Estimation of annual mortality rates caused by early mortality syndromes (EMS) and their impact on salmonid stock–recruit relationships

Catherine G.J. Michielsens, Samu Mäntyniemi, and Pekka J. Vuorinen

Abstract: In this paper, we demonstrate how information from broodstocks can be combined with lab information on alevins to obtain annual stock-specific mortality estimates from early mortality syndromes (EMS) using a probabilistic approach, how a hierarchical model structure can be used to predict these mortality rates for related, partly sampled, or unsampled stocks, and why these estimates should be used to remove the effect of this mortality on stock–recruit estimates. The approach has been illustrated for Atlantic salmon (*Salmo salar*) stocks in the Baltic Sea affected by the M74 syndrome. Results indicate that data on the proportion of M74-affected females, commonly used to approximate M74 mortality, overestimate actual M74-related mortality because of a declining trend in mortality among offspring of these females. The stock-specific M74 mortality estimates are used to account for nonstationarity in the stock–recruitment relationship caused by this fluctuating mortality. Because hierarchical meta-analyses assume exchangeability, the effect of M74 mortality is removed before including these stocks within hierarchical stock–recruit analyses of Atlantic salmon stocks, which are commonly unaffected by M74 mortality. Failure to remove the effect of M74 mortality on the stock–recruit data results in underestimation of the stock's productivity and resilience to exploitation, especially in the case of stocks with steep stock–recruit curves.

Résumé : Notre travail démontre comment de l'information provenant du stock reproducteur peut être combinée par une méthode probabiliste avec des renseignements recueillis en laboratoire sur les alevins, afin d'obtenir des estimations de la mortalité due aux syndromes de mortalité précoce (EMS) et spécifique aux stocks; il montre aussi comment une structure de modèle hiérarchique peut servir à prédire ces taux de mortalité chez des stocks apparentés partiellement échantillonnés ou non échantillonnés et pourquoi ces estimations devraient être utilisées pour retirer les effets de cette mortalité dans les estimations de stock–recrutement. Nous illustrons cette méthodologie avec des stocks de saumons atlantiques (*Salmo salar*) de la Baltique affectés par le syndrome M74. Les résultats indiquent que les données sur la proportion de femelles affectées par M74, couramment utilisées pour obtenir une approximation de la mortalité due à M74, surestiment la mortalité réelle causée par M74 à cause d'une tendance décroissante de la mortalité chez les rejetons de ces femelles. Les estimations de la mortalité au M74 spécifique aux stocks servent à tenir compte de l'état non stationnaire de la relation stock–recrutement causée par cette mortalité fluctuante. Parce que les méta-analyses hiérarchiques présupposent l'échangeabilité, l'effet de la mortalité au M74 est retiré avant que les stocks ne soient inclus dans les analyses hiérarchiques de stock–recrutement des stocks de saumons atlantiques généralement non affectées par la mortalité au M74. Le fait de ne pas retirer les effets de la mortalité au M74 des données de stock–recrutement a pour effet de sous-estimer la productivité du stock et sa résilience à l'exploitation, particulièrement chez les stocks qui ont des courbes de stock–recrutement à forte pente.

[Traduit par la Rédaction]

Introduction

Stock–recruit relationships are among the most important model specifications in fisheries stock assessment, but it is often difficult to obtain the needed data to estimate the appropriate relationship. For salmonids, it is generally accepted that there is a density-dependent relationship between the spawning stock size and subsequent smolt recruitment (Jonsson et al. 1998) and that the most important effects of density dependency occur once the alevins have consumed their yolk sacs, emerged from the spawning gravel as fry, and started feeding (Elliott 1993). Traditionally, stock–

Received 1 November 2005. Accepted 22 April 2006. Published on the NRC Research Press Web site at http://cjfas.nrc.ca on 23 August 2006.
J18972

C.G.J. Michielsens[1] and P.J. Vuorinen. Finnish Game and Fisheries Research Institute, P.O. Box 2, FIN-00791 Helsinki, Finland.
S. Mäntyniemi.[2] University of Helsinki, Rolf Nevanlinna Institute, Department of Mathematics and Statistics, P.O. Box 4, FIN-00014 Helsinki, Finland.

[1]Corresponding author (catherine.michielsens@rktl.fi).
[2]Present address: University of Helsinki, Department of Biological and Environmental Sciences, P.O. Box 56, FI-00014 Helsinki, Finland.

 doi:10.1139/F06-095

recruit relationships require the spawning stock to be expressed in terms of eggs (Beverton and Holt 1956) because of the non-linear relationship between the number of spawners and the number of eggs, which is due to a more diverse age structure at high spawner numbers (Needle 2002). The number of eggs spawned can be estimated from information on the fecundity at age and the proportion of each age. Because stock–recruit analyses are very sensitive to errors in the measurement of the spawning stock size (Walters and Ludwig 1981), it is important to take into account any annual changes in the parameters that determine the spawning stock size. In addition, it is important to remove effects that render the stock–recruit relationship non-stationary (Hilborn and Walters 1992). Among such effects are early life history mortality syndromes, characterised by diet-related deficiencies of thiamine in broodstock and eggs and causing high mortality among yolk-sac fry and alevins (Bengtsson et al. 1999). These syndromes are an international problem, affecting various species of salmon and trout, e.g., the early mortality syndrome (EMS) in the Great Lakes, the M74 syndrome in the Baltic Sea area, and the Cayuga syndrome in the New York Finger Lakes (Fitzsimons et al. 1999). Because the mortality resulting from these syndromes occurs before density dependency takes place, i.e., before the fry emerge from the gravel, it is possible to remove the effect of this annual mortality from the estimates of the number of eggs to be used as an indicator of stock size within the stock–recruit relationship. Because the early mortality caused by these syndromes is dependent on the year, species, and location, in practise it is difficult to obtain time series estimates suitable to adjust the spawning stock size of stock–recruit data for individual stocks. Usually information on the mortality caused by these syndromes comes from observations on adult females, collected as broodstock, or on their offspring, either in hatchery rearing facilities or through lab experiments. It is therefore most common to have information on the early mortality of hatchery-reared salmonid stocks. When monitoring the yearly early life history mortality caused by these syndromes on wild stocks, budgetary concerns would only allow monitoring a limited number of stocks.

This is the first paper to estimate the effect of female-dependent early life mortality caused by EMS syndromes on stock–recruit analyses. Within this paper we demonstrate how different sources of information from broodstocks and alevins can be used to obtain annual estimates of early life mortality, how an hierarchical model approach can be used to extrapolate results from monitored salmonids stocks to unsampled stocks, and why the estimated early life history mortality rates should be taken into account in subsequent stock–recruit analyses. The approach has been illustrated for Atlantic salmon (*Salmo salar*) stock in the Baltic Sea area affected by the M74 syndrome.

Materials and methods

In this section, we first describe the available data. We then continue by describing the models, i.e., a general model for predicting the number of eggs surviving EMS and an estimation model, specific to the Baltic salmon case study, using the M74-monitoring data to provide estimates for parameters needed in the prediction model. To simplify notation, the prediction model has been developed for a single stock without specifying the age structure. The estimation model provides parameter estimates simultaneously for multiple stocks.

Data

The M74 syndrome was first detected in 1974 in Swedish Atlantic salmon hatcheries, where it caused increased mortality among alevins. The cause of the syndrome is still unknown but its occurrence has been linked to low thiamine contents within broodstock and eggs (Börjeson and Norrgren 1997; Romakkaniemi et al. 2003). The occurrence of M74 symptoms among offspring can be monitored by stripping the eggs from the female spawners and observing their development. An overview of the observed proportion of females with M74-affected offspring for different salmon stocks is provided (Table 1). Data on the occurrence of M74 are most commonly obtained when collecting broodstock for the hatcheries. As a result, most data on M74 relate to stocks that are hatchery-reared or supplemented with hatchery-reared salmon. The occurrence of the M74 syndrome has been monitored for seven hatchery-reared stocks and for five wild salmon stocks. Their location within the Baltic Sea area is shown (Fig. 1).

To relate the proportion of females with M74-affected offspring with the actual mortality among offspring, detailed information has been obtained for two wild salmon stocks, i.e., the Tornionjoki and Simojoki rivers (Table 2). By incubating in the lab batches of eggs from females collected from the wild (Vuorinen and Keinänen 1999), additional data on the total yolk-sac-fry mortality (YSFM) among offspring of individual females and on the number of females with 100% mortality among offspring have been obtained.

