

Hierarchical Bayesian analysis of capture–mark–recapture data

E. Rivot and E. Prévost

Abstract: We present a hierarchical Bayesian model (HBM) for capture–mark–recapture (CMR) data analysis. It aims at estimating the probability of capture (θ_i) and the total population size (N_i) in a series of I years $i = 1, \dots, I$. The HBM assumes that the θ_i s and N_i s are sampled from a common probability distribution with unknown parameters. It is compared with the model assuming independence between years in the θ_i s and N_i s (ABM). We show how a transfer of information between years is organized by the HBM. We compare the merits of HBM vs. ABM to estimate the spawning run and smolt run of an Atlantic salmon (*Salmo salar*) population of the River Oir (France) over a period of 17 years. In the spawners case, yearly data are poorly informative. Consequently, the HBM greatly improves posterior inferences compared with the ABM in terms of dispersion and robustness to the choice of prior. In the smolts case, the HBM does not significantly improve inferences compared with the ABM because data are more informative. We discuss why hierarchical modeling should be recommended in any ecological study where the data are collected on several sampling units that share some common features.

Résumé : Nous proposons un modèle Bayésien hiérarchique (HBM) pour analyser des données de capture–marquage–recapture. Ce modèle permet d'estimer la probabilité de capture (θ_i) et la taille de la population (N_i) pour une série de I années $i = 1, \dots, I$. Le HBM suppose que les θ_i s et les N_i s sont issus a priori d'une même distribution de probabilité dont les paramètres sont inconnus. On le compare à un modèle qui suppose l'indépendance des θ_i s et des N_i s a priori (ABM). Nous montrons comment le HBM organise le transfert d'information entre les années. Nous comparons les avantages respectifs du HBM et du ABM pour estimer les migrations d'adultes et de smolts de la population de Saumon atlantique (*Salmo salar*) de la rivière Oir (France). Dans le cas des adultes, les données annuelles sont peu informatives. Le HBM améliore les inférences par rapport au ABM, en terme de dispersion et de robustesse au choix des distributions a priori. Dans le cas des smolts, le HBM n'améliore pas significativement les inférences par rapport au ABM car les données annuelles sont plus informatives. Finalement, nous recommandons d'utiliser un modèle hiérarchique pour analyser des données en écologie dans tous les cas où les données traitées concernent des unités qui partagent une caractéristique commune.

Introduction

The Bayesian approach has a great potential for the treatment of capture–mark–recapture (CMR) models. Although it is one of the main features of Bayesian data analysis, hierarchical modeling seems to have aroused little interest for CMR data. This paper investigates the potential of Bayesian hierarchical modeling applied to estimating the size of a population in a series of years, either consecutive or not, for which yearly CMR data have been collected under the same experimental design.

The Bayesian framework may significantly improve statistical inferences from CMR data analysis (Smith 1991; Pollock 1991; Garthwaite et al. 1995). Bayesian estimators perform better than classical ones in data-poor situations (Gazey and Staley 1986; Chao 1989). Indeed, classical point estimators are mostly based on asymptotic approximations. Therefore, bias can be relatively large when sample sizes are

small (Seber 1982). Confidence intervals around maximum likelihood estimates can also be inaccurate when they do not account for the skewness of the underlying likelihood distribution (Sprott 1981; Smith 1988). In contrast, Bayesian posterior inferences about the quantities of interest naturally reflect the entire shape of their distributions. The development of intensive computation methods now offers numerical procedures for solving a wide variety of inference problems within the Bayesian framework. Importance sampling or Markov chain Monte Carlo (MCMC) methods (Gelman et al. 1995) enable easy sampling from posterior distributions even in the most complex multiple CMR census models (George and Robert 1992; Basu and Ebrahimi 2001). By comparison, standard analytical approaches, such as maximum likelihood techniques, may become rapidly untractable when model complexity increases.

Relatively long but sparse (small sample size) series of data are quite common when dealing with CMR surveys

Received 4 June 2002. Accepted 30 October 2002. Published on the NRC Research Press Web site at <http://cjfas.nrc.ca> on 7 December 2002.
J16922

E. Rivot^{1,2} and E. Prévost. INRA, UMR Ecobiologie et qualité des hydrosystèmes continentaux, 65 rue de St Briec, CS 84215, 35042 Rennes CEDEX, France.

¹Corresponding author (e-mail: rivot@roazhon.inra.fr).

²On leave from ENGREF, Ecole Nationale du Génie Rural des Eaux et des Forêts, 19 av du Maine, 75732 Paris CEDEX 15, France.

aimed at estimating wild populations over a series of years. To analyze such data sets, one may first be tempted by considering the complete independence between years, i.e., a model where the CMR experiment performed each year is totally unrelated to those of the other years. At another extreme, one can ignore the between-year variability by pooling all years. An exchangeable hierarchical model is a sensible compromise between these radically contrasted views (Gelman et al. 1995). It assumes that the years are neither completely identical nor fully unrelated, but considers the yearly unknown variables as a random sample from a common probability distribution. In data-poor situations, the model assuming independence between years may lead to poor posterior inferences. Indeed, little information may be brought by the data in certain years, what may in turn result in highly imprecise or unreliable posterior distributions. Another serious drawback is that results may be highly sensitive to the prior choice (Gazey and Staley 1986; Chao 1989) and thus widely open to criticism because there is not a unique accepted way for assigning prior probabilities.

In this paper, we show two ways in which hierarchical modeling can significantly improve the inferences. First, the hierarchical approach can transfer valuable information from neighboring data to estimate the total population size in years for which CMR data are poor. Second, it offers a sensible solution to the critical problem of the sensitivity of the posterior inferences to the prior choice. Smith (1991) and George and Robert (1992) presented a hierarchical model for the catchability in the context of a multiple mark–recapture census of a closed population. However, the authors do not really discuss how hierarchical structure may improve the analysis of CMR data. Although he explicitly did not set his study in a CMR framework, Raftery (1988) already proposed a hierarchical model for N when it is estimated from a multiple Binomial sampling with a constant probability of “success”. He pointed out that the hierarchical structure is an interesting alternative to classical choices of discrete bounded prior on N . Indeed, it allows a convenient representation of a vague prior on N , avoiding the critical problem of the high sensitivity of inferences to the choice of the bounds. More recently, Su et al. (2001) applied hierarchical modeling to account for similarities in run timing among years for pink salmon (*Oncorhynchus gorbuscha*) escapement. The authors illustrated well how the hierarchical method significantly enhances the ability to estimate escapement for years where only few data are available.

We set a model that can accommodate any case where the CMR experiments are performed over a series of years following a homogeneous design similar to the two-sample Schnabel census for a closed population, i.e., the simple Petersen experiment (Seber 1982). We first concentrate on the conceptual framework of the hierarchical modeling without entering into technical details. The hierarchical model, jointly treating all years, appears as the most general model. We highlight that the model treating all years independently can be considered a particular case of the hierarchical model. We show how the between-year dependency introduced by the hierarchical assumption organizes a transfer of information between years. It enables one to make use of the full historical information to construct a modified prior distribution that should improve yearly inferences.

We apply our methodology to two contrasted numerical case studies. We treat two data sets of CMR surveys designed to estimate the spawning run and the smolt run of an Atlantic salmon (*Salmo salar*) population of the River Oir (France) over a period of 17 years. We focus on the comparison between the posterior inferences obtained under the hierarchical model and the model assuming between-year independence. Posterior inferences are carried out by Markov chain Monte Carlo (MCMC) methods (Gelman et al. 1995). As the problem of setting the priors is always crucial in Bayesian analysis, we carefully investigated the sensitivity of our results to the prior choices in an expanded appendix.

Materials and methods

CMR framework

We consider a data set from a population survey to estimate the size of the same population in a series of I years (typically I would be between 10 and 100). In the following, the subscript i stands for the year i . The data available come from CMR experiments analogous to the well-known two-stage Petersen experiment. A first-capture sample is carried out by means of a trap. Some of the individuals caught are marked and released. Recaptures of marked individuals are then carried out. Hypotheses needed to assume a Petersen-type model for a given year are supposed to be verified (they will be detailed later in the numerical examples). The unknown size of the population in year i is denoted N_i , and the CMR data collected for the same year are denoted d_i . We characterize the trap by a trapping efficiency, denoted θ_i , that can be interpreted as the probability of each individual to be caught. The entire series of unknown variables of interest are denoted $\theta = \{\theta_i\}_{i=1,\dots,I}$ and $N = \{N_i\}_{i=1,\dots,I}$ and data are denoted $\{d_i\}_{i=1,\dots,I}$.

The CMR data are used to make inferences on both the trap efficiencies θ and the population sizes N . In the following, we first develop an exchangeable hierarchical Bayesian model (HBM) that combines all years to derive joint inferences about the entire series (N, θ) . By jointly treating all of the years, the HBM provides a mean to estimate each (N_i, θ_i) by taking advantage of the information coming from the data of the other $I - 1$ years (denoted $\{1, \dots, I\}_{\neq i}$ in the following). Then, we derive a second model that can be considered a particular case of the HBM, referred as the annual Bayesian model (ABM). The ABM is less parsimonious than the HBM. It assumes that the θ_i s and N_i s are independent between years: posterior inferences about variables are derived independently for each year i using only the data from the year in turn. A completely pooled Bayesian model, ignoring the between-year variability, can also be considered an extreme particular case of an exchangeable model. It is the most parsimonious model but is of little interest and is therefore not further investigated.

The following notations hold in the sequel. $P(\cdot)$ denotes a conditional probability distribution and $\pi(\cdot)$ a prior distribution, with arguments determined by the context. Let us denote as X any random variable that we seek to estimate and as y the observations available to estimate X . The unnormalized Bayesian posterior density of $X|y$ is

$$(1) \quad P(X|y) \propto \pi(X)P(y|X)$$

Table 1. Hierarchical Bayesian model.

Notations

$N = \{N_i\}_{i=1,\dots,I}$, $\theta = \{\theta_i\}_{i=1,\dots,I}$, data = $\{d_i\}_{i=1,\dots,I}$

Hyperparameters $\phi = (\delta, \gamma)$

Joint prior distribution for (N, θ, ϕ)

Independence assumption

(T1.1) $\pi(N, \theta, \phi) = \pi(N, \delta) \pi(\theta, \gamma)$

Exchangeability assumption

(T1.2) $\pi(N, \theta, \phi) = \pi(\delta) \prod_{i=1}^I \pi(N_i | \delta) \pi(\gamma) \prod_{i=1}^I \pi(\theta_i | \gamma)$

Likelihood

(T1.3) $L(N, \theta, \phi | \text{data}) = L(N, \theta | \text{data}) = \prod_{i=1}^I L(N_i, \theta_i | d_i) = \prod_{i=1}^I L_i$

Posterior distribution

(T1.4) $P(N, \theta, \phi | \text{data}) \propto \pi(\gamma) \pi(\delta) \prod_{i=1}^I \pi(N_i | \delta) \pi(\theta_i | \gamma) L_i$

(T1.5) $P(N, \theta | \text{data}) \propto \int [\pi(N, \theta, \phi) L(N, \theta, \phi | \text{data})] d\phi$

Note: The notation $d\phi$ means that the integration is made with respect to the variable ϕ .

The symbol \propto means that both terms are proportional. The coefficient of proportionality is a normalizing factor ensuring that the posterior $P(X|y)$ integrates to one. It depends on y only and is thus a constant with regard to X . $P(X|y)$ is the sampling distribution, i.e., the joint distribution of the observations y conditionally on X . When regarded as a function of X , $P(X|y)$ is called the likelihood and is denoted $L(X|y)$.

