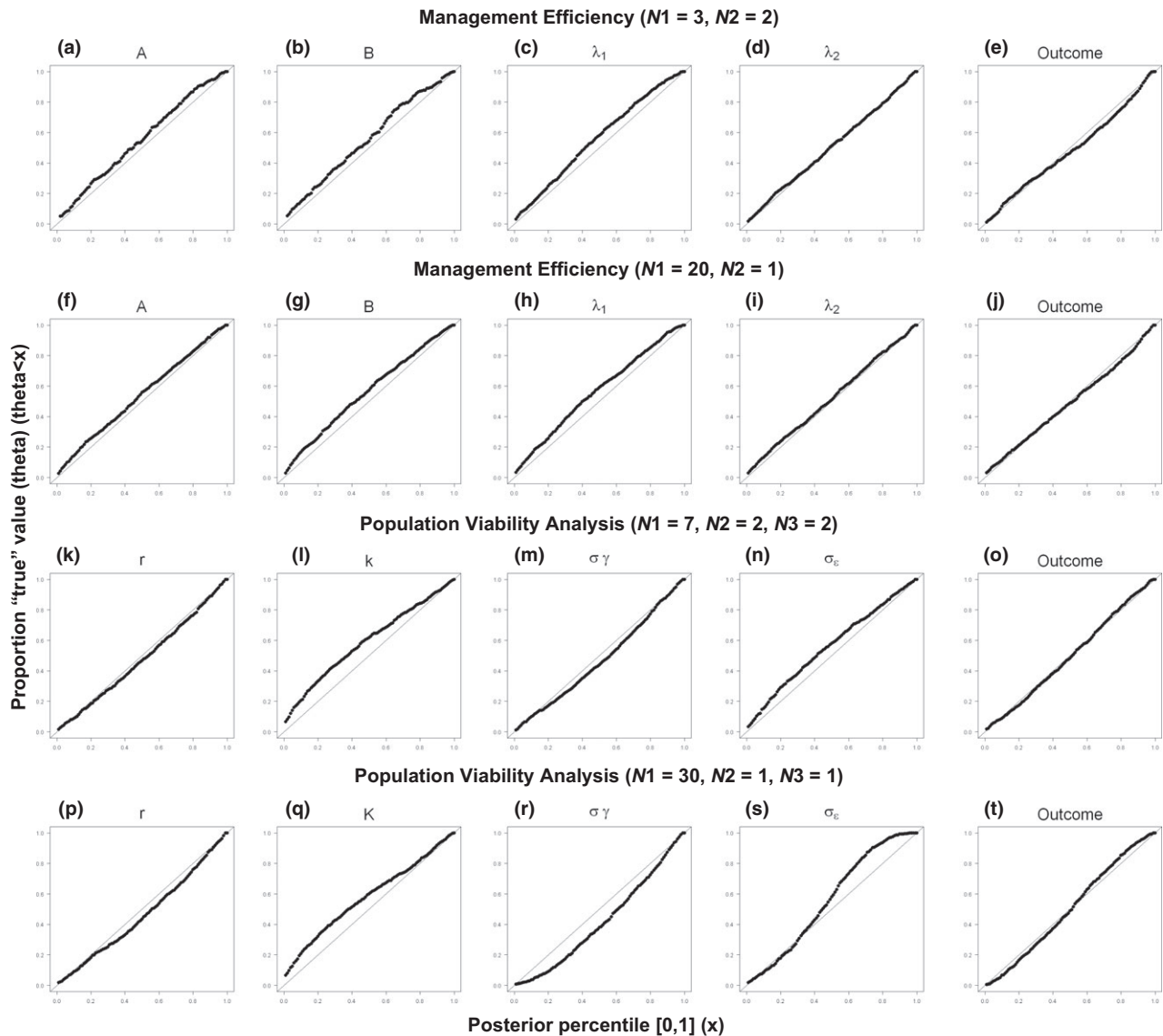


Figure 3 Simulated results of bias in the posterior distribution for four models: a ballast water management model with two samples taken per ship (panels a–e, equations 15–17), and with only one sample taken per ship (panels f–j), and a population viability analysis (PVA) model with two subpopulations and two samples per subpopulation (panels k–o), and with only one subpopulation and one sample per year (panels p–t). For the management model, n_1 is the number of ships, and n_2 is the number of samples per ship. For the PVA model, T is the number of years, n_1 is the number of subpopulations and n_2 is the number of samples per subpopulation per year. ‘Outcome’ denotes comparison between the posterior predictive distribution and new data points generated using the ‘true’ models. For the PVA model, a slight bias still existed even with a large data set (30 years), with a single subpopulation. Compare with Fig. 1 for expected patterns given different biases.