Prediction model

The prediction of the number of eggs, set to survive EMS, starts from the assumption that the number of females (N) contributing to the spawning stock would be known. We first assume that each female has the same chance (ρ) of its offspring being affected by EMS. Then, each female with affected offspring is assumed to have the same chance of losing all offspring (θ) to EMS. Thus, given that the chance parameters are known, the population of females can be divided into three categories, depending on the degree of EMS-related offspring mortality, using a multinomial distribution:

$$(1)\qquad (N_1, N_2, N_3) | N, \theta, \rho, \sim \text{Multinomial}(N, (1-\rho), \rho(1-\theta), \rho\theta)$$

where N_1 is the number of females with healthy offspring, and N_2 and N_3 represent the number of females with EMS-related mortality among part or all of their offspring, respectively.

If the variation in fecundity (f) among female spawners can be assumed to follow a normal distribution defined by a mean fecundity across females (μ_f) and an associated variance (σ_f^2)

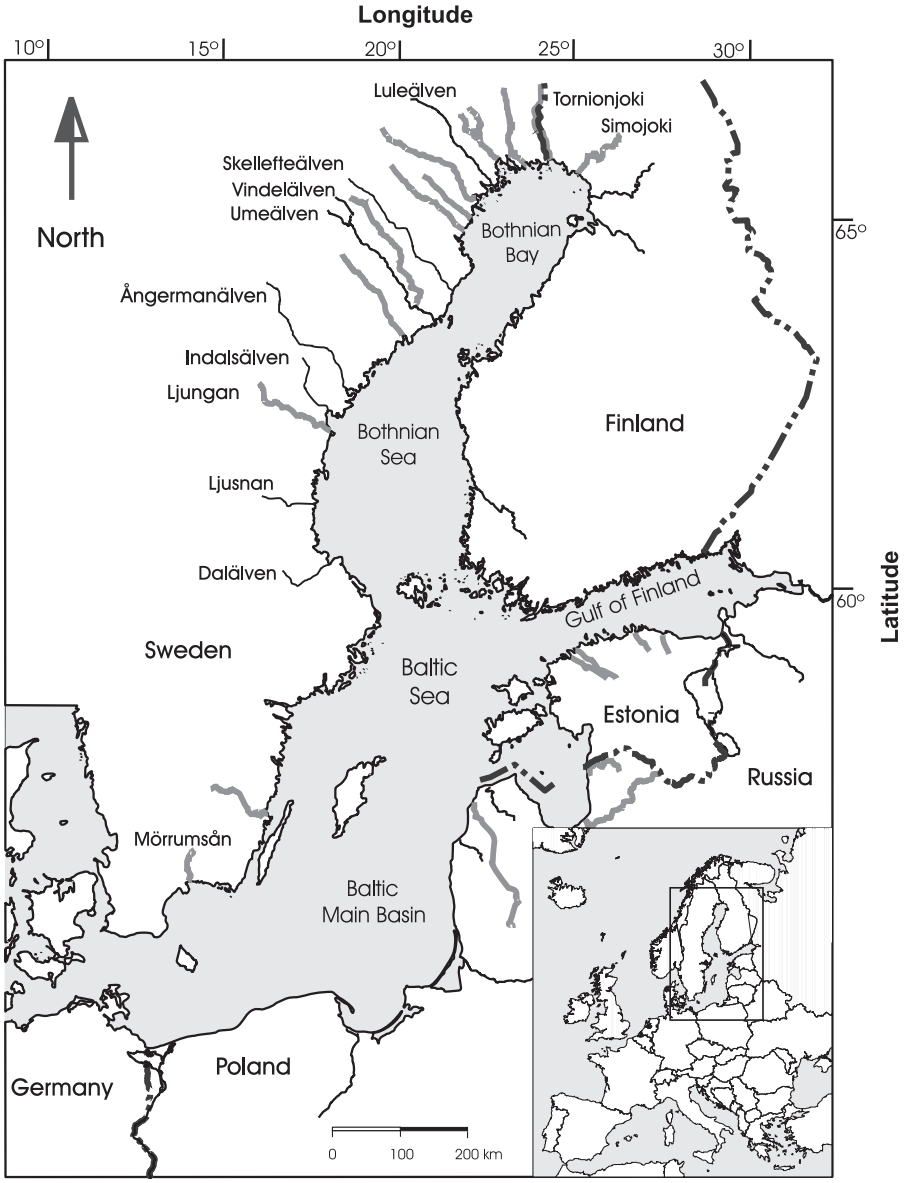
$$(2)\qquad f | \mu_f, \sigma_f^2 \sim \text{Normal}(\mu_f, \sigma_f^2)$$

Table 1. Summary of M74 data for nine different Atlantic salmon (*Salmo salar*) stocks (1985–2004), in terms of the number of females of stock j sampled in year i and with offspring affected by the M74 syndrome ($Y_{i,j}$) in comparison with the total number of females of stock j sampled in year i ($M_{i,j}$).

	Luleälven	Skellelteälven	Ume/Vindelälven	Angermanälven	Indalsälven	Ljungan	Ljusnan	Dalälven	Mörrumsån
1985	NA	NA	14/35	NA	9/219	NA	0/78	19/69	23/50
1986	NA	NA	16/82	NA	18/251	NA	0/49	4/49	24/50
1987	NA	NA	16/64	NA	20/245	NA	0/84	8/88	32/50
1988	NA	NA	12/64	NA	15/202	NA	0/75	16/79	23/50
1989	NA	NA	6/38	NA	6/192	NA	0/78	7/65	29/50
1990	NA	NA	18/59	NA	15/198	NA	0/86	4/45	39/55
1991	NA	NA	32/71	NA	14/196	NA	14/88	16/78	35/55
1992	161/279	16/40	55/71	78/157	85/190	14/22	29/89	50/63	33/60
1993	232/352	44/89	60/68	98/128	149/206	5/5	89/119	69/81	54/60
1994	269/435	54/78	146/164	52/79	148/208	6/12	105/163	70/126	4/5
1995	209/418	38/77	148/215	58/126	97/237	15/27	79/142	22/40	17/27
1996	202/392	54/70	68/87	36/57	107/167	6/22	92/128	102/178	10/18
1997	156/409	8/50	26/71	38/183	39/178	5/17	28/130	360/159	5/22
1998	22/389	2/48	6/37	3/81	2/155	2/20	7/82	14/83	NA
1999	108/316	22/53	27/51	30/108	25/126	5/20	19/46	27/82	NA
2000	67/320	7/57	27/60	29/136	27/125	1/10	29/114	36/131	NA
2001	96/322	9/51	24/62	31/122	7/100	0/10	47/102	27/82	NA
2002	119/300	8/42	20/53	56/122	25/123	6/11	23/60	56/150	NA
2003	12/270	4/60	8/53	15/120	5/128	0/2	17/100	22/164	NA
2004	10/270	0/59	2/56	4/114	0/125	NA	0/47	5/112	NA

Note: NA, not available.

Fig. 1. Map of the Baltic Sea, indicating the location of the Atlantic salmon (*Salmo salar*) stocks included in the hierarchical analysis. Rivers containing wild salmon stocks have been indicated by thick shaded lines.



the number of eggs or offspring in each category (O_c) also has a normal distribution given the number of females in each category (N_c) and the mean and variance of the fecundity per female,

$$(3) \qquad O_c | N_c, \mu_f, \sigma_f^2 \sim \text{Normal}(\mu_f N_c, \sigma_f^2 N_c)$$

Next, the number of eggs set to die at the alevin stage as a result of EMS (D_c) needs to be calculated for each category. Clearly, no offspring will die as a result of EMS in category 1, making $D_1 = 0$. In contrast, eggs from females belonging to category 3 will all die, thus $D_3 = O_3$. However, the chance φ that an egg from a female belonging to category 2 would die at the alevin stage as a result of EMS is assumed to vary between individuals according to a beta distribution with mean μ_φ and variation parameter η_φ, whereby parameter η_φ increases as the variance decreases, and vice versa. The number of dead eggs from each female of category 2 can then be described by a beta–binomial distribution

$$(4) \qquad D_2 | N_2, \mu_\varphi, \eta_\varphi, f \sim \text{Beta-Bin}(fN_2, \alpha, \beta)$$

$$(5) \qquad \alpha = \mu_\varphi \left(\frac{(\eta_\varphi + 1) f N_c - (\eta_\varphi + f)}{f - 1} \right)$$

$$(6) \qquad \beta = (1 - \mu_\varphi) \left(\frac{(\eta_\varphi + 1) f N_c - (\eta_\varphi + f)}{f - 1} \right)$$

The total number of dead eggs from all females then becomes modeled as a mixture of beta–binomial (eq. 4) and normal (eq. 2) distributions (details on the derivation of these equations can be found in Appendix A). It can be seen from these equations that to predict the number of eggs dying as a result of EMS, information about the number of females and

Table 2. Summary of M74 data for Atlantic salmon (*Salmo salar*) stocks of the Simojoki and Tornionjoki rivers (1985–2004) indicating the percentage of sampled females with offspring that display M74 symptoms (%), the total average yolk-sac-fry mortality among offspring of sampled females (%), and the percentage of sampled females with 100% mortality among offspring (%).