Exchangeable HBM

The HBM assumes a hierarchical structure on both the trapping efficiencies and the total population sizes. This model combines all years together through a methodology of estimation that explicitly accounts for the similarities in the CMR experiments and for the dependence among the population sizes.

Basically, the hierarchical assumption is a conditional structure on the probability distribution (Gelman et al. 1995). Let us first consider the hierarchical hypothesis on the trapping efficiencies. The HBM assumes that the θ_i s are sampled from a common population distribution $\pi(\theta_i | \gamma)$, conditional on unknown hyperparameters γ . In turn, a prior $\pi(\gamma)$ is assigned to the hyperparameters. The θ_i s arise from analog experiments using the same trapping gear and the same experimental protocol. The hierarchical structure of the probability distribution sets the dependency between the θ_i s by expressing both similarity and heterogeneity among the θ_i s. It allows for between-year variations of the θ_i s resulting from unpredictable changes in environmental conditions or fish behavior (e.g., river discharge or temperature). The HBM considers γ as random with a prior that will be updated by the data of all years. This updating of the hyperparameters distribution (details follow) allows transferring of information between years.

A hierarchical structure is also imposed on the total population size $N = \{N_i\}_{i=1,\dots,I}$. The population investigated during the period $i = 1, \dots, I$ is the same population but observed

in different years. The size of the population is generated and constrained by some common ecological processes whatever the year. Thus, the estimation that one may derive for the years $\{1, \dots, I\}_{\neq i}$ are also partially informative with regards to the population size of a given year i . It is then appropriate to model a priori the N_i s as issuing from a common population distribution $\pi(N_i | \delta)$ with unknown random hyperparameters δ with their own prior $\pi(\delta)$.

The HBM jointly treats the entire time series $i = 1, \dots, I$. All variables (N_i, θ_i) depend on the hyperparameter vector denoted $\phi = (\gamma, \delta)$. The joint posterior $P(N, \theta, \phi | \text{data})$ in eq. 2 (detailed in Table 1) is obtained by the combination of the joint prior $\pi(N, \theta, \phi)$ and the likelihood term $L(N, \theta, \phi | \text{data})$

(2) $P(N, \theta, \phi | \text{data}) \propto \pi(N, \theta, \phi) L(N, \theta, \phi | \text{data})$

The definition of the joint prior $\pi(N, \theta, \phi)$ relies on two common hypotheses. We assume first the independence between (θ, γ) and (N, δ) in T1.1 and second the exchangeability of the θ_i s and the N_i s in T1.2 (Gelman et al. 1995). Let us first detail the exchangeability for the θ_i s. The joint prior for (θ, γ) reflects the dependence among the θ_i s. Exchangeability is a weaker concept than statistical independence. It is based on the hypothesis that before seeing the results of the CMR experiments, i.e., in the absence of data, there is no argument to distinguish a priori the θ_i s. Mathematically, exchangeability means that the joint distribution of the θ_i s remains unchanged by any permutation of the i indices. Unlike in a sequential Bayesian model, the order by which data have been gathered does not bear any significance. As suggested by Gelman et al. (1995), the most appropriate exchangeable distribution for (θ, γ) considers each θ_i as independently sampled from the conditional population distribution governed by γ , $\pi(\theta_i | \gamma)$. We made the same exchangeability assumption for the N_i s. Combined with the independence between (θ, γ) and (N, δ) , the exchangeability assumption leads to the joint prior in T1.2.

The likelihood term $L(N, \theta, \phi | \text{data})$ is the product of the yearly likelihood functions $L(N_i, \theta_i, \phi | d_i)$ denoted L_i in the following. The L_i s stem from the stochastic model underlying the sampling process of the CMR experiments. Thus, the likelihood does not depend on the hyperparameter vector ϕ , and the simplification in T1.3 holds. The sampling distributions that we use in this paper are the standard Binomial and Hypergeometric distributions (see, for instance, Gazey and Staley 1986). The likelihood functions and the underlying hypotheses are detailed in the section devoted to the numerical applications.

The resulting full joint posterior distribution of (N, θ, ϕ) is detailed in T1.4. The posterior distribution of the quantities of interest in T1.5 is obtained by integrating out the full joint posterior with respect to the hyperparameter vector ϕ (considered nuisance parameters).

ABM

The ABM that assumes the between-year independence of both the θ_i s and N_i s can be considered a particular case of the most general exchangeable hierarchical model. The ABM assumes a priori that the θ_i s and N_i s are fully unrelated. Under this hypothesis, the CMR data of the year i are only relevant to estimating the corresponding θ_i and N_i . The

Table 2. Hierarchical Bayesian model.

Predictive

Marginal predictive prior distribution for unknown observable variables $(\tilde{N}, \tilde{\theta})$

$$(T2.1) \quad \pi(\tilde{N}, \tilde{\theta}) = \int [\pi(N, \theta | \phi) \pi(\phi)] d\phi$$

Posterior predictive distribution conditional on the observed data

$$(T2.2) \quad P(\tilde{N}, \tilde{\theta} | \text{data}) = \int [\pi(N, \theta | \phi) P(\phi | \text{data})] d\phi$$

Marginal posterior distribution of year i

For each i in $1, \dots, I$, the marginal of the variables of interest writes

$$(T2.3) \quad P(N_i, \theta_i | \text{data}) = \int [P(N, \theta, \phi | \text{data})] d\{N_k, \theta_k\}_{k \neq i} d\phi$$

Extracting the term in “ i ” from the posterior $P(N, \theta, \phi | \text{data})$ yields

$$(T2.4) \quad P(N, \theta, \phi | \text{data}) \propto \pi(\phi) \pi(N_i, \theta_i | \phi) L_i \prod_{k \neq i} [\pi(N_k, \theta_k | \phi) L_k]$$

We denote as data_i^* the data of all years except i . The term L_i in T2.4 does not depend on ϕ . By extracting it from the integral on ϕ in T2.3, it becomes

$$(T2.5) \quad P(N_i, \theta_i | \text{data}) \propto \pi^*(N_i, \theta_i, \text{data}_i^*) L_i$$

where the term $\pi^*(N_i, \theta_i, \text{data}_i^*)$ equals

$$(T2.6) \quad \pi^*(N_i, \theta_i, \text{data}_i^*) = \int [\pi(N_i, \theta_i | \phi)] \int [\pi(\phi) \prod_{k \neq i} (\pi(N_k, \theta_k | \phi) L_k)] d\{N_k, \theta_k\}_{k \neq i} d\phi = \int [\pi(N_i, \theta_i | \phi) P(\phi | \text{data}_i^*)] d\phi$$

Note: Predictive distribution and expression of the marginal $P(N_i, \theta_i | \text{data})$. The notations $d\phi$ and $d\{N_k, \theta_k\}_{k \neq i}$ mean that the integrations are made with respect to the variable ϕ and $\{N_k, \theta_k\}_{k \neq i}$, respectively.

ABM still needs to assign a prior to each θ_i and N_i , but it no longer assumes random hyperparameters common to all years. By analogy with the HBM setting, those annual priors are denoted $\pi(N_i | \delta_i^0)$ and $\pi(\theta_i | \gamma_i^0)$, where the parameters $\phi_i^0 = (\delta_i^0, \gamma_i^0)$ are fixed to particular values, either equal or not between years. This impedes any transfer of information between years, and inferences about a particular year i fully ignore what can be learned from experiments carried out in the years $\{1, \dots, I\}_{\neq i}$. When the same population is surveyed during several years, assuming the between-year independence is counterintuitive and can result in a loss of information. In contrast, by structuring the between-year dependency, the HBM acknowledges that the CMR experiments of years $\{1, \dots, I\}_{\neq i}$ are partially informative with regards to the trapping efficiency and the population size of year i . The pattern of dependence and the amount of information that is transferred between years will be set a posteriori through the updating procedure of the hyperparameters.

One may also build “hybrid” models that make the independence assumption for only one of the two components θ and N , the other one still being modeled under a hierarchical structure. We compare inferences obtained under the HBM, the ABM, and the two “hybrid” models to investigate which of the two components, θ or N , is the most sensitive to independence vs. hierarchical modeling.

Choice of priors and sensitivity analysis

Setting the prior distributions is a critical point of any Bayesian analysis. All priors $\pi(\theta_i)$ and $\pi(N_i)$, conditional priors $\pi(\theta_i | \gamma)$ and $\pi(N_i | \delta)$, and hyperpriors $\pi(\gamma)$ and $\pi(\delta)$ used in numerical applications are detailed in Appendix A. We investigated how the posterior inferences obtained in the ABM

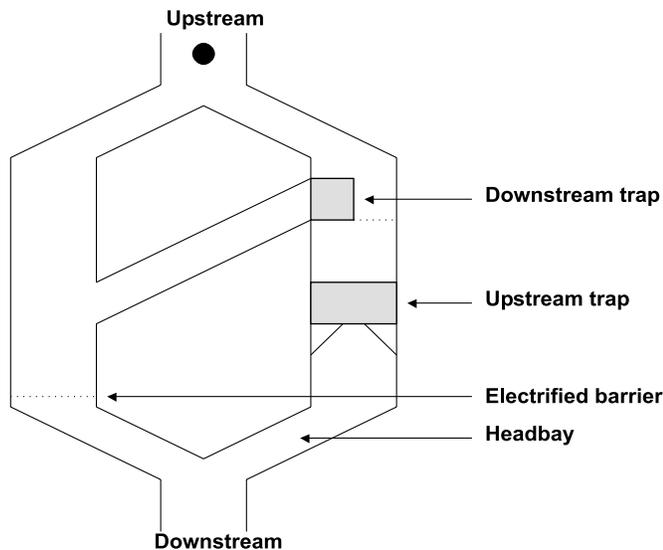
and HBM are sensitive to the choice of priors. This issue is also deferred to Appendix A.

Interpreting the HBM

Inferences about new observable variables denoted $(\tilde{N}, \tilde{\theta})$ can be derived through the predictive distribution conditional on the observed data (Gelman et al. 1995). It is an average of the conditional population distribution $P(N, \theta | \phi)$ over the posterior distribution of the hyperparameter ϕ conditioned on all observed data. The difference between the marginal prior in T2.1 and the posterior predictive in T2.2 reflects the amount of information brought by the data of all years to update the prior distribution common to all years (Table 2).

In the HBM, the data of all years are included to estimate the variables of interest (N_i, θ_i) for any particular year i . As shown in T2.3–T2.5, the marginal posterior $P(N_i, \theta_i | \text{data})$ in the HBM can be rewritten as the product of two terms: the likelihood term L_i and a complex integral term denoted $\pi^*(N_i, \theta_i, \text{data}_i^*)$. In this integral, all random variables except (N_i, θ_i) are integrated out, and the integration depends on data of years $\{1, \dots, I\}_{\neq i}$, denoted data_i^* . Thus, it can be considered a function of (N_i, θ_i) and data_i^* . The term $\pi^*(N_i, \theta_i, \text{data}_i^*)$ should be interpreted as a “modified” prior on (N_i, θ_i) . Equation T2.6 points out that $\pi^*(N_i, \theta_i, \text{data}_i^*)$ is the marginal prior of (N_i, θ_i) obtained by integrating over ϕ when ϕ is distributed as $P(\phi | \text{data}_i^*)$, i.e., the posterior obtained with data of each year k in $\{1, \dots, I\}_{\neq i}$. Note that the contribution of each of the years $\{1, \dots, I\}_{\neq i}$ to this modified prior is judiciously weighed according to L_k , i.e., according to how informative each year is about the population size and the trapping efficiency. It is through the modified prior $\pi^*(N_i, \theta_i, \text{data}_i^*)$ that the information from the data of years $\{1, \dots, I\}_{\neq i}$ is trans-

Fig. 1. Scheme of the Cerisel trapping facilities working as a double upstream and downstream partial counting fence. ● indicates the released site of tagged smolts and spawners. The released site and the downstream trap are 1 km away from each other.



ferred into the estimation of the variables (N_i, θ_i) . The higher the information carried by the posterior $P(\phi | \text{data}_i^*)$, the more informative the modified prior distribution $\pi^*(N_i, \theta_i, \text{data}_i^*)$.

