A hierarchical model to estimate the abundance and biomass of salmonids by using removal sampling and biometric data from multiple locations

Philippe Ruiz, Christophe Laplanche

A hierarchical model to estimate the abundance and biomass of salmonids by using removal sampling and biometric data from multiple locations

Philippe Ruiz and Christophe Laplanche

Abstract: We present a Bayesian hierarchical model to estimate the abundance and the biomass of brown trout (*Salmo trutta fario*) by using removal sampling and biometric data collected at several stream sections. The model accounts for (*i*) variability of the abundance with fish length (as a distribution mixture), (*ii*) spatial variability of the abundance, (*iii*) variability of the catchability with fish length (as a logit regression model), (*iv*) spatial variability of the catchability, and (*v*) residual variability of the catchability with fish. Model measured variables are the areas of the stream sections as well as the length and the weight of the caught fish. We first test the model by using a simulated dataset before using a 3-location, 2-removal sampling dataset collected in the field. Fifteen model alternatives are compared with an index of complexity and fit by using the field dataset. The selected model accounts for variability of the abundance with fish length and stream section and variability of the catchability with fish length. By using the selected model, 95% credible interval estimates of the abundances at the three stream sections are (0.46,0.59), (0.90,1.07), and (0.56,0.69) fish/m^2. Respective biomass estimates are (9.68, 13.58), (17.22, 22.71), and (12.69, 17.31) g/m^2.

Résumé : Nous présentons un modèle hiérarchique bayésien pour estimer l'abondance et la biomasse de truites brunes (*Salmo trutta fario*) basé sur un échantillonnage par retraits et des données biométriques récoltées sur plusieurs sections de cours d'eau. Le modèle tient compte de (*i*) la variabilité de l'abondance en fonction de la longueur des poissons (comme une distribution de mélange), (*ii*) la variabilité spatiale de l'abondance, (*iii*) la variabilité de la capturabilité en fonction de la longueur du poisson (comme modèle de régression logit), (*iv*) la variation spatiale de la capturabilité et (*v*) la variabilité résiduelle de la capturabilité en fonction des poissons. Les variables mesurées du modèle incluent les surfaces des sections de cours d'eau, ainsi que la longueur et la masse des poissons capturés. Nous testons le modèle à l'aide d'une banque de données simulées avant d'utiliser un ensemble de données récoltées à trois sites et en deux échantillonnages par retraits en nature. Nous comparons quinze modèles de rechange avec un indice de complexité et d'ajustement aux données de terrain. Le modèle retenu tient compte de la variabilité de l'abondance en fonction de la longueur des poissons et de la section de cours d'eau, ainsi que de la variabilité de la capturabilité en fonction de la longueur des poissons (sans la variabilité spatiale ou résiduelle de la capturabilité). Dans le modèle retenu, les estimations de l'intervalle crédible au niveau de 95 % de l'abondance dans les trois sections de cours d'eau sont (0,46; 0,59), (0,90; 1,07) et (0,56; 0,69) poissons/m^2. Les estimations de biomasses correspondantes sont (9,68; 13,58), (17,22; 22.71) et (12,69; 17,31) g/m^2.

[Traduit par la Rédaction]

Introduction

Alpine freshwater systems are subject to anthropogenic disturbances, e.g., power dams (Petts 1984; The World Commission on Dams 2000; Petts and Gurnell 2005) and global warming (Xenopoulos et al. 2005; Buisson et al. 2008; Jonsson and Jonsson 2009). Fish can be used as a bioindicator of the health of freshwater sytems, and for that purpose freshwater ecologists have considered various variables indicative of fish populations, for instance abundance (number of fish per area unit of stream) (Heimbuch et al. 1997; Wyatt 2002), biomass (mass of fish per area unit of stream) and growth rate (Penczak et al. 1981), mortality rate (Gouraud et al. 2001), recruitment (Lobón-Cerviá 2009), or production (Kwak and Waters 1997).

Fish ecologists can use removal sampling by electrofishing to derive estimates of fish abundance (Lobón-Cerviá

P. Ruiz. Ecolab (UMR 5245), Université Paul Sabatier, 29 Rue Jeanne Marvig, B.P. 24349, 31055 Toulouse, France.
C. Laplanche.[1] Ecolab (UMR 5245), École Nationale Supérieure Agronomique de Toulouse, Avenue de l'Agrobiopole, B.P. 32607 Auzeville-Tolosane, 31326 Castanet, France.