	Proportion of females with offspring affected by M74 (%)		Total average yolk-sac-fry mortality among offspring (%)		Proportion of females without surviving offspring (%)	
	Simojoki	Tornionjoki	Simojoki	Tornionjoki	Simojoiki	Tornionjoki
1985	0	NA	6	NA	NA	NA
1986	0	NA	2	NA	NA	NA
1987	0	NA	6	5	NA	NA
1988	0	NA	3	6	NA	NA
1989	0	NA	14	1	12	NA
1990	0	NA	4	29	NA	NA
1991	53	NA	52	70	47	NA
1992	74	NA	75	76	74	NA
1993	53	89	55	84	53	55
1994	92	76	76	66	58	49
1995	86	NA	67	NA	50	NA
1996	91	NA	71	NA	50	NA
1997	31	25	19	26	6	19
1998	59	61	54	62	38	56
1999	44	34	38	34	22	24
2000	41	41	41	35	26	21
2001	47	69	33	61	25	46
2002	7	3	2	4	0	0
2003	7	0	4	2	0	0
2004	3	NA	5	NA	3	NA

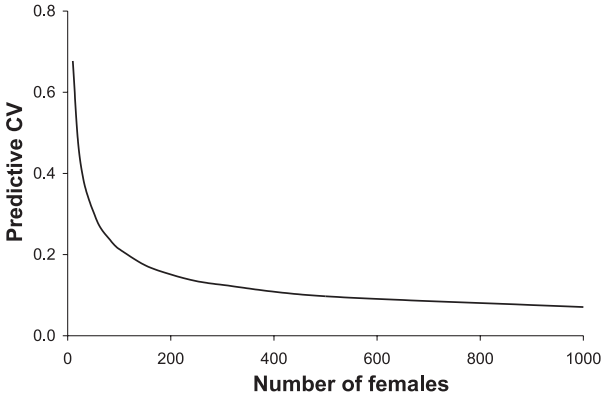
Note: NA, not available.

their fecundity is needed together with the parameters describing the occurrence and severity of the EMS-related mortality among spawner offspring. The expected value of the total number of eggs that die as a result of EMS can calculated as

$$(7)\qquad E(D_1 + D_2 + D_3) = E(D_1) + E(D_2) + E(D_3)$$
$$= \rho\mu_f N((1-\theta)\mu_\varphi + \theta)$$

The variance, however, is more difficult to compute. We used a Monte Carlo simulation to obtain the coefficient of variation (CV) of the total number of eggs set to die at the alevin stage due to EMS as a function of the number of females while keeping the other parameters fixed (Fig. 2). This illustrates that early life history mortality syndromes affecting the survival of eggs can make the estimated stock size very uncertain and the future of the population difficult to predict, especially when the population has declined to a very small number of spawners that each produce a large number of eggs.

Fig. 2. Coefficient of variation (CV) of the predictive distribution of the number of eggs that will die because of the early mortality syndrome (EMS) as a function of the number of female spawners when using the predictive model described in the text with fixed values for the EMS parameters ($\rho = 0.2$, $\theta = 0.8$, $\eta_\varphi = 20$, $\mu_\varphi = 0.3$) and fecundity values ($\mu_f = 3000$, $\sigma_f = 300$).

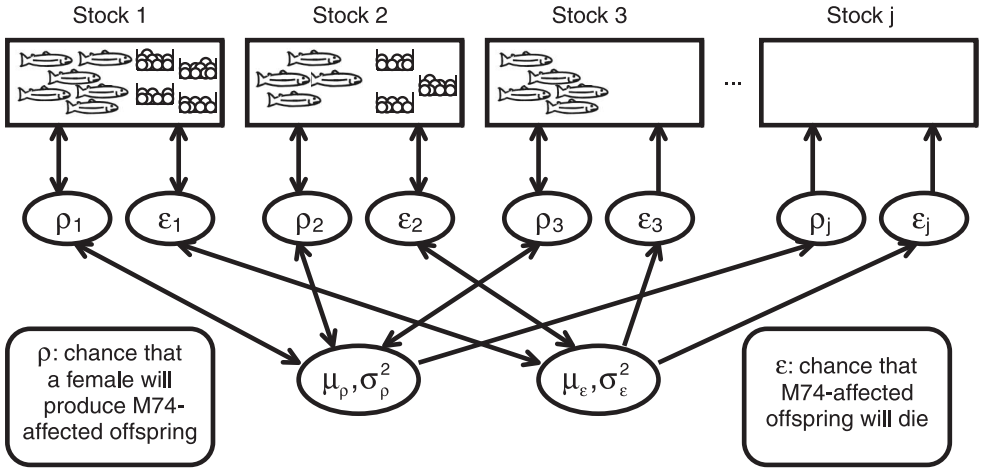


Estimation model

To provide information on the EMS parameters needed in the prediction model, an estimation model, specific to the Baltic salmon case study, has been developed and fitted to M74-monitoring data. Information about the number of females and their fecundity should be obtained by using other models and data sets. Following the idea of the Bayesian approach to statistical inference, we endeavor to specify the model structure based on our knowledge before obtaining the M74 data used to fit the model. We therefore base our assumptions about the syndrome on the early part of the time series (from the year 1974 when the syndrome was first found until the year 1984), after which a more detailed monitoring program was launched. The model is then fitted to time series of observations from 1985 to 2003.

A particular feature of the model proposed in this paper is the use of hierarchical structures (Gelman et al. 1995). Hierarchical structures can be used to combine data from several

Fig. 3. Graphical presentation showing the flow of information within the hierarchical model. Information from broodstocks is used to estimate the chance that a female will produce offspring with M74 syndromes (ρ), whereas lab information is used to estimate the chance that M74-affected offspring will die (ε). Because of the transfer-of-information, the hierarchical model can predict these chances for partially sampled or unsampled stocks.



stocks allowing the estimation of model parameters both at the stock-specific level and at a higher, cross-stock level. A hierarchical model structure can be applied for a particular parameter through the use of a mean across stocks (μ) and a parameter describing the parameter variation across stocks (η). Using this model structure, it is possible to predict parameters of a stock for which no data are observed, based on the estimates across stocks and the similarities between individual stocks. The hierarchical model proposed in this paper allows obtaining stock-specific estimations for the chance that a female is affected by M74 and the chance that the offspring of an affected female would die (Fig. 3).

We define the model structure by specifying conditional probability density functions (pdfs) for all model parameters and observable quantities given their immediate parents in the directed acyclic graph (DAG; Fig. 4) (Pearl 2000). An overview of all the model parameters and their symbols is provided (Table 3). Each node in the graph is conditionally independent, given its parents, of all the other nodes that are not its descendants, and the joint density of all the variables in the graph is the product of the densities of each node conditional on its parents. Parameters without parents are given unconditional prior pdfs (Table 4).

Within the estimation model, the survival of each egg from a female is assumed to be conditionally independent of the survival of other eggs from the same female, given that the survival chance of the egg is known. Then, given that the number of sampled eggs ($N_{i,j,k}$) from individual female k and the chance of survival ($\gamma_{i,j,k}$) were known, our beliefs about the number of eggs ($X_{i,j,k}$) that survive can be described by the binomial distribution

$$X_{i,j,k} \mid \gamma_{i,j,k}, N_{i,j,k} \sim \mathrm{Bin}(N_{i,j,k}, \gamma_{i,j,k}) \quad (8)$$

where i denotes the year and j denotes the stock. The chance of survival is assumed to depend on the normal YSFM and the additional YSFM caused by M74. If the offspring of female k show no signs of M74, then the chance of survival depends only on the normal YSFM. Mathematically this is implemented as

$$\gamma_{i,j,k} = \begin{cases} \lambda_{i,j,k} & , \text{ if } l_{i,j,k} = 0 \\ \lambda_{i,j,k}\,\delta_{i,j,k} & , \text{ if } l_{i,j,k} = 1 \end{cases} \quad (9)$$

where $\lambda_{i,j,k}$ denotes the chance to survive normal YSFM, which is assumed to be exchangeable between all individuals in every year and in every stock and modeled by

$$\lambda_{i,j,k} \mid \mu_\lambda, \eta_\lambda \sim \mathrm{Beta}(\mu_\lambda \eta_\lambda, (1-\mu_\lambda)\eta_\lambda) \quad (10)$$

where μ_λ is the mean YSFM of all individuals and η_λ describes the variation between individuals. The parameter $\delta_{i,j,k}$ denotes the chance that the eggs survive M74 and $l_{i,j,k}$ denotes the occurrence of M74 symptoms among the offspring of a female

$$l_{i,j,k} = \begin{cases} 1, \text{ if the offspring of a female} \\ \quad \text{show M74 symptoms} \\ 0, \text{ otherwise} \end{cases} \quad (11)$$

All females of stock j in year i are assumed to have equal chance $\rho_{i,j}$ of producing offspring with M74 symptoms. This implies that the indicator variable $l_{i,j,k}$ follows the Bernoulli distribution

$$l_{i,j,k} \mid \rho_{i,j} \sim \mathrm{Bern}(\rho_{i,j}) \quad (12)$$

Further, all annual stock-specific chances $\rho_{i,j}$ are assumed to be exchangeable, i.e., there exists no prior information indicating that certain stocks are more likely to be affected by M74 than others during certain years. This means that we can assume the parameters $\rho_{i,j}$ to be a conditionally independent sample from a distribution with unknown parameters. A beta distribution is chosen to represent the distribution

$$\rho_{i,j} \mid \mu_{\rho,i}, \eta_\rho \sim \mathrm{Beta}(\mu_{\rho,i}\eta_\rho, (1-\mu_{\rho,i})\eta_\rho) \quad (13)$$

where $\mu_{\rho,i}$ is the mean in year i, and η_ρ describes the amount of variation between stocks, assumed to be the same every year.