In the ABM, the posterior of (N_i, θ_i) is the combination of L_i with a prior $\pi(N_i, \theta_i | \phi_i^0)$ where ϕ_i^0 is fixed to a particular value. Under such a model, the analyst often faces the dilemma between choosing ϕ_i^0 so that $\pi(N_i, \theta_i | \phi_i^0)$ is uninformative and taking the risk of unreliable inferences in data-poor situations, and choosing ϕ_i^0 that brings substantial information a priori and taking the risk of influencing the inferences by personal judgment not formally related to the data in hand. The HBM proposes an intermediate approach as a sensible solution to this dilemma by assigning higher degrees of credibility to some values of hyperparameters according to the data available.

Numerical examples

An Atlantic salmon population survey

We apply the framework of the previous section to a data set coming from the survey of the Atlantic salmon population on the Oir River (Normandy, France) since 1984. The Oir River is a spawning tributary of the Sélune River, which flows into the English Channel. The characteristics of the watershed are described in Baglinière et al. (1993). Spawners enter into Oir River to spawn during late fall and early winter (October–December). Smolts migrate out to the sea in the springtime (mainly in April). Both smolt and spawner runs are composed of several age classes (see Baglinière et al. 1993 and Prévost et al. 1996), but the age structure is not considered in this study.

The Cerisel trapping facility is located 2.3 km upstream from the confluence with the Sélune River and 12.3 km downstream from an impassable dam. The trapping facility consists of a double upstream and downstream partial counting fence (Baglinière et al. 1993) (Fig. 1). The inward and

Table 3. Capture–mark–recapture data for spawners by spawning migration year.

Year	c_i	x_i	m_i	r_i	rm_i
1984	167	13	154	22	12
1985	264	48	216	25	21
1986	130	37	93	9	5
1987	16	4	12	24	2
1988	226	43	183	12	12
1989	235	36	199	56	56
1990	15	8	7	17	2
1991	44	0	44	24	23
1992	31	11	20	9	4
1993	100	19	81	7	4
1994	32	14	18	5	1
1995	109	7	102	46	39
1996	70	15	55	82	25
1997	56	22	34	15	12
1998	34	4	30	36	6
1999	154	6	148	35	23
2000	53	0	53	37	4

Note: c_i denotes the number of spawners caught at the upstream trap; x_i denotes the number of spawners that died during manipulation or that are removed for experimental use or hatchery production; m_i is the number of spawners marked and released ($m_i = c_i - x_i$); r_i is the recapture sample; and rm_i the number of fish marked among r_i . See text for more details about the experiments.

outward salmon migrations to and from the entire area above the trap are monitored. The downstream trap is used to capture smolts during seaward migration but also spent adults after spawning. The upstream trap is used to capture spawners swimming upstream. Both upstream and downstream traps are operated every day during the migration time so that no pulse of migration is missed. Data currently available are from 1986 to 2001 for the smolt run and from 1984 to 2000 for the spawning run (Tables 3 and 4). They have been consistently gathered under a homogeneous experimental design through the entire study period.

The CMR experiments are different in their details for the smolts and the spawners, thus leading to distinct models. The two following sections describe the data, the stochastic models, and the subsequent sampling distributions that define the likelihood terms L_i (Table 5).

Spawning run

For each year i from 1984 to 2000, c_i denotes the number of spawners caught at the upstream trap. A number x_i of fish trapped is not released upstream either because they die during manipulation or because they are removed for experimental use or hatchery production. m_i denotes the number of fish marked and released ($m_i = c_i - x_i$). The fish released upstream from the trap are individually tagged. The recapture sample is gathered during and after spawning. Let us denote as r_i the sum of all those fish recaptured or observed (see thereafter) and as rm_i the fish marked among r_i . Our model leading to the likelihood L_i in T5.3 requires a few somewhat simplifying hypotheses (H1–H4). H1: all of the N_i spawners are assumed to be independently and equally catchable in the trap, with a probability θ_i . θ_i is considered constant over the migration season. H2: the spawner population upstream

Table 4. Capture–mark–recapture data for smolts by migration year.

Year	c_i	m_i	rm_i
1986	887	135	91
1987	283	31	24
1988	307	59	43
1989	553	65	43
1990	746	38	35
1991	151	35	27
1992	580	50	43
1993	209	26	24
1994	329	17	10
1995	618	63	53
1996	767	76	58
1997	205	63	31
1998	511	63	31
1999	195	59	45
2000	1849	300	232
2001	688	264	123

Note: c_i denotes the number of smolts caught at the downstream trap; m_i is the number of smolts marked and released; and rm_i the number of marked recaptured. See text for more details about the experiments.

from the trap is closed to change in size. No spawner runs downstream after getting over the trap. There is neither mortality following marking nor natural mortality between the time of marking and recapture for marked or unmarked spawners. H3: there is no tag shedding. H4: the recapture (or resighting) probability is the same for all the fish whether or not marked.

Under the “equal catchability assumption” H1, one can consider the migration of the N_i spawners as independent Bernoulli experiments with probability of “success” θ_i . Accordingly, c_i is the observed result of a Binomial experiment in which each of the N_i spawners has a probability θ_i of being caught (T5.1).

Hypotheses H2–H3 enable one to consider that the number of spawners present on the spawning grounds above the trap is $N_i - x_i$ and that m_i are marked among them. In practice, the recapture sample is obtained by three methods: electrofishing on the spawning grounds, collection of dead fish after spawning, and trapping of spent fish at the downstream trap of the Cerisel facility. Each of the three recapture methods randomly samples within a fraction of the $N_i - x_i$ fish. Electrofishing samples the fish alive on the spawning grounds. Recovery of carcasses samples fish dead after spawning not previously caught alive. Downstream trapping samples the fish having survived to spawn and not previously caught alive. As a first approximation to this more complicated scheme, our modeling approach works as if the aggregated recapture sample r_i was randomly drawn from the entire population $N_i - x_i$. Under this premise and further assuming H4, one can consider the recapture sample r_i containing rm_i marked fish as a random sample from a population of total size $N_i - x_i$ in which m_i fish are marked. Because every fish recaptured is pelvic-fin clipped, recapture is a sampling without replacement. The recapture experi-

ment is thus modeled by a Hypergeometric distribution (T5.2).

Smolt run

For each year i between 1986 and 2001, c_i denotes the number of smolts caught in the downstream trapping facility during the migration time. Among the c_i smolts captured, a number $m_i \leq c_i$ have been tagged (mostly fin-clipped) and released upstream from the trapping facility used for capture (Fig. 1). Some of them will be recaptured at the same downstream trap. Note that the recapture is not conducted at a location downstream from the trap as it is often the case in mark–recapture experiments. We denote as rm_i the number of tagged and released smolts that are recaptured ($rm_i \leq m_i$). We make four classical assumptions analogous to H1–H4. H1': we assume equal catchability at the downstream trap (hypothesis analogous to H1). H2': the population is closed during the migration time. There is neither mortality induced by the capture–marking procedure nor natural mortality between the time of marking and recapture for either marked or unmarked smolts. H3': there is no tag shedding and all smolts marked and released will migrate out. H4': all m_i marked and released smolts have the same probability of being recaptured at the downstream trap. We suppose that the capture and marking does not affect the behavior of the smolts in a way that would change their vulnerability to the trap. Thus, the probability of recapture of previously marked and released smolts is the same as the probability of capture of unmarked smolts exposed to the trap for the first time.

Under H1', the first term of the likelihood L_i corresponds to a Binomial sampling distribution (T5.1'). Under H2'–H4', one can model the recapture experiment as a Binomial one with m_i trials and rm_i successes. H4' allows one to use in the recapture Binomial model (T5.2') the same probability of “success” θ_i as in the capture Binomial model. It is a crucial hypothesis for the estimation procedure. Intuitively, T5.2' will be used to estimate θ_i , and T5.1' will be used to estimate N_i conditionally on θ_i .

MCMC simulations and implementation with the software WinBUGS®

Once the step of describing the stochastic model is achieved, our model is easily implemented and fitted using an MCMC algorithm. We used the WinBUGS® software (Spiegelhalter et al. 2000, available at <http://www.mrc-bsu.cam.ac.uk/bugs>) that samples from the joint posterior distribution using the Metropolis-within-Gibbs sampling algorithm. The WinBUGS® code for all models described in this paper is available from the authors upon request. The basic idea behind the Gibbs sampling algorithm is to generate random draws for all unknown variables by cyclically sampling from the unidimensional full posterior conditional distributions of all variables. The latter are obtained by extracting from the joint posterior distribution the term involving only the variable in turn while treating all others terms as constant. Meyer and Millar (1999) reported on significant progress made in facilitating the routine implementation of Bayesian analysis thanks to the WinBUGS® software (see also the review by Gentleman 1997). It simplifies the implementation of the Gibbs sampling by alleviating

Table 5. Observation model and sampling equations for the capture–mark–recapture (CMR) experiments for spawners and smolts.

Spawners	Smolts
Observations	
$d_i = (c_i, x_i, m_i, r_i, rm_i)$	$d_i = (c_i, m_i, rm_i)$
Capture models	
(T5.1) $c_i \sim \text{Binomial}(N_i, \theta_i)$	(T5.1') $c_i \sim \text{Binomial}(N_i, \theta_i)$
Recapture models	
(T5.2) $rm_i \sim \text{Hypergeometric}(N_i - x_i, m_i, r_i)$	(T5.2') $rm_i \sim \text{Binomial}(m_i, \theta_i)$
Likelihood for year i $L(N_i, \theta_i d_i) = L_i$	
(T5.3) $L_i = C_{N_i}^{c_i} \theta_i^{c_i} (1 - \theta_i)^{N_i - c_i} \frac{C_{m_i}^{rm_i} C_{r_i - rm_i}^{r_i - rm_i}}{C_{N_i - x_i}^{r_i}}$	(T5.3') $L_i = C_{N_i}^{c_i} \theta_i^{c_i} (1 - \theta_i)^{N_i - c_i} C_{m_i}^{rm_i} \theta_i^{rm_i} (1 - \theta_i)^{m_i - rm_i}$

Note: The subscript i indicates the year of migration; CMR data d_i for spawners are given in Table 3; c_i denotes the number of spawners caught at the upstream trap; x_i denotes the number of spawners that died during manipulation or that are removed for experimental use or hatchery production; m_i is the number of spawners marked and released ($m_i = c_i - x_i$); r_i is the recapture sample; and rm_i is the number of fish marked among r_i . CMR data d_i for smolts are given in Table 4; c_i denotes the number of smolts caught at the downstream trap; m_i is the number of smolts marked and released; and r_i is the number of marked recaptured. For spawners, N_i denotes the number of upstream migrating fish and q_i denotes the upstream trap efficiency. For smolts, N_i denotes the number of downstream migrating fish and q_i denotes the downstream trap efficiency. Notation $X \sim Z$ denotes that the variable X has the probability distribution Z . See text for details about the hypotheses underlying the model.

the requirement to derive by hand the full conditional distributions. It only requires the declaration of the full Bayesian model and the specification of the data. Model declaration consists of the deterministic equations linking the variables, the prior density of unknown variables, and the sampling distributions relating data and variables. Initial values for all unknowns are also needed to start the sampling process.