matches the output needed for risk assessment with the implicit logic of decision theory operations, such as integration across the posterior distribution, weighting by probability and costs which may be asymmetric (Moschini & Hennessey, 2001), or acceptance of some threshold of risk of ‘bad’ events occurring (Keller *et al.*, 2007). In the absence of data, the extent to which subjective beliefs are incorrect, not surprisingly, determines the extent to which the advice from decision theory also deviates from reality. Importantly, with only a few data points, these biases can be ameliorated. We note, however, that we do not explore model misspecification

in this article. Although important, model misspecification is a problem not just for Bayesian analyses, but for any parametric statistical approach. Further, model misspecification can remain a problem even with large sample sizes and thus is tangential to the purpose of this article, which is to identify when sample size should be limiting for a given model structure. Other considerations related to model misspecification are of course important as well, albeit beyond the scope of this article. For instance, analysis of the robustness of estimators (Berger, 1994), PPC (Gelman *et al.*, 1996) or measures of error rates in predictions (e.g. Boyce *et al.*, 2002)

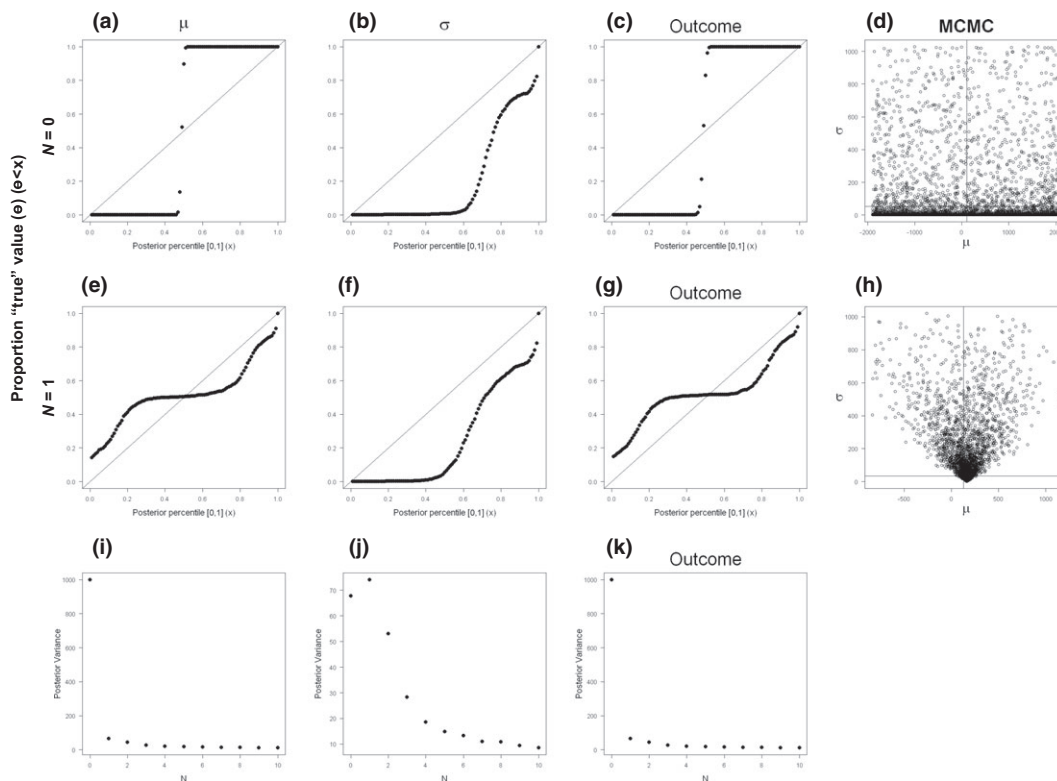


Figure 4 Analyses using no data (only priors, panels a–d) and a single data point (panels e–h) to estimate a normal distribution, and the posterior mean squared error (MSE; i.e. magnitude of uncertainty; panels i–k). Row $N = 0$ and $N = 1$ are for no data and a single data point, respectively. Compare with Fig. 1, to interpret biases. ‘Outcome’ denotes comparison between the posterior predictive distribution and new data points generated using the ‘true’ models. Panels (d, h) show the Markov Chain Monte Carlo (MCMC) estimates of parameter combinations across all iterations for one simulation, with μ on the x -axis and σ on the y -axis. Note the combinations of μ and σ excluded with even a single data point (compare panel d versus h). Priors were given diffuse bounds to facilitate convergence (see text).

and Bayesian model averaging (Hoeting *et al.*, 1999) all play a role in testing and managing model misspecification.

Our second argument is that even if one only has a single data point, and the posterior distribution will theoretically be biased, it may still be beneficial to consider statistical analysis. This is contrary to concepts of identifiability (Rannala, 2002) and requires a philosophical shift (although subjective Bayesians may have less difficulty, see Goldstein, 2006). While a single data point will not be sufficient to appropriately identify the value of a parameter, it may help rule out unlikely combinations of parameter values (Fig. 4g,h). Further, the magnitude of reduction in uncertainty is often greatest at the lowest sample sizes, with diminishing returns as sample size increases (Fig. 4e–f). Thus, if the options are to conduct a data-free analysis solely based on expert opinion or to incorporate a single empirical measurement in addition to expert opinion, the latter should be preferred. We note that within the Bayesian context, expert opinion can naturally be integrated with data, through the prior (Martin *et al.*, 2012).