Early monitoring data indicate that the M74 syndrome can have two levels of severity: either all offspring die or part of the offspring survive. To take this phenomenon into account,

Fig. 4. Directed acyclic graph (DAG) of the hierarchical model estimating annual early life history mortality among Atlantic salmon (*Salmo salar*) stocks in the Baltic Sea using a combination of data on the proportion of female spawners with offspring exhibiting M74 syndromes and data on the chance of survival of affected offspring. Oval nodes represent stochastic variables, square nodes represent constants, arrows represent stochastic dependencies, and double arrows indicate a logical function. A list of the symbols and their descriptions can be found in Table 3.

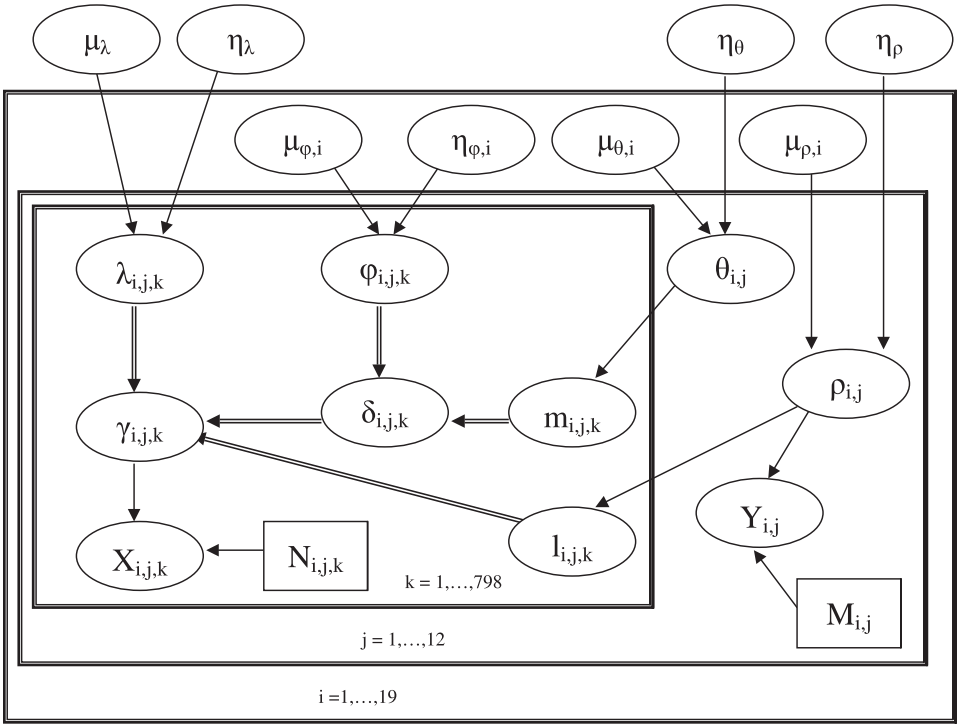


Table 3. List of symbols used within the estimation model.

Indices	
i	Year
j	Stock
k	Female
Data	
$X_{i,j,k}$	Number of surviving eggs for a sample of eggs taken in year i from female k of stock j
$N_{i,j,k}$	Total number of eggs sampled from female k of stock j in year i
$Y_{i,j}$	Number of females of stock j sampled in year i and with offspring affected by the M74 syndrome
$M_{i,j}$	Total number of females of stock j sampled in year i
Model parameters	
$l_{i,j,k}$	Indicates whether offspring from female k of stock j show symptoms of M74 in year i
$m_{i,j,k}$	Indicates whether the probability of surviving M74 mortality is equal to 0 for offspring of female k of stock j in year i
$\gamma_{i,j,k}$	Chance of an egg from female k of stock j surviving yolk-sac-fry mortality (YSFM) in year i (M74-related YSFM and "normal" YSFM)
$\lambda_{i,j,k}$	Chance of an egg from female k of stock j surviving "normal" yolk-sac-fry mortality in year i
$\delta_{i,j,k}$	Chance of an egg from female k of stock j surviving M74 mortality in year i
$\theta_{i,j}$	Chance that all offspring die, given that the offspring show M74 symptoms
$\varphi_{i,j,k}$	Chance of an egg from female k of stock j dying as a result of M74 in year i when M74 mortality of the offspring is not 100%
$\varepsilon_{i,j}$	Chance of offspring from females of stock j dying when affected by M74 in year i
$\rho_{i,j}$	Chance that offspring from females of stock j are affected by M74 in year i

we use variable $m_{i,j,k}$ to indicate the extent at which the offspring will be affected by M74. We define

$$\delta_{i,j,k} = \begin{cases} \varphi_{i,j,k}, & \text{if } m_{i,j,k} = 0 \\ 0, & \text{if } m_{i,j,k} = 1 \end{cases} \quad (14)$$

where $\varphi_{i,j,k}$ is the chance of the offspring surviving if M74 mortality is not 100%. We assume that chance parameters $\varphi_{i,j,k}$ are exchangeable within each year. This leads to the model

$$\varphi_{i,j,k} | \mu_{\varphi,i}\eta_{\varphi,i} \sim \text{Beta}(\mu_{\varphi,i}\eta_{\varphi,i}, (1-\mu_{\varphi,i})\eta_{\varphi,i}) \quad (15)$$

Table 4. Prior probability density functions (pdfs) for model parameters and corresponding 95% probability interval (PI).

Parameter	Density function	95% PI
μ_{ρ_i}	Beta(2,2)I(0.01,0.99)	0.09–0.9
η_ρ	Unif(2,1000)	27–975
μ_{θ_i}	Beta(2,2)I(0.05,0.95)	0.11–0.89
η_θ	Unif(10,1000)	12–97.5
μ_{φ_i}	Beta(2,2)I(0.01,0.99)	0.09–0.9
η_{φ_i}	Unif(2,1000)	27–975
μ_λ	Beta(2,2)I(0.01,0.99)	0.09–0.9
η_λ	Unif(2,1000)	27–975

Note: For the prior pdfs of the means, it is assumed that the values near 0 and 1 are less likely than the values in between.

where $\mu_{\varphi,i}$ represents the mean across stocks and $\eta_{\varphi,i}$ describes the respective variation across stocks. The chance $\theta_{i,j}$ that a female with offspring exhibiting M74 symptoms will lose all offspring as a result of M74 is assumed to be equal for all females of stock j in year i and exchangeable between stocks in each year. Thus

$$m_{i,j,k} \mid \theta_{i,j} \sim \text{Bern}(\theta_{i,j}) \quad (16)$$

$$\theta_{i,j} \mid \mu_{\theta,i}, \eta_\theta \sim \text{Beta}(\mu_{\theta,i}\eta_\theta, (1-\mu_{\theta,i})\eta_\theta) \quad (17)$$

where $\mu_{\theta,i}$ is the annual mean and η_θ describes the variation of $\theta_{i,j}$ between stocks in every year.

In most rivers, the fate of offspring from sampled females is not recorded. Instead, data from these rivers are comprised solely of the total number of females sampled ($M_{i,j}$) and the number of females with offspring showing M74 symptoms ($Y_{i,j}$). Because

$$Y_{i,j} = \sum_{k=1}^{M_{i,j}} l_{i,j,k} \quad (18)$$

and indicators $l_{i,j,k}$ are Bernoulli distributed given $\rho_{i,j}$, the number of females with offspring showing M74 symptoms has a conditional binomial distribution

$$Y_{i,j} \mid M_{i,j}, \rho_{i,j} \sim \text{Bin}(M_{i,j}, \rho_{i,j}) \quad (19)$$

The constructed model allows us to use both the detailed data on M74 mortality among salmon offspring available for a few stocks and the more common, aggregated data on the number of females with offspring affected by M74-related mortality. This model construction allows us to transfer information about the chance to survive M74 ($\delta_{i,j,k}$) across stocks while still accounting for differences in severity of M74-related mortality between stocks. It is important to note that for each stock and year, data values can be given to $Y_{i,j}$ or to the pairs ($X_{i,j,k}$, $l_{i,j,k}$) but not to both at the same time.