In any MCMC method, a critical issue is whether the simulated chain has converged to its ergodic target distribution. Following the recommendations of Spiegelhalter et al. (2000), we ran three different chains starting at contrasting initial values spread over the space of the unknown variables. We first ran the three MCMC chains for 5000 initial iterations. To diminish the effect of the starting points, these first 5000 iterations were considered a burn-in period and were discarded. We generated another 20 000 random draws to derive posterior inferences. We checked convergence of the chains using the Gelman and Rubin (1992) diagnostics as modified by Brooks and Gelman (1998) proposed by WinBUGS®. For most of the models fitted, Gelman–Rubin diagnostics indicated convergence for the three chains after those 20 000 iterations. For the others, we reran 20 000 or 40 000 additional simulations to obtain convergence.

The output after convergence of the MCMC chains consists of a joint sample from the posterior distribution of all unknowns of the model. The integration needed to derive the marginal posterior distribution of any subset of unknowns is straightforward. Indeed, the MCMC sample of the subset of interest is to be considered a sample from its marginal posterior distribution. We summarize marginal posterior distributions of interest by simple statistics directly computed from the MCMC samples: mean, median, and 95% Bayesian posterior credibility interval.

Results

Spawning run

Our results show that the inferences obtained under the

ABM may be unreliable and highly sensitive to the choice of prior distributions. Posterior distributions of the θ_i s and N_i s derived from the ABM, with reference noninformative priors, vary highly between years in terms of both position and dispersion (Figs. 2a, 2b). The posterior estimates of the θ_i s are particularly variable between years (Fig. 2a). However, the high dispersion of θ_i s observed for years such as 1992 and 1993 does not necessarily have strong repercussion on the N_i s (Fig. 2b). Years with very few adults caught at the trap are indicative of a low population size, as far as the recapture data allow discarding of very low trapping probabilities. Some years have precise and symmetric posterior distributions of N_i , whereas other years may conversely have highly dispersed and skewed posteriors (Fig. 2b). For the latter, the upper bounds of the 95% Bayesian credibility intervals appear unrealistically high given the size of the Oir River and the available knowledge on the biology and ecology of Atlantic salmon. Inferences on the N_i s stem from the hypergeometric recapture experiments and thus depend on the recapture sample above all. Sparse mark–recapture data, i.e., low number of marked released or, more importantly, low number of recaptures of previously marked fish yield imprecise inferences (e.g., years 1987, 1990, and 1994). Posterior inferences under the ABM are rather robust to the choice of the prior on the θ_i s. In contrast, posterior inferences on the N_i s are highly sensitive to the prior choice. Changing the prior distribution of the N_i s can have a strong influence on posterior distributions, especially for data-poor years (see Appendix A).

Hierarchical modeling greatly improves posterior inferences (Fig. 2): the most precise inferences are obtained under the HBM. It is the hierarchical structure on the N_i s that is the major source of improvement. Compared with the ABM that assumes independence of both the N_i s and the θ_i s, the hybrid model assuming a hierarchical structure on the θ_i s alone does not really improve the inferences on the variable of ultimate interest, N_i , and the sensitivity to the choice of the prior on the N_i s is not alleviated (results not shown).

Fig. 2. Spawners case. Marginal posterior distributions of (a) the $\theta_{i,s}$ (upstream trap efficiency) and (b) the $N_{i,s}$ (number of upstream migrating fish) for years 1984 to 2000 obtained under four model structures. Line styles indicate the model investigated: solid line, hierarchical Bayesian model, with hyperpriors π_3 in Table A4 and π_2 in Table A5; broken line, hybrid model with hierarchical structure on the $\theta_{i,s}$ only with the hyperprior π_3 in Table A4, and the prior π_2 in Table A2 on the $N_{i,s}$ with $N_{\max} = 2000$; broken-dotted line, hybrid model with hierarchical structure on the $N_{i,s}$ only with the hyperprior π_2 in Table A5 and the prior π_1 in Table A1 on the $\theta_{i,s}$; dotted line, annual Bayesian model with the prior π_1 in Table A1 on the $\theta_{i,s}$ and the prior π_2 in Table A2 on the $N_{i,s}$ with $N_{\max} = 2000$. \blacktriangle and \blacktriangledown , 2.5 and 97.5 percentiles, respectively; \blacksquare , mean; +, median.

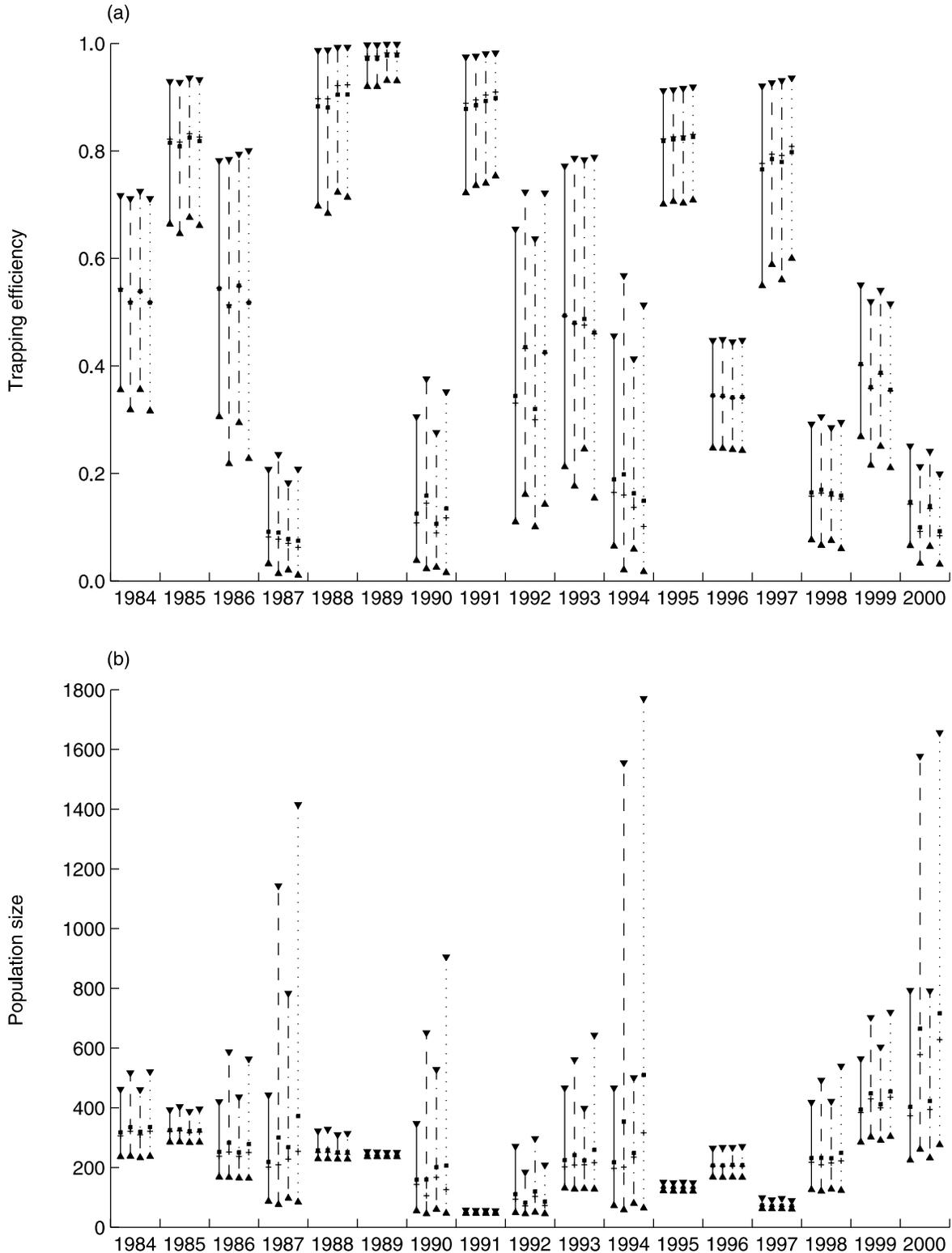
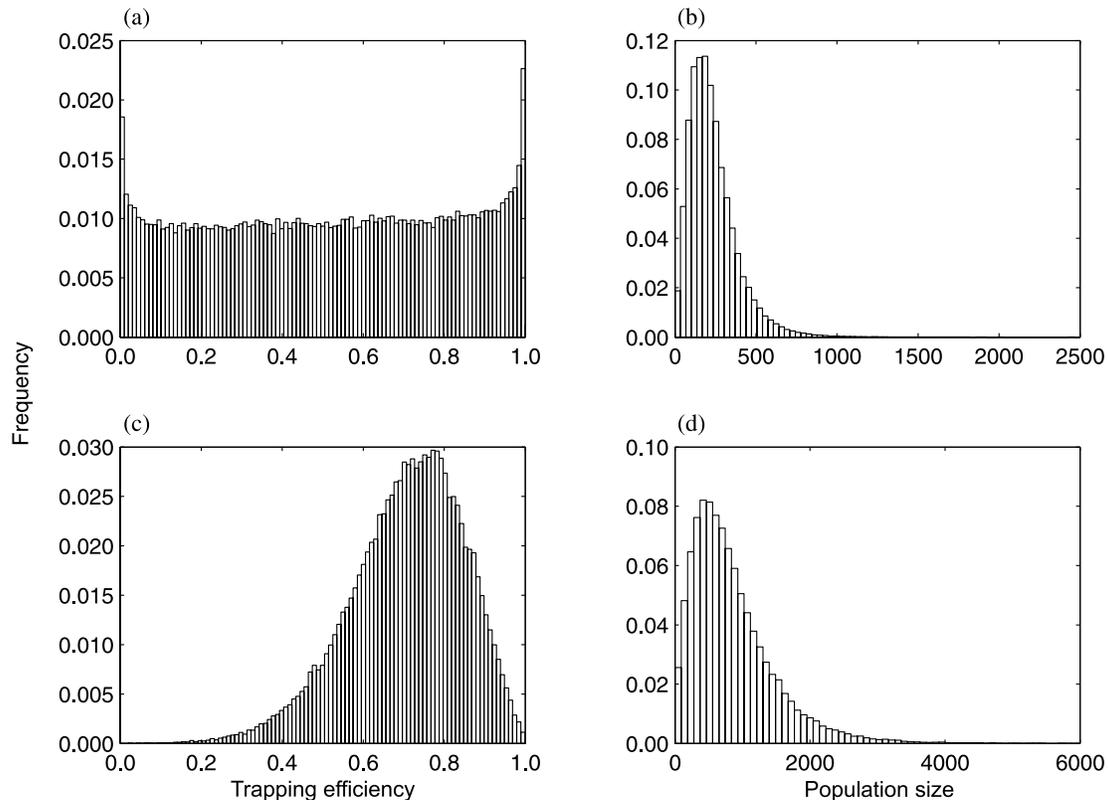


Fig. 3. Profiles of posterior predictive distributions $P(\tilde{\theta}|\text{data})$ and $P(\tilde{N}|\text{data})$ derived from the hierarchical Bayesian model in the spawners case (a and b, respectively) and in the smolts case (c and d, respectively). Hyperpriors used are π_3 in Table A4 and π_2 in Table A5. Profiles are frequency histograms issued from the Markov chain Monte Carlo samples.