[1] Corresponding author (e-mail: laplanche@gmail.com).

1991). Fish ecologists have created various Bayesian hierarchical models to statistically relate removal sampling data to abundance. The most simple model considers removal sampling data collected at a single stream section and bears on the assumption that catchability is equal among fish and across removals (Moran 1951; Cowx 1983; Bedrick 1994). This model has been popular among fish ecologists since the work of Carle and Strub (1978) who presented an algorithm to efficently compute a maximum likelihood estimate of the abundance. This approach actually provides inaccurate estimates of fish abundance (Riley and Fausch 1992; Peterson et al. 2004; Laplanche 2010). The main source of inaccuracy is unaccounted variability of the catchability or of the abundance. Several authors have therefore presented more advanced hierarchical models to estimate fish abundance from removal sampling data. Such models consider variability of the catchability between stream sections (Wyatt 2002, 2003), residual among fish (Dorazio et al. 2005; Mäntyniemi et al. 2005; Laplanche 2010), and variability of the abundance between stream sections (Wyatt 2002, 2003; Dorazio et al. 2008).

Catchability (Mahon 1980; Anderson 1995) and abundance (Pauly and Moreau 1997; Pitcher 2002) also fluctuate with fish length. The motivations to account for such variabilities within a Bayesian hierarchical model to derive abundance estimates are 3-fold. First, unaccounted variability of the catchability with fish length could lead to inaccurate abundance estimates. In the aim of providing accurate estimates, such a variability should be included (or at least its significance evaluated) within a removal sampling model. Second, the Bayesian approach makes possible the handling of complex mixed-effect, multilevel, nonlinear models (Congdon 2006). In this case, it is possible to consider multilevel, nonlinear length–catchability and length–abundance relationships. Variability of the abundance with fish length can be modeled as a mixture of distributions (Pauly and Moreau 1997; Pitcher 2002). Freshwater biologists usually decompose the length–abundance plots by using the method suggested by Bhattacharya (1967). Such a decomposition could be efficiently carried out within a Bayesian framework (Marin et al. 2005). Third, the description of the abundance as a function of fish length is the key to derive other variables indicative of fish populations, for instance growth rate (Pauly and Moreau 1997). Since weight plots from samples of fish populations show a well-pronounced linear relationship on a log–log scale (Cone 1989; Reiss 1989), i.e., fish weight is proportional to a power of fish length, biomass can be computed as the cross-product of the estimated abundance by the expected fish weight (Kwak and Waters 1997; (Lobón-Cerviá 2009).

We present a hierarchical Bayesian model that we use to estimate the abundance and the biomass of fish in an alpine stream by sampling at several stream sections. The model accounts for (i) variability of the abundance with fish length, (ii) random spatial variability of the abundance, (iii) variability of the catchability with fish length, (iv) random spatial variability of the catchability, and (v) residual variability of the catchability with fish. The data that is required to run the model is first described. The model is the combination of three submodels (abundance, weight, biomass) that are presented one after the other. We use the model with two data-

sets: a simulated dataset and a dataset collected in the field. We use the simulated dataset to compare parameter estimates with true values. We use the dataset collected in the field to illustrate the capability of the model to estimate the abundance and the biomass of brown trout in alpine streams. The model could be extended to estimate additional variables indicative of fish populations, such as growth rate, which is discussed.

Materials and methods

Measured variables

We consider s sections of a stream. Each stream section is depleted by electrofishing in r removals (Lobón-Cerviá 1991). Let A_h (m^2) be the area of the stream section $h \in \{1, \ldots, s\}$. Let $C_{h,j}$ be the number of fish caught during removal $j \in \{1, \ldots, r\}$ in section h and $C_h = \sum_{j=1}^{r} C_{h,j}$ be the total number of fish caught in section h. Let $L_{h,j,f}$ (mm) and $W_{h,j,f}$ (g) be the length and the weight of the fish $f \in \{1, \ldots, C_{hj}\}$ caught in section h during removal j. Fish are grouped by length class of width Δ_l (mm). The class width Δ_l can be equal to the measurement accuracy of the fish length (e.g., 1 mm) or larger. Let m be the number of length classes, $[(i - 1)\Delta_l, i\Delta_l[$ the classes, and $L_i = (i - 1/2)\Delta_l$ the class centers ($i \in \{1, \ldots, m\}$). Let $C_{h,i,j}$ be the number of fish of length class i caught during removal j in section h. The measured variables considered in the following are A_h, $C_{h,i,j}$, L_i, $L_{h,j,f}$, and $W_{h,j,f}$ (Table 1).