Computation and model checking

Posterior distributions of the model parameters are not available in closed-form expressions and are therefore approximated by Markov chain Monte Carlo (MCMC) simulation (Gilks et al. 1995). The MCMC simulation is implemented using the WinBUGS 1.4 software package (Spiegelhalter et al. 2003). The convergence of the simulated Markov chains is checked using the CODA software package (Best et al. 1995). All the modeling results described in this paper have undergone tests to remove the "burn-in" (Brooks and Gelman 1998), eliminate autocorrelation, and assess convergence, and it is assumed that the reported distributions are representative of the underlying stationary distributions.

The goodness of model fit is examined by calculating mixed predictive p-values as proposed by Marshall and Spiegelhalter (2003). If the model fits well to the observed data set, the set of mixed predictive p-values should resemble a sample from a uniform distribution. This is verified by a Q–Q plot of the p values against the corresponding quantiles of the uniform distribution (Fig. 5).

Results

When estimating the mortality among offspring of females affected by M74, the resulting posterior pdfs indicate a decreasing trend over the years (Fig. 6). Between 1991 and 1995, on average around 80% of the offspring died because of M74 when the female was affected by M74, whereas between 1999 and 2003, this percentage had dropped to around 50%. Because information on the chance of offspring survival is available only for the Simojoki and Tornionjoki stocks, most stocks are unsampled, and the corresponding estimates rely on the estimated mean and variance for the these two stocks (Fig. 6*c*).

Combining the chance that the offspring of a female would be affected by the M74 syndrome with the chance that the offspring would die as a consequence results in the chance that an egg from that female is set to die as a result of M74 when reaching the alevin stage (Fig. 7). These estimates of the M74 mortality (%) have been compared with the data traditionally assumed to approximate M74 mortality, i.e., the percentage of females with offspring affected by M74 and the total average YSFM among offspring. In general, the percentage of females with offspring affected by M74 overestimates the M74 mortality because part of the offspring will die as a result of normal YSFM, unrelated to M74. In addition, not all offspring necessarily die when affected by M74. Because of the decreasing trend in mortality among offspring of females affected by M74 (Fig. 6), the data on proportion of females affected by M74 especially overestimate M74 mortality in recent years. For some of the stocks (e.g., Figs. 7*f* and 7*h*), the proportion of female spawners with M74-affected offspring is relatively high in 2002, whereas the estimated M74 mortality indicates a further decrease compared with previous years. Data on the total average YSFM are much better at tracking the general trend but overestimate the actual M74 mortality because these data do not distinguish between normal YSFM and YSFM caused by the M74 syndrome, i.e., the actual cause of mortality cannot be determined for each individual. In addition, these data are hard to come by given their costs.

The hierarchical model proposed in this paper allows the chance of M74 mortality among alevins of different stocks to be estimated. Because of differences in the amount of information available for each stock, the uncertainty in the M74 mortality estimates differ between stocks. The value-

Fig. 5. Q–Q plots of the mixed predictive *p*-values for (*a*) the yolk-sac-fry mortality among offspring from individual females and (*b*) the percentage of females among the broodstock with M74 symptoms. The closer the points lie towards the 45° reference line, the better is the fit of the model to the data.

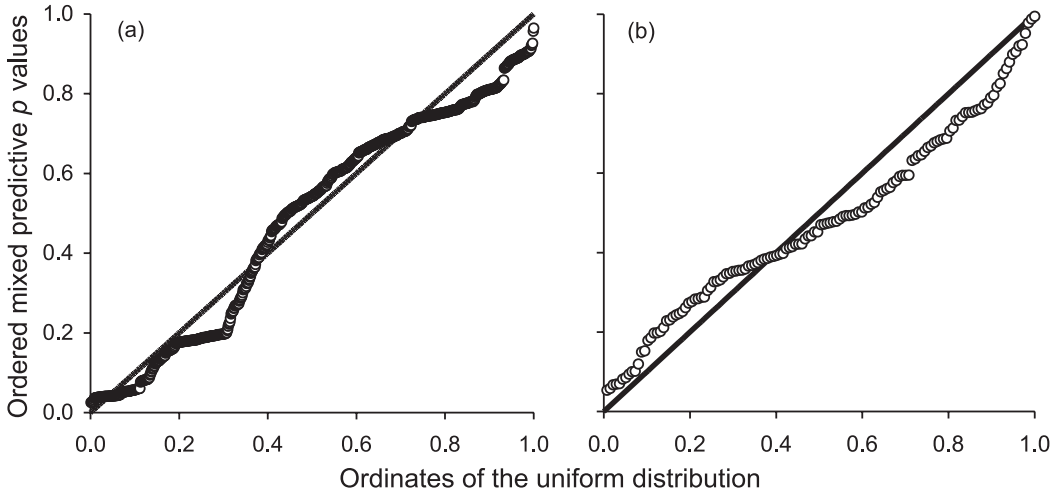
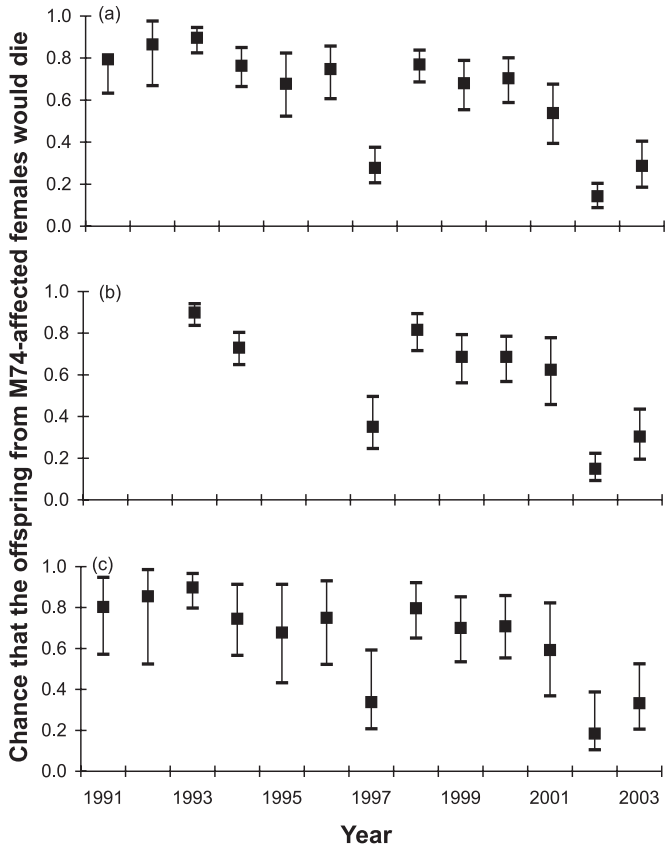


Fig. 6. Posterior probability density functions (median and 95% probability interval) for $\varepsilon_{i,j}$, i.e., the chance that M74-affected offspring would die (1991–2003): (*a*) Simojoki stock, (*b*) Tornionjoki stock, and (*c*) an unsampled Atlantic salmon (*Salmo salar*) stock within the Baltic Sea area.



of-information is illustrated for the Simojoki stock (Fig. 8). If information on the proportion of females with offspring affected by M74 is available, the posterior pdf for the M74 mortality rate becomes substantially more informative than the prior pdf. The M74 mortality estimates become even more certain if information on both the proportion of females with offspring affected by M74 and subsequent offspring mortality are available. The same figure also illustrates the transfer-of-information when using hierarchical model structures. Even when no stock-specific information is available, the hierarchical structure still allows a considerable update of the prior pdf (Fig. 8). The more stock-specific information becomes available, the less important the hierarchical structure becomes for that particular stock.