Posterior mean values of the θ_i s are only slightly shrunk toward the overall mean at around 0.5 (Fig. 2a). In contrast, setting a hierarchical structure on the N_i s strongly reduces the skewness and uncertainty in the estimation of the N_i s (Fig. 2b). The tails of the posterior distributions of the N_i s are drastically shortened. These results appear much more sensible and realistic than those derived from the ABM. The reduction in uncertainty is particularly effective for years with poor CMR data. As expected, the gain in precision is insignificant for years with CMR data informative enough to enable precise estimation of the N_i s under the ABM. Posterior inferences under the HBM are quite robust to the choice of priors for the hyperparameters (see Appendix A).

The posterior predictive of the mean trapping efficiency $P(\tilde{\theta}|\text{data})$, under the HBM (Fig. 3a), appears rather uninformative. The transfer of information between years for the trapping efficiencies is weak. This observation is consistent with the insignificant effect of setting a hierarchical structure on the θ_i s. Unlike that of the trapping efficiency, the posterior predictive of the mean population size, $P(\tilde{N}|\text{data})$, is informative with a mean value around 230 and 95% of its density in the range [40,610] (Fig. 3b). The data of all years combined allow discarding a priori the possibility of very high spawner population size (i.e., greater than a thousand) in any additional year.

Smolt run

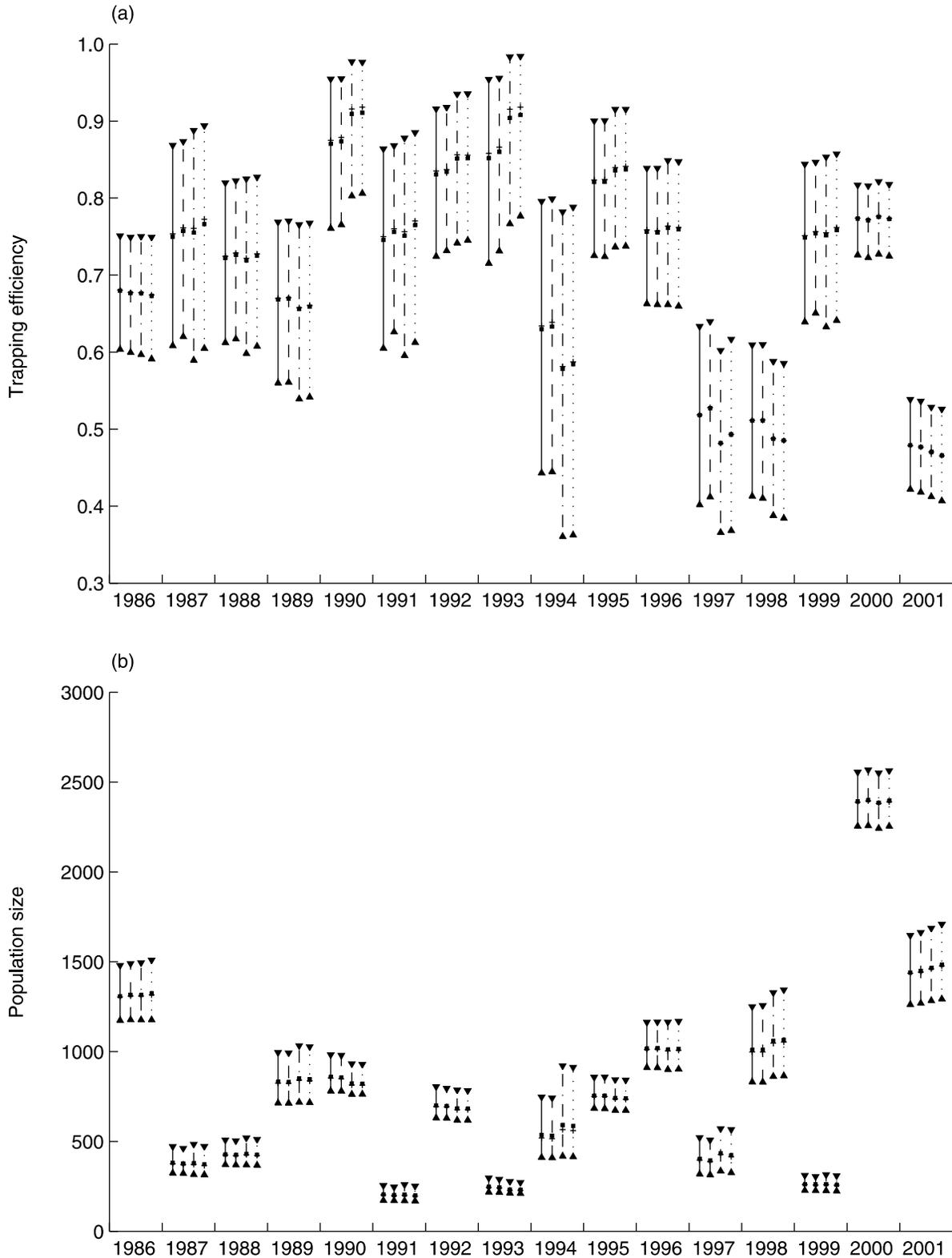
In contrast with the spawners case, the CMR data for the smolts appear to be much more informative and the ABM

framework enables precise estimations that are quite robust to the prior choice. Posterior distributions of the θ_i s are less variable between years than for spawners (Fig. 4a). The uncertainty around posterior estimates is generally much lower than in the spawners case. Posterior distributions of the N_i s are also less dispersed and less skewed than for the spawners (Fig. 4b). Accordingly, the posterior inferences derived under the ABM are rather insensitive to the choice of priors (see Appendix A).

Unlike in the spawners case, the HBM does not significantly improve posterior inferences compared with the ABM (Fig. 4). The hierarchical structure has more effect on posterior inferences for θ_i s than for N_i s. The hierarchical structure on θ_i has a slight shrinkage effect that reduces between-year variation (Fig. 4a): the most extreme estimates of θ_i are barely pulled up to the values of the other years (e.g., 1990, 1993, 1994, 1997, and 1998). Posterior distributions of N_i s and θ_i s under the HBM are robust to the choice of priors for the hyperparameters (see Appendix A).

In contrast with the spawners, both the posterior predictive $P(\tilde{\theta}|\text{data})$ and $P(\tilde{N}|\text{data})$ for the smolts are informative (Figs. 3c, 3d). On the basis of the data collected from 1986 to 2001, the probability of capture at the smolt trap is expected to average 0.7 with a 95% posteriori credibility interval [0.39,0.94]. The smolt run is expected to have 95% probability of being in the range [90,2500] with a mean of 850. However, the transfer of information between years remains of little use because yearly data are always informative enough.

Fig. 4. Smolts case. Marginal posterior distributions of (a) the $\theta_{i,s}$ (downstream trap efficiency) and (b) the $N_{i,s}$ (number of downstream migrating fish) for years 1986 to 2001 obtained under four model structures. Line styles indicate the model investigated: solid line, hierarchical Bayesian model, with hyperpriors π_3 in Table A4 and π_2 in Table A5; broken line, hybrid model with a hierarchical structure on the $\theta_{i,s}$ only with the hyperprior π_3 in Table A4, and the prior π_2 in Table A2 on the $N_{i,s}$ with $N_{\max} = 4000$; broken-dotted line, hybrid model with hierarchical structure on the $N_{i,s}$ only with the hyperprior π_2 in Table A5 and the prior π_1 in Table A1 on the $\theta_{i,s}$; dotted line, annual Bayesian model with the prior π_1 in Table A1 on the $\theta_{i,s}$ and the prior π_2 in Table A2 on the $N_{i,s}$ with $N_{\max} = 4000$. \blacktriangle and \blacktriangledown , 2.5 and 97.5 percentiles, respectively; \blacksquare , mean; $+$, median.



Discussion

Assessment of founding hypotheses

Like any model for a CMR experiment, our approach relies on the agreement with the classical hypotheses H1–H4 and H1'–H4' (see Materials and methods section). Warren and Dempson (1995), Arnason et al. (1996), and Schwarz and Taylor (1998) discussed how some of those hypotheses can be relaxed without strong consequences on the population size estimates. However, departure from those hypotheses may yield inaccurate estimates. Stratified models accounting for nonconstant catchability over time or heterogeneity in the capture probability between individuals may be more appropriate than pooled models (Pollock 1981, 1991; Plante et al. 1998). They have been applied to estimate Atlantic salmon migrations (Dempson and Stansbury 1991) or Pacific salmon migrations (Schwarz and Dempson 1994; Miyakoshi and Kudo 1999; Newcomb and Coon 2001). However, in many instances, as in our illustrative case study, the detailed information required to fit those models is not available and pooling is necessary. The robustness of pooled models to the departure from their underlying hypotheses is beyond the scope of this paper, which focuses on the contribution of hierarchical modeling. Yet, it is worth keeping in mind that some hypotheses may be questionable.

Hierarchical modeling can significantly improve inferences

Our paper is focused on the comparative merits of HBM versus ABM in the estimation of the quantities of interest and primarily the $N_{i,t}$. Formal statistical procedures, e.g., the deviance information criteria (DIC; Spiegelhalter et al. 2002), may provide some valuable indication for selecting the most appropriate model. However, we followed Spiegelhalter et al. (2002) in believing that such a formal selection approach could be of little relevance when other features of the models can be taken into account. We thus preferred to found the comparison between the models on two main criteria: the dispersion of posterior inferences of quantities of ultimate interest, i.e., the $N_{i,t}$, and the sensitivity to the choice of prior.

In the spawners case, the HBM clearly yields more reliable inferences compared with the ABM. However, the hybrid model assigning the $\theta_{i,t}$ a common hierarchical distribution while letting the $N_{i,t}$ be independent only slightly modifies the inferences when compared with the ABM. The yearly CMR data indicate that $\theta_{i,t}$ have little in common between years. This is probably because they are influenced by the water discharge during the migration season. In this case, the hierarchical hypothesis on $\theta_{i,t}$ has no effect because there is very little information transferred among years through the hyperparameters γ . The spread of the posterior distribution of γ remains wide. As a consequence, the marginal prior on $\theta_{i,t}$ is as diffuse as if $\theta_{i,t}$ were assigned independent uninformative priors, and subsequent inferences are not markedly different.

A noticeable positive effect is obtained by imposing a hierarchical structure on $N_{i,t}$. The spawners CMR data set typifies a survey where data are poor for some years. Inferences on the $N_{i,t}$ based on the ABM are of little reliance when the number of marks released or the size of the recapture sample

are too small. Consequently, posterior distributions of estimates are widely dispersed, and some 95% posterior credibility intervals may spread over ecologically unrealistic values. Another detrimental effect is high sensitivity to the prior choice because the data do not contain sufficient information to dominate the prior. For certain years, two prior distributions, both meant to be uninformative, yield markedly different posterior estimates of $N_{i,t}$ (see Appendix A). Several studies have already highlighted such sensitivity (Raftery 1988; Smith 1991; Garthwaite et al. 1995). Inferences on $N_{i,t}$ are also sensitive to minor changes in the data (Chao 1989) and consequently may be nonrobust to errors made when collecting data. Thanks to its ability to take advantage of the whole historical information available, the HBM limits these undesirable effects and significantly improves the inferences. The posteriors of the $N_{i,t}$ are less dispersed, more coherent, and more robust to changes in the prior assumptions. In the HBM, the posterior distribution of each year in turn is obtained by the combination of the likelihood with a modified prior that uses the data from all other years. This modified prior is informative about the $N_{i,t}$ and the information that it brings is significant compared with that of a single year with poor data.