Our third argument concerns a technical point related to Bayesian statistics. One of the main difficulties has been the choice of priors, especially for more complex hierarchical models (Gelman, 2006). There has been debate regarding

which priors should be chosen [e.g. robust to different formulations (Berger, 1994) and robust to arbitrary transformation (Bernardo, 2005)]. These choices can be important given that different priors can have strong effects on the posterior distribution particularly with small sample sizes, even when they are supposedly uninformative (Van Dongen, 2006). In this study, we do not make inferences about the general properties of vague priors, but instead focus on context-specific consequences of priors. This is particularly relevant given that applied ecological models are often constructed to characterize specific ecological systems and therefore may not have well-studied priors for those models. We argue that the fact that priors yield different results is not itself the most relevant consideration in conservation risk analyses. Clearly, in our results, the vague priors we examined differed substantially in their effect (compare Fig. 2 with Appendix S1 in Supporting Information), yet one was unbiased for our specific models, whereas others showed strong bias. As a practical decision in conservation biology, this can help us to exclude priors that do not work in theoretical simulations of a given system where we can control confounding factors, as it seems unlikely they will work in the real system. We note that while we focused on vague

priors, sometimes there may be good ecological reasons based on either logic or previous experience to use (weakly) informative priors (Martin *et al.*, 2012). Similar theoretical simulations could assess the effect of these informative priors given different scenarios as well.

Following from the above three arguments, our fourth argument is that the use of simulations as a means of theoretical evaluation should become standard practice in conservation biology. This is often carried out in statistics to test the behaviour of statistical models (e.g. Glass *et al.*, 1972; Berkof *et al.*, 2003; Lambert *et al.*, 2005). For simpler models (e.g. the first three examined here), analytic solutions using conjugate priors are well known (e.g. Lee, 2004). However, given the complexity and idiosyncrasy of many environmental problems, conjugate priors are often not known, behaviour of priors unexplored and biases analytically intractable. The simulation approach used here can provide a reasonably simple means to assay for bias across the entire posterior distribution, specifically for the numerical methods (MCMC), which are commonly employed in ecology. We argue that we should not simply assume that the statistical methods we use will yield the expected output for untested model structures. For instance, the adoption of Bayesian approaches without evaluation has been criticized (Lele & Dennis, 2009), and we agree with these concerns. Specifically in the context of risk analysis, we argue that theoretical simulations should be analysed using P–P plots.