The results from the estimation model can be used within the prediction model to predict the number of eggs set to survive M74 mortality. The effect of M74 mortality on the estimated stock–recruit relationship has been illustrated for the salmon stocks in the Tornionjoki and Simojoki rivers for which stock–recruit estimates, hereafter called stock–recruit data, have been obtained by combining smolt abundance estimates (International Council for the Exploration of the Sea (ICES) 2005) with survival rate estimates from smolt to spawner stage (Michielsens et al. 2006) and fecundity estimates (ICES 2005) (Fig. 9). The resulting stock–recruit data with or without taking into account the M74 mortality have been included within an hierarchical meta-analysis of Atlantic salmon stock–recruit data. Because of the low probability of the Ricker stock–recruit function given the available Atlantic salmon stock–recruit data (Michielsens and McAllister 2004), only the Beverton–Holt stock–recruit function has been fitted to the data. The steepness parameter (z), i.e., the proportion of the long-term unfished recruitment obtained when the stock abundance is reduced to 20% of the virgin level (Mace and Doonan 1988) (Table 5), can be estimated using the hierarchical model. The steepness parameter, by definition restricted between 0.2 and 1, is greater for stocks that are more productive and resilient to exploitation. Results indicate that for the Tornionjoki stock the posterior steepness is much lower and more uncertain (mean z = 0.52, CV = 0.3) when ignoring M74 mortality in the stock–recruit data than when M74 mortality is accounted for (mean z = 0.69, CV = 0.18). For the Simojoki stock, the difference be-

Fig. 7. Posterior probability density functions (median (solid squares) and 95% probability interval (error bars)) for the M74 mortality (%) among offspring from different Atlantic salmon (*Salmo salar*) stocks (1985–2004): (*a*) Simojoki, (*b*) Tornionjoki, (*c*) Luleälven, (*d*) Skellefteälven, (*e*) Ume/Vindelälven, (*f*) Ångermanälven, (*g*) Indalsälven, (*h*) Ljungan, (*i*) Ljusnan, (*j*) Dalälven, (*k*) Mörrumsån, and (*l*) an unsampled stock within the Baltic Sea area. This figure also shows the data assumed to approximate M74 mortality, i.e., the proportion of females with offspring affected by M74 (open diamonds) and the total average yolk-sac-fry mortality among offspring (open circles).

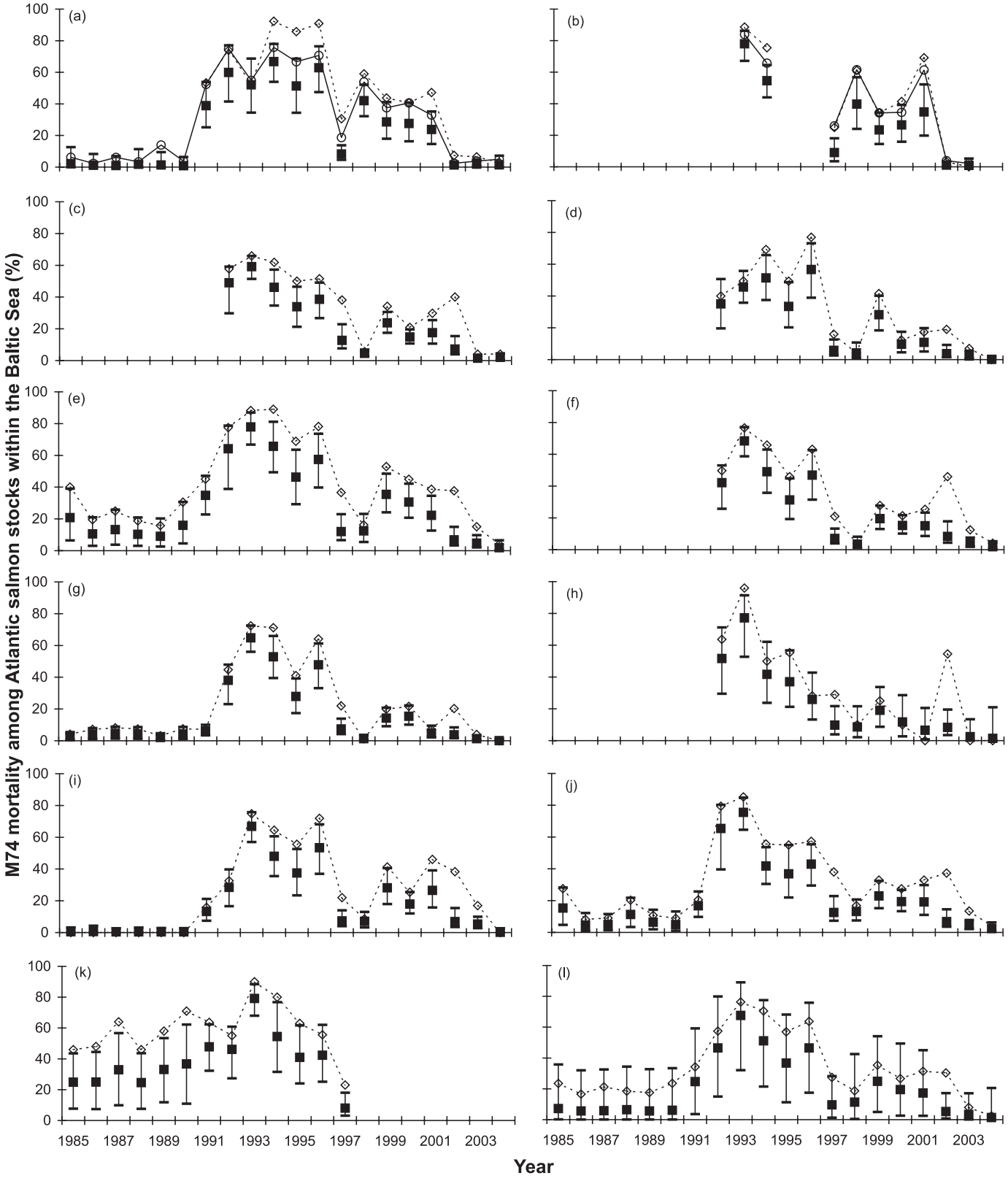
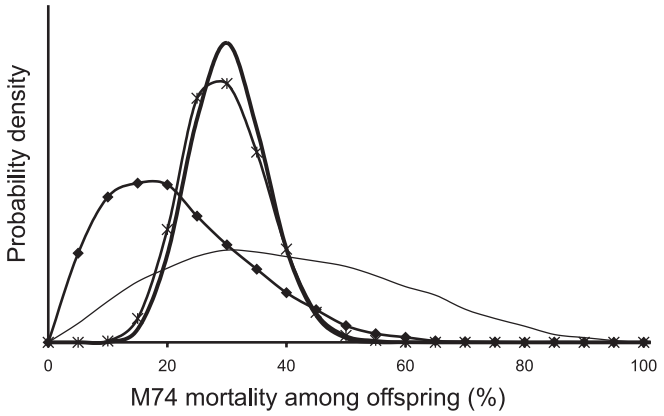


Table 5. Estimated mean and coefficient of variation (CV) for the prior and posterior probability density functions (pdfs) of the steepness for the Beverton–Holt stock–recruit function of Atlantic salmon (*Salmo salar*) stocks in the Tornionjoki and Simojoki rivers.

	Tornionjoki		Simojoki	
	Mean	CV	Mean	CV
Prior pdfs from hierarchical analysis of Atlantic salmon stock–recruit data	0.7	0.23	0.7	0.23
Posterior pdfs based on stock–recruit data with unaccounted M74 mortality	0.52	0.3	0.57	0.32
Posterior pdfs based on stock–recruit data with M74 mortality accounted for	0.69	0.18	0.59	0.28

Note: The prior pdf is obtained from a hierarchical analysis of Atlantic salmon stock–recruit data (Michielsens and McAllister 2004), whereas the posterior pdfs are obtained by including stock–recruit data, with or without accounting for M74 mortality, from these two stocks within this hierarchical analysis.

Fig. 8. Illustrative figure indicating the value-of-information for M74 data from females and (or) offspring and the transfer-of-information when using hierarchical model structures to estimate the percentage of M74 mortality among Simojoki offspring in the year 2000. Thin solid line, prior probability density function (pdf); solid diamonds, posterior pdf based on a hierarchical model structure with data from other stocks; asterisks, posterior pdf using a hierarchical model structure with M74 data of females; thick solid line, posterior pdf using a hierarchical model structure with M74 data of females and offspring.

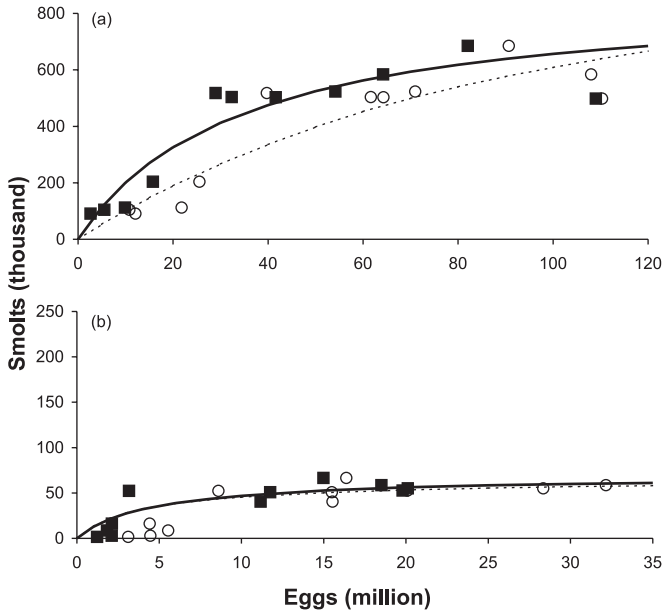


tween the results without (mean z = 0.57, CV = 0.32) or with (mean z = 0.59, CV = 0.28) accounting for M74 mortality is limited.