In contrast with the spawners case, fitting the HBM to the smolts data does not significantly improve inferences on the $N_{i,t}$ compared with the ABM. The hierarchical structure on both $\theta_{i,t}$ and $N_{i,t}$ has no effect because the likelihood always dominates the prior. Indeed, unlike in the spawners case, the yearly CMR data for smolts are highly informative. The information brought by all years combined is negligible compared with that of any single year. Consequently, the use of a noninformative prior, as in the ABM, or of an informative prior based on historical information, as in the HBM, makes little difference.

The contrasted results obtained for the spawners and the smolts shows that the contribution of the hierarchical modeling is positive or null. In the best case, setting a hierarchical structure improves the inferences; in the worst case, it remains neutral.

General interest of the hierarchical approach to the analysis of ecological data

We believe that hierarchical Bayesian modeling has a great potential of application in the analysis of ecological data. The interesting properties revealed by our specific study could be extended to many cases considering a collection of units in time or space, yet partial commonality can be assumed across units. Whenever there are enough units to derive posterior inferences about the hyperparameters, we encourage the use of an exchangeable hierarchical model.

Hierarchical modeling is almost universally applicable even when some available knowledge would prevent the similarities and (or) differences between units from being modeled as draws from a common probability distribution. Indeed, any information available to distinguish different units can be encoded in covariates and accounted for in a conditional independent model, which assumes partial exchangeability (Gelman et al. 1995). In our case, it might have been possible to model the variations of $\theta_{i,t}$ with the water discharge, $Q_{i,t}$, by any function of the form $\theta_{i,t} = f(Q_{i,t}) + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ are random residuals. The partial exchange-

ability can be invoked by assuming that ε_i s are exchangeable with hyperparameters Ψ . Exchangeability is then modeled by the conditional independence in eq. 3:

$$(3) \quad \pi(\theta|Q_1, \dots, Q_I, \Psi) = \pi(\Psi) \prod_{i=1}^I \pi(\theta_i|Q_i, \Psi)$$

The analysis of data in a hierarchical framework is now simplified by the use of MCMC algorithms within modern software such as WinBUGS[®]. Sensitivity analysis of posterior inferences to the choice of priors is also facilitated. George and Robert (1992) already showed that CMR models lend themselves to Bayesian treatment using the Gibbs sampler. However, they faced the practical restrictions that the prior and the likelihood must lead to conditional posterior distributions, which are easily amenable to simulations. WinBUGS[®] allows for a wide choice of prior and sampling distributions, even if they lead to highly nonstandard conditional posteriors (Gentleman 1997; Meyer and Millar 1999). Formulations that would have been avoided in the past because of analytical and numerical difficulties can now be easily considered for practical applications. Prior and sampling distributions can be modified without having to respecify analytically the full conditional probability distributions and without getting bogged down by details about the densities used.

Because of the transfer of information between years, a hierarchical framework can accommodate missing data, provided they are few. In our example concerning smolts, let us imagine that recapture data is missing for a particular year i . Even in the absence of yearly data to update θ_i , the prior $\pi(\theta_i)$ will be updated by the data of all years except i , i.e., data _{i} ^{*}. This “updated” informative prior allows us to derive more precise posterior inferences on N_i than the initial prior $\pi(\theta_i)$.

To conclude, an exchangeable hierarchical model appears as a superior methodology to treat many ecological data sets. As we pointed out, a model that considers each unit independently may be considered a particular case of an exchangeable hierarchical model. When the data set contains poorly informative or even missing data, the independence assumption between the sampling units may lead to widely dispersed posterior inferences. To reduce uncertainty, one may be tempted to use informative priors. However, this generally consists of adapting the prior to the amount of information brought by the data, e.g., the less informative the data, the more informative the prior introduced. This may result in the observations playing a role in determining the prior distribution, and this practice must be strongly discouraged. We encourage others to consider a hierarchical model as the default choice. As noted by Gelman et al. (1995), even if the unknown variables to estimate vary highly among units, it is not irrelevant to estimate them in a hierarchical framework. For each unit, an updated prior is build by combining information from all experiments in a consistent and parsimonious framework. This informative prior is scaled to the between-units (e.g., years in our case study) variability and each unit’s contribution is judiciously weighed, i.e., depending on how informative its corresponding data are. The inferences about quantities of interest can be significantly

improved in terms of both precision and robustness to the choice of prior distributions on hyperparameters.

Acknowledgements

We gratefully acknowledge Frederic Marchand and Richard Delanoë for their invaluable work in collecting and computing the trapping data. We are especially indebted to Eric Parent for useful discussion about statistical modeling and for helpful comments on earlier versions of this paper. We also wish to thank Jean-Luc Baglinière and Jean-Marc Roussel for their thoughtful comments.

References

- Arnason, A.N., Kirby, C.W., Schwarz, C.J., and Irvine, J.R. 1996. Computer analysis of data from stratified mark–recovery experiments for estimation of salmon escapement and other populations. *Can. Tech. Rep. Fish. Aquat. Sci. No. 2106*.
- Baglinière, J.L., Maisse, G., and Nihouarn, A. 1993. Comparison of two methods of estimating Atlantic salmon (*Salmo salar*) wild smolt production. *In* Production of juvenile Atlantic salmon, *Salmo salar*, in natural waters. *Edited by* R.J. Gibson and R.E. Cutting. *Can. Spec. Publ. Fish. Aquat. Sci. No. 118*. pp. 189–201.
- Basu, S., and Ebrahimi, N. 2001. Bayesian capture–recapture methods for error detection and estimation of population size: heterogeneity and dependence. *Biometrika*, **88**(1): 269–279.
- Box, G.E.P., and Tiao, G.C. 1992. Bayesian inference in statistical analysis. *Col. Wiley Classics Library*, John Wiley and Sons, New York and London.
- Brooks, S.P., and Gelman, A. 1998. General methods for monitoring convergence of iterative simulations. *J. Comp. Graph. Stat.* **7**: 434–455.
- Chao, A. 1989. Estimating population size for sparse data in capture–recapture experiments. *Biometrics*, **45**: 427–438.
- Dempson, J.B., and Stansbury, D.E. 1991. Using partial counting fences and a two-sample stratified design for mark–recapture estimation of an Atlantic salmon smolt population. *N. Am. J. Fish. Manag.* **11**: 27–37.
- Garthwaite, P.H., Yu, K., and Hope, P.B. 1995. Bayesian analysis of multiple-recapture model. *Commun. Statist.-Theory Meth.* **24**(9): 2229–2247.
- Gazey, W.J., and Staley, M.J. 1986. Population estimation from mark–recapture experiments using a sequential Bayes algorithm. *Ecology*, **67**(4): 941–951.
- Gelman, A., and Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* **7**: 457–511.
- Gelman, A., Carlin, J.B., Stern, H., and Rubin, D.B. 1995. Bayesian data analysis. *Col. Texts in Statistical Science*, Chapman & Hall, London and New York.
- Gentleman, R. 1997. A review of BUGS: Bayesian inferences using Gibbs sampling. *Chance*, **10**(2): 48–51.
- George, E.I., and Robert, C.P. 1992. Capture–recapture estimation via Gibbs sampling. *Biometrika*, **79**(4): 677–683.
- Meyer, R., and Millar, R.B. 1999. BUGS in Bayesian stock assessments. *Can. J. Fish. Aquat. Sci.* **56**: 1078–1086.
- Miyakoshi, Y., and Kudo, S. 1999. Mark–recapture estimation of escapement of Masu salmon *Oncorhynchus masu* with a comparison to a fence. *N. Am. J. Fish. Manag.* **19**: 1108–1111.

- Newcomb, T.J., and Coon, T. 2001. Evaluation of three methods of estimating numbers of steelhead smolts emigrating from great lakes. *N. Am. J. Fish. Manag.* **21**: 548–560.
- Plante, N.L., Rivets, L.-P., and Tremblay, G. 1998. Stratified capture–recapture estimation of the size of a closed population. *Biometrics*, **54**: 47–60.
- Pollock, K.H. 1981. Capture–recapture models: a review of current methods, assumptions and experimental design. *Stud. Avian Biol.* **6**: 426–435.
- Pollock, K.H. 1991. Modeling capture, recapture and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. *J. Am. Stat. Assoc.* **86**(413): 225–238.
- Prévoist, E. 2000. Le stock de saumon atlantique du Scorff (Morbihan). Bilan synthétique d'évaluation à l'issue du premier contrat "Retour aux sources" (1994–1998). Programme "Restauration du Scorff pour les poissons migrateurs", Convention No. 6968682, Rapport final, mars 2000.
- Prévoist, E., Baglinière, J.L., Maise, G., and Nihouarn, A. 1996. Premiers éléments d'une relation stock/recrutement chez le Saumon atlantique (*Salmo salar*) en France. *Cybiurn*, **20**(Suppl. 3): 7–26.
- Raftery, A.E. 1988. Inference for the binomial N parameter: a hierarchical Bayes approach. *Biometrika*, **75**(2): 223–228.
- Schwarz, C.J., and Dempson, J.B. 1994. Mark–recapture estimation of a salmon smolt population. *Biometrics*, **50**: 98–108.
- Schwarz, C.J., and Taylor, C.G. 1998. Use of the stratified-Petersen estimator in fisheries management: estimating the number of pink salmon (*Oncorhynchus gorbuscha*) spawners in the Fraser River. *Can. J. Fish. Aquat. Sci.* **55**: 281–296.
- Seber, G.A.F. 1982. The estimation of animal abundance and related parameters. 2nd ed. Charles Griffin & Company Ltd., London and High Wycombe.
- Smith, P.J. 1988. Bayesian methods for multiple capture–recapture surveys. *Biometrics*, **44**: 1177–1189.
- Smith, P.J. 1991. Bayesian analysis for a multiple capture–recapture model. *Biometrika*, **78**(2): 399–407.
- Spiegelhalter, D., Thomas, A., and Best, N. 2000. WinBUGS. Version 1.3. User Manual. MRC and Imperial College of Science, Technology and Medicine (<http://www.mrc-bsu.cam.ac.uk/bugs>).
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Van der Linde, A. 2002. Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B*, **64**(Part 3): 1–34.
- Sprott, D.A. 1981. Maximum likelihood applied to a capture–recapture model. *Biometrics*, **37**: 371–375.
- Su, Z., Adikson, M.D., and Van Halen, B.W. 2001. A hierarchical Bayesian model for estimating historical salmon escapement and escapement timing. *Can. J. Fish. Aquat. Sci.* **58**: 1648–1662.
- Warren, W.G., and Dempson, J.B. 1995. Does temporal stratification improve the accuracy of mark–recapture estimates of smolts production? A case study based on the Cone River, Newfoundland. *N. Am. J. Fish. Manag.* **15**: 126–136.

Appendix A

Specification of the priors and sensitivity analysis

To investigate the robustness of our conclusions to the choice of priors, we compared posterior distributions obtained with several priors with various mean and dispersion features, either informative or not. We first give the distribution that we tested in the annual Bayesian model (ABM) for both the θ_i s and the N_i s (Tables A1 and A2). Then, we de-

Table A1. Prior distributions for the θ_i s used in the ABM for both spawners and smolts.