In summary, risk analyses are constrained by information, time and resources. We should make our choice of analysis based upon the available alternatives, given that decisions need to be made despite limited information and high uncertainty. Thus, we argue that there are substantial advantages to incorporating even a few data points into risk analyses. There is no magic bullet to dispel uncertainty, beyond additional research. However, we may choose analyses that minimize the magnitude of uncertainty and/or bias, where the estimated distribution of outcomes theoretically matches the true probability of outcomes. Where necessary given limited data, we argue that we should philosophically shift from ‘what are the parameter values?’ to ‘which parameter values can we exclude?’ Thus, even with only a single data point, such that ecological models must necessarily be over-parameterized, advantages may still exist compared with data-free options. In combination, our suggestions are aimed to promote rapid implementation, and the most powerful analyses possible given available information, time and resources.

ACKNOWLEDGEMENTS

We thank the National Sciences and Engineering Research Council (NSERC) discovery grant to BL.

REFERENCES

- Anderson, M.C. (2005) Potential applications of population viability analysis to risk assessment for invasive species. *Human and Ecological Risk Assessment*, **11**, 1083–1095.
- Bailey, S.A., Deneau, M., Jean, L., Wiley, C., Leung, B. & MacIsaac, H.J. (2011) Has management reduced the risk of ballast-mediated species introductions to the Great Lakes? *Environmental Science and Technology*, **45**, 2554–2561.
- Bartlett, J.E., Kotrlik, J.W. & Higgins, C.C. (2001) Organizational research: determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, **19**, 43–50.
- Berger, J.O. (1994) An overview of robust Bayesian analysis. *Test*, **3**, 4–124.
- Berkof, J., van Mechelen, I. & Gelman, A. (2003) A Bayesian approach to the selection and testing of mixture models. *Statistica Sinica*, **13**, 423–442.
- Bernardo, J.M. (2005) Reference analysis. *Handbook of statistics 25: Bayesian thinking, modeling and computation* (ed. by D.K. Dey and C.R. Rao), pp. 17–90. North-Holland, Amsterdam.
- Boyce, M.S., Vernier, P.R., Nieson, S.E. & Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300.
- Briski, E., Bailey, S.A., Cristescu, M.E. & MacIsaac, H.J. (2010) Efficacy of ‘saltwater flushing’ in protecting the Great Lakes from biological invasions by invertebrate eggs in ships’ ballast sediment. *Freshwater Biology*, **55**, 2414–2424.
- Brook, B.W., O’Grady, J.J., Chapman, A.P., Burgman, M.A., Akcakaya, H.R. & Frankham, R. (2000) Predictive accuracy of population viability analysis in conservation biology. *Nature*, **404**, 385–387.
- Burgman, M.A. (2005) *Risks and decisions for conservation and environmental management*. University Press, Cambridge.
- Camerer, C.F. & Johnson, E.J. (1997) The process-performance paradox in expert judgment: how can experts know so much and predict so badly? *Research on judgment and decision making: currents, connections and controversies* (ed. by W.M. Goldstein and R.M. Hogarth), pp. 342–364. Cambridge University Press, Cambridge.
- Clark, J.S., Ferraz, G., Ogue, N., Hays, H. & Dicostanzo, J. (2005) Hierarchical bayes for structured, variable populations: from recapture data to life-history prediction. *Ecology*, **86**, 2232–2244.
- Coulson, T., Mace, G.M., Hudson, E. & Possingham, H. (2001) The use and abuse of population viability analysis. *Trends in Ecology and Evolution*, **16**, 219–221.
- Dennis, B. (2002) Allee effects in stochastic populations. *Oikos*, **96**, 389–401.
- Estévez, R.A., Walshe, T. & Burgman, M.A. (2013) Capturing social impacts for decision-making: a Multicriteria Decision Analysis perspective. *Diversity and Distributions*, **19**, 608–616.
- Finnoff, D., Shogren, J.F., Leung, B. & Lodge, D.M. (2007) Take a risk – preferring prevention over control of biological invaders. *Ecological Economics*, **62**, 216–222.
- Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **3**, 515–533.
- Gelman, A. & Rubin, D. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.