Discussion

Stock–recruit relationships are very important within fisheries stock assessment, even though they are very uncertain, and a large amount of effort is spent investigating this relationship (Cushing 1988). Current hierarchical methodologies for the analysis of stock–recruit data rely on the use of multiple data sets from related populations to decrease the uncertainty (Liermann and Hilborn 1997; Myers 2001; Michielsens and McAllister 2004). One major requirement when using hierarchical model structures is that the stock-specific parameters within one analysis need to be assumed exchangeable (Gelman et al. 1995). This means that there would be no information other than the stock–recruit data indicating differences in stock–recruit relationships. This renders any data series from salmonids stocks plagued by early life history mortality syndromes inappropriate for use within hierarchical analyses, even when the associated mortality is

Fig. 9. Median values of the probability density functions for the stock–recruit data of the (*a*) Tornionjoki and (*b*) Simojoki Atlantic salmon (*Salmo salar*) stocks with (■) and without (○) taking account of the early life history mortality caused by the M74 syndrome. The graphs also indicate the median values of the posterior probability distributions for the Beverton–Holt stock–recruit function obtained by including the stock–recruit data within a hierarchical stock–recruit analysis of Atlantic salmon stocks (Michielsens and McAllister 2004). To facilitate comparison between the two stocks, the *Y* axis of the graph for the Simojoki stock has been rescaled to represent similar egg-to-smolt ratios as for the Tornionjoki stock.



constant over the years. In addition, the extent to which salmonid populations are affected by early life history mortality syndromes often fluctuates over time, making the data series non-stationary and inappropriate when estimating the stock–recruit relationship. Because early life history mortality for salmonids can be assumed to take effect before any density dependency occurs (Elliott 1993), estimates for early life history mortality syndromes can be taken into account in stock–recruit analyses as a factor affecting stock size.

In case M74 mortality is ignored when including Baltic salmon stock–recruit data within the hierarchical meta-analysis of Atlantic salmon stocks, the resulting steepness estimates for the two Baltic salmon stocks are much lower

than the estimates obtained by Michielsens and McAllister (2004) for the other Atlantic salmon stocks, suggesting that Atlantic salmon stocks in the Baltic Sea would have a much lower productivity and resilience to exploitation than other Atlantic salmon stocks. When accounting for M74 mortality, the steepness estimate for the Tornionjoki stock is similar to the steepness estimates of other Atlantic salmon stocks. These latter results also correspond to the biological knowledge of the productivity and resilience of this stock to exploitation. In the late 1980s, when exploitation rates were still very high (Michielsens et al. 2006), several Atlantic salmon stocks in the Baltic Sea area were close to extinction and some even completely disappeared (Romakkaniemi et al. 2003), whereas the Tornionjoki stock still produced around 60 000 wild salmon smolts annually. Once exploitation rates had decreased, wild smolt production in the Tornionjoki river increased rapidly to a current production of around one-half million wild salmon smolts, about one-third of the total wild smolt production in the Baltic Sea (ICES 2005). For the Simojoki stock, the steepness estimate remains low, even when M74 mortality is accounted for within the stock–recruit analysis. The lower steepness of this stock can be explained by the large amount of hatchery-reared salmon released within the Simojoki river to supplement the wild salmon population and aid stock recovery, thus reducing the productivity of the stock (Kallio-Nyberg and Koljonen 1997; Jutila et al. 2003). These results indicate that accounting for the mortality caused by early mortality syndromes is especially important for productive stocks.

Because stock–recruit data themselves are already very uncertain, it is important to use all available information when estimating the mortality associated with early life history mortality syndromes. Data on early life history mortality syndromes, however, are difficult to collect. In the most simplified cases, it can be estimated from spawner-related observations. As a result, the occurrence of M74 has traditionally been expressed in terms of the proportion of females with offspring affected by M74 (ICES 2005). Although initially all offspring from a female would succumb to the M74 syndrome (Börjeson and Norrgren 1997), partial M74 mortality, i.e., M74 mortality affecting only part of the offspring from one female, seems more common in recent years. This renders the occurrence of M74-related symptoms among spawner offspring no longer a suitable indicator for the actual mortality among offspring. Instead, the general information on the occurrence of M74 symptoms among offspring from individual females needs to be combined with information on the actual survival of affected offspring. Because of the cost associated with this type of research, it is not possible to do these analyses for each stock. If the stocks selected for such analyses can be considered random samples from the available stocks, a hierarchical methodology can be applied to predict the mortality among the offspring of female spawners affected by M74 for sampled and unsampled stocks.

The resulting M74 mortality estimates show a similar pattern over the years across the different stocks, with low or relatively low M74 mortality before 1991, high M74 mortality from 1992 to 1996, and decreasing M74 mortality thereafter. The decrease in M74 mortality after 1996 can potentially be explained by natural selection due to genetic differences in the survival capability of female spawners and their respective offspring (Langefors et al. 1998). Selection may therefore have reduced the proportion of females affected by the M74 syndrome in the second generation, after the high levels of M74 mortality removed a large proportion of non-M74-resistant salmon from the gene pool. In addition, differences in genetic background have been pinpointed as the cause of the differences in survival among M74-affected yolk-sac fry (Vuori et al. 2006), possibly explaining the decreasing mortality of offspring from M74-affected females in recent years.

Even though natural selection may have improved the survival probability of the population, it does not remove the actual cause of the M74 syndrome and future outbreaks are still possible. The model presented in this paper can be extended further to include explanatory variables, allowing use of the model to predict future levels of M74 mortality. Short-term prognoses about future M74 mortality can be made based on the thiamine content in the eggs (Vuorinen and Keinänen 1999). The cause of the thiamine deficiency, however, is less certain and several factors may be involved, including changes in the ecosystem and the food-web dynamics (Börjeson and Norrgren 1997). This and the fact that M74 mortality seems very low or absent in Latvian salmon stocks (ICES 2005) indicate that it might be necessary to include a spatial dimension in the model by grouping different salmon stocks depending on similar migration routes or feeding areas (Ikonen 2006). An additional hierarchical structure could then be added across different stock groups.

Acknowledgements

We thank Hans Börjeson (Fiskhälsan FH AB, Sweden) for providing data and information on the occurrence of the M74 syndrome in Swedish salmon stocks and for useful comments on the manuscript. We also thank Lars Karlsson (National Board of Fisheries, Sweden) for helpful suggestions on the model structure and Marja-Liisa Koljonen (Finnish Game and Fisheries Research Institute, Finland) and two anonymous reviewers for their valuable comments on the paper. This study was partly funded by EU Project No. 502516; EFIMAS (Operational Evaluation Tools for Fisheries Management Options).

References

Bengtsson, B., Hill, C., Bergman, A., Brandt, I., Johansson, N., Magnhagen, C., Sodergren, A., and Thulin, J. 1999. Reproductive disturbances in Baltic fish: a synopsis of the FiRe Project. Ambio, **28**: 2–8.

Best, N., Cowles, M.K., and Vines, K. 1995. CODA manual. Version 0.30. MRC Biostatistics Unit, Cambridge, UK.

Beverton, R.J.H., and Holt, S.J. 1956. The theory of fishing. *In* Sea fisheries: their investigation in the United Kingdom. *Edited by* M. Graham. Edward Arnold (Publishers) Ltd., London. pp. 372–441.

Brooks, S.P., and Gelman, A. 1998. Alternative methods for monitoring convergence of iterative simulations. J. Comp. Graph. Stat. **7**: 434–455.

Börjeson, H., and Norrgren, L. 1997. M74 syndrome: a review of potential etiological factors. *In* Chemically induced alterations in functional development and reproduction of fishes. Proceedings from a Session at the 1995 Wingspread Conference. *Edited by*

R.M. Rolland, M. Gilbertson, and R.E. Peterson. The Society of Environmental Toxicology and Chemistry (SETAC), Pensacola, Florida. pp. 153–166.

Cushing, D.H. 1988. The study of stock and recruitment. *In* Fish population dynamics. *Edited by* J.A. Gulland. Wiley, New York. pp. 105–128.

Elliott, J.M. 1993. A 25-year study of production of juvenile sea trout, *Salmo trutta*, in an English Lake District stream. Can. Spec. Publ. Fish. Aquat. Sci. No. 118. pp. 109–122.

Fitzsimons, J.D., Brown, S.B., Honeyfield, D.C., and Hnath, J.G. 1999. A review of early mortality syndrome (EMS) in great lakes salmonids: relationship with thiamine deficiency. Ambio, **28**: 9–15.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, R.B. 1995. Bayesian data analysis. Chapman and Hall, London, UK.