$\pi_1(\theta_i) \sim \text{Beta}(1/2, 1/2)$
$\pi_2(\theta_i) \sim \text{Beta}(1, 1)$
$\pi_3(\theta_i) \sim \text{Beta}(0.01, 1)$
$\pi_4(\theta_i) \sim \text{Beta}(1, 0.01)$

scribe the conditional prior distributions $\pi(\theta_i|\gamma)$ and $\pi(N_i|\delta)$ (Table A3) and the diffuse hyperpriors on γ and δ (Tables A4 and A5). Almost all of the priors that we used for the N_i s in the ABM and for hyperparameters in the hierarchical Bayesian model (HBM) are bounded. By restricting the priors to a finite domain of variation, we ensure that all priors are proper. We thereby avoid improper posterior distributions that may arise from the use of improper priors (Gelman et al. 1995). In the HBM, the ranges used for the hyperparameters were taken wide enough that the posterior mass was not accumulated close to the bounds and that increasing the range did not affect the posterior distributions. We thus ensured that we did not affect a priori null density to a domain where the likelihood was non-null.

ABM: prior on the θ_i s (Table A1)

Beta distributions were assigned to the θ_i s. The Beta distribution allows very different shapes and is the natural conjugate for the classical Binomial likelihood (Box and Tiao 1992). We tested four different sets for the hyperparameters (α, β) that reflect different prior opinion. Setting $\alpha = \beta$ in π_1 and π_2 yields symmetric distribution around 0.5. π_1 is the standard noninformative prior (Box and Tiao 1992), and π_2 corresponds to a Uniform on [0,1] (used by Raftery 1988). π_3 and π_4 give strong weight to values of θ_i close to 0 and close to 1, respectively. π_3 is closed to an improper prior $\propto 1/\theta_i$ used as a noninformative by Garthwaite et al. (1995) for their multicapture–recapture census model.

ABM: prior on the N_i s (Table A2)

We first tested different discrete little informative priors. They are adapted to situations where only vague prior knowledge about the N_i s is available and when the aim is to extract the useful information from the data. The bounded Uniform discrete prior π_1 is motivated by its practical convenience (used by Gazey and Staley 1986 or Chao 1989). The only information that it brings into the analysis is that a zero probability is assigned to values greater than the truncation point N_{\max} . The classical discrete prior π_2 is used as an uninformative prior by Raftery (1988), Smith (1991), George and Robert (1992), and Garthwaite et al. (1995). For both π_1 and π_2 , we assessed the sensitivity of the posterior inferences to the choice of the bound by testing different values of N_{\max} .

To avoid the subjective choice of the bounds, we also tried the Inverse Binomial (or Negative Binomial) as a robust alternative to the Poisson law proposed by Raftery (1988). As suggested by Smith (1991), we used a two-step prior density to simulate the Inverse Binomial. If the conditional distribution of N_i is considered to be Poisson with mean λ , and if λ has the natural conjugate distribution Gamma(a, b), then the unconditional distribution of N_i is Inverse Binomial with parameters (a, b) (see Raftery 1988 or Smith 1991). We first

Table A2. Prior distributions for the N_i s used in the ABM.

	Form	Bounds or parameters	
		Spawners	Smolts
$\pi_1(N_i)$	Uniform on $[0, N_{\max}]$	$N_{\max} = 1000, 2000, 3000$	$N_{\max} = 4000, 7000, 10\ 000$
$\pi_2(N_i)$	$1/N$ on $[1, N_{\max}]$	$N_{\max} = 1000, 2000, 3000$	$N_{\max} = 4000, 7000, 10\ 000$
$\pi_3(N_i)$	Poisson(λ)	$\lambda \sim$ Inverse Gamma(0.001, 0.001)	$\lambda \sim$ Inverse Gamma(0.001, 0.001)
$\pi_4(N_i)$	Inverse Binomial(a, b)	$(E_N, \sigma_N) = (100, 20)$	$(E_N, \sigma_N) = (1500, 300)$

Note: For both spawners and smolts, (E_N, σ_N) values are grounded on previous knowledge gathered on a nearby river, the Scorff River (Brittany) (see Baglinière et al. 1993 for the smolts and Prévost 2000 for the spawners).

Table A3. Beta and Inverse Binomial conditional prior distributions for θ_i and N_i (Gelman et al. 1995).

Beta distribution for θ_i

For all i in $1, \dots, I$

$$(TA3.1) \quad \pi(\theta_i | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1} (1 - \theta_i)^{\beta-1}, \quad \alpha > 0, \beta > 0, \theta_i \in [0, 1]$$

Mean and variance

$$(TA3.2) \quad \begin{cases} E_{\theta}(\alpha, \beta) = \frac{\alpha}{\alpha + \beta} \\ V_{\theta}(\alpha, \beta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \end{cases}$$

Inverse transformation under the constraint $V_{\theta} < E_{\theta}(1 - E_{\theta})$

$$(TA3.3) \quad \begin{cases} \alpha(E_{\theta}, V_{\theta}) = E_{\theta} \left(\frac{E_{\theta}(1 - E_{\theta})}{V_{\theta}} - 1 \right) \\ \beta(E_{\theta}, V_{\theta}) = (1 - E_{\theta}) \left(\frac{E_{\theta}(1 - E_{\theta})}{V_{\theta}} - 1 \right) \end{cases}$$

Inverse Binomial distribution for N_i

For all i in $1, \dots, I$

$$(TA3.4) \quad \pi(N_i | a, b) = C_{N_i+a-1}^{a-1} \left(\frac{b}{b+1} \right)^a \left(\frac{1}{b+1} \right)^{N_i}, \quad a > 0, b > 0, N_i = 0, 1, \dots$$

Mean and variance

$$(TA3.5) \quad \begin{cases} E_N(a, b) = \frac{a}{b} \\ V_N(a, b) = \frac{a(b+1)}{b^2} \end{cases}$$

Inverse transformation under the constraint $V_N > E_N$

$$(TA3.6) \quad \begin{cases} a(E_N, V_N) = \frac{E_N^2}{(V_N - E_N)} \\ b(E_N, V_N) = \frac{E_N}{(V_N - E_N)} \end{cases}$$

Table A4. Priors on the natural hyperparameters $\gamma = (\alpha, \beta)$ or transformed ones used in the HBM for both spawners and smolts.

	Parameters	Form	Bounds
π_1	Initial (α, β)	Uniform on $[\alpha_{\min}, \alpha_{\max}] \times [\beta_{\min}, \beta_{\max}]$	$[\alpha_{\min}, \alpha_{\max}] = [0, 100]$ $[\beta_{\min}, \beta_{\max}] = [0, 100]$
π_2	$u = \ln(\alpha/\beta)$ $v = \ln(\alpha + \beta)$	Uniform on $[u_{\min}, u_{\max}] \times [v_{\min}, v_{\max}]$	$[u_{\min}, u_{\max}] = [-5, 10]$ $[v_{\min}, v_{\max}] = [-5, 10]$
π_3	(E_{θ}, V_{θ})	$E_{\theta} \sim$ Beta(1/2, 1/2) $V_{\theta} E_{\theta} \propto 1/V_{\theta}$ on $[V_{\theta \min}, V_{\theta \max} E_{\theta}]$	$V_{\theta \min} = 1 \times 10^{-20}$ $V_{\theta \max} E_{\theta} = E_{\theta}(1 - E_{\theta})$

used the unbounded uninformative (improper) prior π_3 obtained when λ has an improper prior distribution proportional to $1/\lambda$. Practically, we approximated this distribution

by first drawing λ from an Inverse Gamma(0.001, 0.001) (see Meyer and Millar 1999 or Su et al. 2001). π_4 reflects subjective information about the N_i s. A priori informative

Fig. A1. Spawners case. (a) Sensitivity analysis of marginal posterior distributions of the $\theta_{i,s}$ to the choice of the prior on the $\theta_{i,s}$ in the annual Bayesian model (ABM). Line styles indicate the four priors on the $\theta_{i,s}$ investigated (Table A1): solid line, π_1 ; broken line, π_2 ; dotted line, π_3 ; broken-dotted line, π_4 . Prior used on the $N_{i,s}$ is π_2 in Table A2 with $N_{\max} = 2000$. (b) Sensitivity analysis of marginal posterior distributions of the $N_{i,s}$ to the choice of the prior on the $N_{i,s}$ in the ABM. Line styles indicate the four priors on the $N_{i,s}$ investigated (Table A2): solid line, π_1 with $N_{\max} = 2000$; broken line, π_2 with $N_{\max} = 2000$; dotted line, π_3 ; broken-dotted line, π_4 . Prior used on the $\theta_{i,s}$ is π_1 in Table A1. \blacktriangle and \blacktriangledown , 2.5 and 97.5 percentiles, respectively; \blacksquare , mean; +, median.