- Gelman, A., Ming, X.L. & Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**, 733–807.
- Gilks, W.R., Richardson, S. & Spiegelhalter, D.J. (eds.) (1996) *Markov Chain Monte Carlo in practice*. Chapman and Hall, London.
- Glass, G.V., Peckham, P.D. & Sanders, J.R. (1972) Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, **42**, 239–288.
- Goldstein, M. (2006) Subjective Bayesian analysis: principles and practice. *Bayesian Analysis*, **3**, 403–420.
- Hilborn, R. & Mangel, M. (1997) *The ecological detective: confronting models with data*. Princeton University Press, Princeton, NJ.
- Hinrichsen, R.A. (2009) Population viability analysis for several populations using multivariate state-space models. *Ecological Modelling*, **220**, 1197–1202.
- Hoeting, J.A., Madigan, D., Raftery, A.E. & Volinsky, C.T. (1999) Bayesian model averaging: a tutorial. *Statistical Science*, **14**, 382–417.
- Hosack, G.R., Hayes, K.R. & Dambacher, J.M. (2008) Assessing model structure uncertainty through an analysis of system feedback and Bayesian networks. *Ecological Applications*, **18**, 1070–1082.
- Keller, R.P., Drake, J.M. & Lodge, D.M. (2007) Fecundity as a basis for risk assessment of nonindigenous freshwater mollusks. *Conservation Biology*, **21**, 191–200.
- Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R. & Jones, D.R. (2005) How vague is vague? A simulation study of the impacts of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, **24**, 2401–2428.
- Lee, P.M. (2004) *Bayesian statistics: an introduction*. Wiley, West Sussex.
- Lele, S.R. & Dennis, B. (2009) Bayesian methods for hierarchical models: are ecologists making a Faustian bargain? *Ecological Applications*, **19**, 581–584.
- Leung, B., Roura-Pascual, N., Bacher, S., Heikkilä, J., Brotons, L., Burgman, M.A., Dehnen-Schmutz, K., Essl, F., Hulme, P.E., Richardson, D.M., Sol, D. & Vilà, M. (2012) TEASIng apart alien species risk assessments: a framework for best practices. *Ecology Letters*, **15**, 1475–1493.
- Locke, A., Reid, D.M., Van Leeuwen, H.C., Sprules, W.G. & Carlton, J.T. (1993) Ballast water exchange as a means of controlling dispersal of freshwater organisms by ships. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 2085–2093.
- Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M. & Mengersen, K. (2012) Eliciting expert knowledge in conservation science. *Conservation Biology*, **26**, 29–38.
- Moschini, G. & Hennessey, D.A. (2001) Uncertainty, risk aversion, and risk management for agricultural producers. *Handbook of agricultural economics, volume 1* (ed. by B. Gardner and G. Rausser), pp. 88–153. Elsevier Science B.V., Amsterdam.
- Ovaskainen, O. & Meerson, B. (2010) Stochastic models of population extinction. *Trends in Ecology and Evolution*, **25**, 643–652.
- Rannala, B. (2002) Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Systematic Biology*, **51**, 754–760.
- Shatkin, J.A. (2008) Informing environmental decision making by combining life cycle assessment and risk analysis. *Journal of Industrial Ecology*, **12**, 278–281.
- Van Dongen, S. (2006) Prior specification in Bayesian statistics: three cautionary tales. *Journal of Theoretical Biology*, **242**, 90–100.
- Yemshanov, D., Koch, F.H., Ducey, M. & Koehler, K. (2013) Mapping ecological risks with a portfolio-based technique: incorporating uncertainty and decision-making preferences. *Diversity and Distributions*, **19**, 567–579.

SUPPORTING INFORMATION

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

Appendix S1 Simulations using alternative vague priors to demonstrate the potential consequences on bias.

BIOSKETCHES

Brian Leung's research interests include ecological forecasting, risk analysis and decision theory particularly as applied to invasive species and ecosystem management, given limited information.

Russel Steele's research interests include Bayesian modelling and computational statistics, missing data and measurement problems.

Author contributions: BL and RS conceptualized the problem. BL performed the simulation analyses and wrote the manuscript with contributions from RS.

Editor: Mark Burgman