Gilks, W., Richardson, S., and Spiegelhalter, D. (*Editors*). 1995. Introducing Markov chain Monte Carlo. *In* Markov chain Monte Carlo in practice. Chapman & Hall, London, UK. pp. 1–19.

Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman and Hall, New York.

Ikonen, E. 2006. The role of the feeding migration and diet of Atlantic salmon (*Salmo salar* L.) in yolk-sac-fry mortality (M74) in the Baltic Sea. Ph.D. thesis, University of Helsinki, Helsinki, Finland.

International Council for the Exploration of the Sea (ICES). 2005. Report of the Baltic salmon and trout working group (WGBAST). ICES, Copenhagen, Denmark.

Jonsson, N., Jonsson, B., and Hansen, L.P. 1998. The relative role of density-dependent and density-independent survival in the life cycle of Atlantic salmon, *Salmo salar.* J. Anim. Ecol. **67**: 751–762.

Jutila, E., Jokikokko, E., and Julkunen, M. 2003. Management of Atlantic salmon in the Simojoki River, northern Gulf of Bothnia: effects of stocking and fishing regulation. Fish. Res. **64**: 5–17.

Kallio-Nyberg, I., and Koljonen, M. 1997. The genetic consequence of hatchery-rearing on life-history traits of the Atlantic salmon (*Salmo salar* L.): a comparative analysis of sea-ranched salmon with wild and reared parents. Aquaculture, **153**: 207–224.

Langefors, Å., Von Schantz, T., and Widegren B. 1998. Allelic variation of Mhc class II in Atlantic salmon: a population genetic analysis. Heredity, **80**: 568–575.

Liermann, M., and Hilborn, R. 1997. Depensation in fish stocks: a hierarchical Bayesian meta-analysis. Can. J. Fish. Aquat. Sci. **54**: 1976–1984.

Mace, P.M., and Doonan, I.J. 1988. A generalized bioeconomic simulation model for fish population dynamics. N.Z. Fish. Assoc. Res. Doc. No. 88/4.

Marshall, E.C., and Spiegelhalter, D.J. 2003. Approximate cross-validatory predictive checks in disease-mapping models. Stat. Med. **22**: 1649–1660.

Michielsens, C.G.J., and McAllister, M.K. 2004. A Bayesian hierarchical analysis of stock–recruit data: quantifying structural and parameter uncertainties. Can. J. Fish. Aquat. Sci. **61**: 1032–1047.

Michielsens, C.G.J., McAllister, M.K., Kuikka, S., Pakarinen, T., Karlsson, L., Romakkaniemi, A., Perä, I., and Mäntyniemi, S. 2006. Bayesian state–space mark–recapture model to estimate fishing mortality rates within a mixed-stock fishery. Can. J. Fish. Aquat. Sci. **63**: 321–334.

Myers, R.A. 2001. Stock and recruitment: generalizations about maximum reproductive rate, density dependence and variability using meta-analytic approaches. ICES J. Mar. Sci. **58**: 937–951.

Needle, C.L. 2002. Recruitment models: diagnosis and prognosis. Rev. Fish Biol. Fish. **11**: 95–111.

Pearl, J. 2000. Causality: models, reasoning, and inference. Cambridge University Press, Cambridge, UK.

Romakkaniemi, A., Perä, I., Karlsson, L., Jutila, E., Carlsson, U., and Pakarinen, T. 2003. Development of wild Atlantic salmon stocks in the rivers of the northern Baltic Sea in response to management actions. ICES J. Mar. Sci. **60**: 329–342.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. 2003. WinBUGS. Version 1.4 user manual. MRC Biostatistics Unit, Cambridge, UK.

Vuori, K., Koskinen, H., Krasnov, A., Koivumaki, P., Afanasyev, S., Vuorinen, P.J., and Nikinmaa, M. 2006. Development disturbances in early life stage mortality (M74) of Baltic salmon fry as studied by changes in gene expression. BMC Genomics, **7**: 56.

Vuorinen, P.J., and Keinänen, M. 1999. Environmental toxicants and thiamine in connection with the M74 syndrome in Baltic salmon (*Salmo salar*). *In* Nordic research cooperation on reproductive disturbances in fish. Report from the Redfish Project. *Edited by* B.E. Bengtsson, C. Hill, and S. Nellbring. Nordic Council of Ministers, Copenhagen, Denmark. pp. 25–37.

Walters, C.J., and Ludwig, D. 1981. Effects of measurement errors on the assessment of stock–recruit relationships. Can. J. Fish. Aquat. Sci. **38**: 704–710.

Appendix A

In this appendix we explain the use of a beta–binomial distribution for predicting the number of eggs that die because of an early life history mortality syndrome (EMS), under the assumption that the survival probability varies between females.

We denote the number of females having EMS symptoms by N and the number of eggs produced by each female $i = 1,\ldots, N$ by O_i. Eggs from a female are assumed to have a female-specific probability φ_i to die because of EMS. Assuming that all eggs from one female die independently of each other, the number of dying eggs from female i, d_i, follows a binomial distribution, given that the number of eggs and the probability of dying are known, i.e.,

$$d_i \mid O_i, \varphi_i \sim \mathrm{Bin}(O_i, \varphi_i)$$

Assuming that parameters φ_i are exchangeable, we can act as if they were an independent sample from some density function. Here we choose to use the beta distribution family to describe the variation between females:

$$\varphi_i \mid \alpha, \beta \sim \mathrm{Beta}(\alpha, \beta)$$

with $\alpha = \mu_\varphi \eta_\varphi$ and $\beta = (1 - \mu_\varphi)\eta_\varphi$. Parameter μ_φ denotes the mean across females and parameter η_φ describes the variation between females. Increasing parameter η_φ decreases the variance, and vice versa. According to standard distribution theory, the conditional distribution of d_i given parameters O_i, μ_φ, and η_φ is the mixture of beta and binomial distributions, i.e., the beta–binomial distribution

$$d_i \mid O_i, \mu_\varphi, \eta_\varphi \sim \mathrm{Beta\text{–}Bin}(O_i, \mu_\varphi, \eta_\varphi)$$

The mean of a beta–binomial distribution is given by

$$E(d_i \mid O_i, \mu_\varphi, \eta_\varphi) = O_i \frac{\alpha}{\alpha + \beta} = O_i \mu_\varphi$$

and the variance is given by

$$V(d_i \mid O_i, \mu_\varphi, \eta_\varphi) = O_i \frac{\alpha\beta(\alpha+\beta+O_i)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$= O_i \frac{\mu_\varphi(1-\mu_\varphi)(\eta_\varphi + O_i)}{\eta_\varphi + 1}$$

Then, assuming that each female has the same number of eggs ($O_a = O_1 = O_2 = \ldots = O_N$) and that the deaths among offspring of different females are conditionally independent given parameters O_i, μ_φ, and η_φ, the mean and variance of the total number of dying eggs $D = \sum_{i=1}^{N} d_i$ can be calculated as

$$E(D \mid O_a, \mu_\varphi, \eta_\varphi, N) = \sum_{i=1}^{N} E(d_i \mid O_a, \mu_\varphi, \eta_\varphi) = N O_a \mu_\varphi$$

$$V(D \mid O_a, \mu_\varphi, \eta_\varphi, N) = \sum_{i=1}^{N} V(d_i \mid O_a, \mu_\varphi, \eta_\varphi, N)$$

$$= N O_a \frac{\mu_\varphi(1-\mu_\varphi)(\eta_\varphi + O_a)}{\eta_\varphi + 1}$$

Then, by introducing a new parameter

$$\eta^* = \frac{(n_\varphi + 1) O_a N - (\eta_\varphi + O_a)}{O_a - 1}$$

the variance of the total number of dying eggs can be expressed also in the form

$$V(D \mid O_a, \mu_\varphi, \eta^*, N) = N O_a \frac{\mu_\varphi(1-\mu_\varphi)(\eta^* + N O_a)}{\eta^* + 1}$$

which corresponds to the variance of the beta–binomial distribution

$$D \mid O_a, N, \alpha^* \beta^* \sim \text{Beta-Bin}(N O_a, \alpha^*, \beta^*)$$

$$\alpha^* = \mu_\varphi \left(\frac{(\eta_\varphi + 1) O_a N - (\eta_\varphi + O_a)}{O_a - 1} \right)$$

$$\beta^* = (1-\mu_\varphi) \left(\frac{(\eta_\varphi + 1) O_a N - (\eta_\varphi + O_a)}{O_a - 1} \right)$$

which gives the correct mean, variance, and range for D. Therefore, the above distribution can be used to approximate the distribution of $D = \sum_{i=1}^{N} d_i$. In simulation experiments, a beta–binomial distribution parameterised in this way has given a very close approximation for the true distribution of the sum of independent beta–binomials.