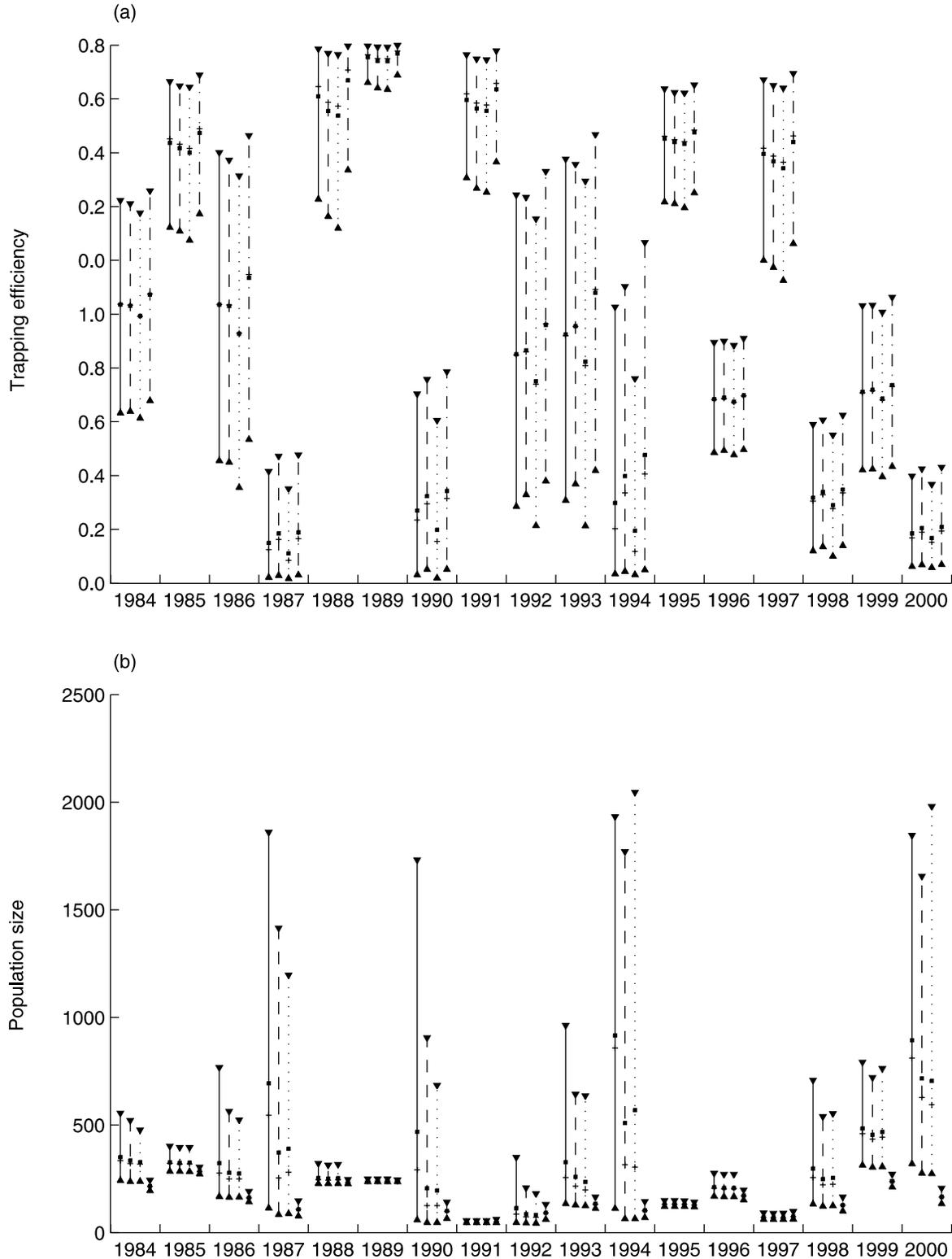


Fig. A2. Smolts case. (a) Sensitivity analysis of marginal posterior distributions of the θ_s s to the choice of the prior on the θ_s s in the annual Bayesian model (ABM). Line styles indicate the four priors on the θ_s s investigated (Table A1): solid line, π_1 ; broken line, π_2 ; dotted line, π_3 ; broken-dotted line, π_4 . Prior used on the N_i s is π_2 in Table A2 with $N_{\max} = 4000$. (b) Sensitivity analysis of marginal posterior distributions of the N_i s to the choice of the prior on the N_i s in the ABM. Line styles indicate the four priors on the N_i s investigated (Table A2): solid line, π_1 with $N_{\max} = 4000$; broken line, π_2 with $N_{\max} = 4000$; dotted line, π_3 ; broken-dotted line, π_4 . Prior used on the θ_s s is π_1 in Table A1. \blacktriangle and \blacktriangledown , 2.5 and 97.5 percentiles, respectively; \blacksquare , mean; $+$, median.

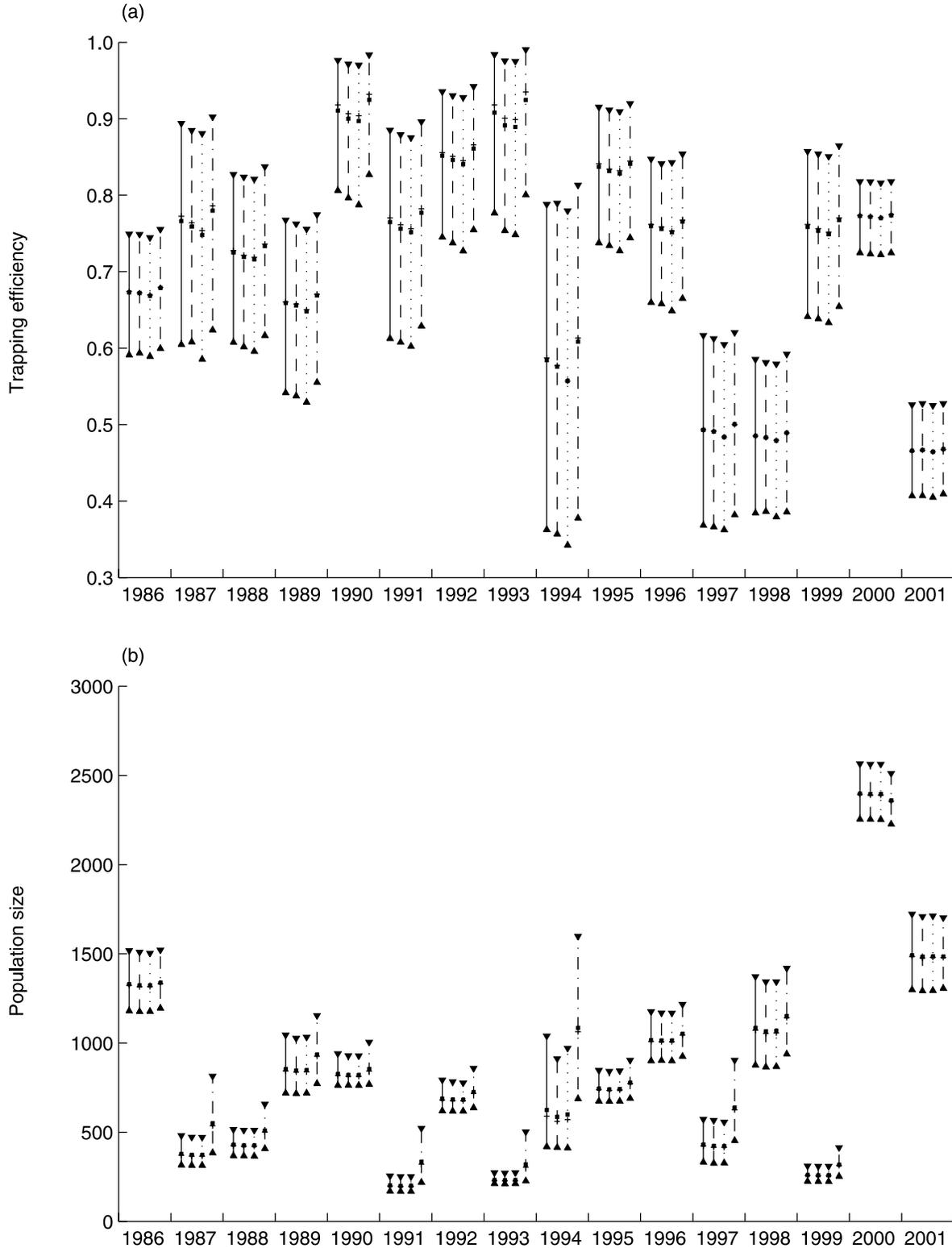


Table A5. Priors on the natural hyperparameters $\delta = (a,b)$ or transformed ones used in the HBM for both spawners and smolts.

	Parameters	Form	Bounds
π_1	Initial (a,b)	Uniform on $[a_{\min}, a_{\max}] \times [b_{\min}, b_{\max}]$	$[a_{\min}, a_{\max}] = [0,50]$; $[b_{\min}, b_{\max}] = [0,50]$
π_2	(E_N, V_N)	$E_N \sim \text{Uniform on }]0, E_{N \max}]$; $V_N E_N \propto 1/V_N$ on $]E_N, V_{N \max}]$	$E_{N \max} = 6000$; $V_{N \max} = 1.0 \times 10^6$

values are chosen for E_N and σ_N , the mean and standard deviation for the N_i s. To determine the corresponding (a,b) parameters of the Inverse Binomial, we use the inverse transformation between the mean and standard deviation and the natural parameters in TA3.5.

HBM: conditional priors $\pi(\theta_i|\gamma)$ and $\pi(N_i|\delta)$ and hyperpriors $\pi(\gamma)$ and $\pi(\delta)$

In the HBM, the natural choice for the conditional prior distributions $\pi(\theta_i|\gamma)$ and $\pi(N_i|\delta)$ are the Beta and the Inverse Binomial distributions, respectively, in Table A3. The parameters for which a prior distribution must be specified are the hyperparameters of the Beta and Inverse Binomial, denoted $\gamma = (\alpha, \beta)$ and $\delta = (a, b)$, respectively. These are higher level parameters and are generally assigned a diffuse hyperprior distribution to reflect our ignorance about them. As posterior inferences may be sensitive to the hyperpriors, we also compared results obtained with several hyperpriors $\pi(\gamma)$ and $\pi(\delta)$, which illustrate different interpretations of our little knowledge a priori (Tables A4 and A5).

Hyperprior $\pi(\gamma)$ (Table A4)

Different forms intended to reflect vague prior information about the capture probability have been proposed by Gelman et al. (1995), George and Robert (1992), and Smith (1991). The simplest that we tested is π_1 , a bounded Uniform distribution on the initial Beta parameter (α, β) . Instead of defining a prior directly on (α, β) , it could be relevant to set a prior on some appropriate one-to-one transformed parameters and then to go back to (α, β) via the inverse transformation. π_2 considers a bounded Uniform on the transformed parameters $(\log(\alpha/\beta), \log(\alpha + \beta))$ (Gelman et al. 1995). π_3 is a non-informative on (E_θ, V_θ) , respectively the mean and variance of the Beta distribution. The equations TA3.3 give the one-to-one relationship of (α, β) in terms of (E_θ, V_θ) . To ensure that the parameters (α, β) are >0 , V_θ must satisfy the constraint $V_\theta < E_\theta(1 - E_\theta)$. Practically, to simulate π_3 , we use a two-step conditional prior density. We first draw E_θ in a Beta($1/2, 1/2$). Then, to respect the previous constraint, we draw V_θ in a noninformative prior $\propto 1/V_\theta$ over the restricted range $[V_{\theta \min}, E_\theta(1 - E_\theta)]$ conditioned on the value E_θ previously drawn.

Hyperprior $\pi(\delta)$ (Table A5)

The prior distribution π_1 is a bounded Uniform on initial parameters (a, b) . π_2 corresponds to a noninformative prior

on (E_N, V_N) , respectively the mean and variance of the Inverse Binomial distribution. Equation TA5.6 gives the expression of (a, b) in terms of (E_N, V_N) . To simulate π_2 , we used a two-step conditional prior density as for $\pi_3(E_\theta, V_\theta)$. We assigned a bounded Uniform distribution over $]0, E_{N \max}]$ to E_N . To ensure that (a, b) are positive, we draw V_N in a prior $\propto 1/V_N$ over the range $]E_N, V_{N \max}]$.

Results of the sensitivity analysis

Spawning run

Posterior distributions of the θ_i s in the ABM are slightly sensitive to the choice of the prior $\pi(\theta_i)$ (Fig. A1a, Table A1). This sensitivity has only moderate repercussions on the posterior of the N_i s (results not shown). By contrast, changing prior distribution on the N_i s in the ABM (Table A2) has strong influence on posterior estimations (Fig. A1b). Two priors, both meant to be uninformative (π_1 and π_2 , Table A2), yield markedly different estimates. Moreover, both π_1 and π_2 (Table A2) lead to very skewed posteriors with tails that are sensitive to the choice of the truncation point N_{\max} (results not shown). Sensitivity to the prior is especially acute for data-poor years (i.e., 1987, 1990, 1994). For these years, an informative prior such as π_4 (Table A2) dominates the likelihood, thus leading to markedly different results than vague priors. Posterior inferences derived under the HBM are robust to the choice of priors on hyperparameters (Tables A4 and A5) (results not shown).

Smolt run

Unlike in the spawners case, the results obtained in the ABM are rather robust to the choice of prior. The sensitivity to the choice of the prior on the θ_i s is slight (Fig. A2a) and has hardly any repercussion on the posterior inferences on the N_i s (results not shown). The influence of the priors on the N_i s is much less important than for spawners. The three different diffuse priors π_1 , π_2 , and π_3 (Table A2) yield similar posterior distributions (Fig. A2b). We have not detected any significant sensitivity to the choice of the bounds N_{\max} for π_1 and π_2 in Table A2 (results not shown). The likelihood always dominates the prior, even when the informative prior is used (π_4 in Table A2). Posterior inferences are robust to the choice of priors on hyperparameters (Tables A4 and A5) (results not shown).