Model structure

The model is structured in six hierarchical levels: stream (the all-embracing level), stream section ($h \in \{1, \ldots, s\}$), mixture component ($k \in \{1, \ldots, q\}$), length class ($i \in \{1, \ldots, m\}$), removal ($j \in \{1, \ldots, r\}$), and fish ($f \in \{1, \ldots, C_{h,j}\}$). The mixture component level (defined later) is part within the stream and section levels. Parameter notations are brought together (Table 1). Parameters at the stream (stream–component), section (section–component), length, and removal levels are of dimension 1 (q), s (sq), sm, and smr, respectively.

The model is the combination of three submodels (abundance, weight, and biomass) that are presented successively. The abundance and the biomass submodels are illustrated as directed acyclic graphs (DAG) (Figs. 1 and 2, respectively). The measured variables A_h, $C_{h,i,j}$, and L_i are used in the abundance submodel. The measured variables $L_{h,j,f}$ and $W_{h,j,f}$ are used in the weight submodel. The biomass submodel uses the measured variables A_h and L_i plus six variables (τ_k, $\tau_{h,k}$, $n_{h,i}$, c_h, d_h, s_h^2) of the abundance and weight submodels.

Abundance submodel

The submodel (Fig. 1) combines 22 mono- and multi-dimensional stochastic nodes: 13 at the stream level (λ, τ_k, μ_k, σ_k, a, b, η, π_λ, π_τ, π_μ, π_σ, π_a, π_b), 6 at the section level (λ_h, $\tau_{h,k}$, $\mu_{h,k}$, $\sigma_{h,k}$, a_h, b_h), 1 at the length level ($n_{h,i}$), and 1 at the removal level ($C_{h,i,j}$). The model comprehends 8 additional deterministic nodes: 1 at the stream level (λ_k), 2 at the section level (A_h, $\lambda_{h,k}$), 3 at the length level (L_i, $\lambda_{h,i}$, $\lambda_{h,i,k}$), and 2 at the removal level ($n_{h,i,j}$, $p_{h,i,j}$).

Table 1. Model parameters.

Parameter	Stream	Section	Size class	Component	Removal	Fish
Index	—	$h \in \{1, \ldots, s\}$	$i \in \{1, \ldots, m\}$	$k \in \{1, \ldots, q\}$	$j \in \{1, \ldots, r\}$	$f \in \{1, \ldots, n_{h,ij}\}$
Measured variables*						
Area	—	A_h	—	—	—	—
Catch	—	C_h	$C_{h,i}$	—	$C_{h,i,j}$	—
Class center	—	—	L_i	—	—	—
Fish length	—	—	—	—	—	$L_{h,i,j,f}$
Fish weight	—	—	—	—	—	$W_{h,i,j,f}$
Abundance†						
Population size	—	n_h	$n_{h,i}$	—	$n_{h,i,j}$	—
Abundance	λ, λ_k	$\lambda_h, \lambda_{h,k}$	$\lambda_{h,i}, \lambda_{h,i,k}$	$\lambda_k, \lambda_{h,k}, \lambda_{h,i,k}$	—	—
Comp. proportion	τ_k	$\tau_{h,k}$	—	$\tau_k, \tau_{h,k}$	—	—
Comp. center	μ_k	$\mu_{h,k}$	—	$\mu_k, \mu_{h,k}$	—	—
Comp. bandwidth	σ_k	$\sigma_{h,k}$	—	$\sigma_k, \sigma_{h,k}$	—	—
Precisions	$\pi_\lambda, \pi_\tau, \pi_\mu, \pi_\sigma$	—	—	—	—	—
Catchability†						
Catchability	—	—	—	—	$p_{h,i,j}$	$p_{h,i,j,f}$
Slope, intercept	a, b	a_h, b_h	—	—	—	—
Variability	$1/\eta$	$1/\eta_h$	$1/\eta_{h,i}$	—	$1/\eta_{h,i,j}$	—
Precisions	π_a, π_b	—	—	—	—	—
Biomass†						
Allometric param.	c, d	c_h, d_h	—	—	—	—
Residual variance	s^2	s_h^2	—	—	—	—
Biomass	$\text{biom}, \text{biom}_k$	$\text{biom}_h, \text{biom}_{h,k}$	$\text{biom}_{h,i}$	$\text{biom}_k, \text{biom}_{h,k}$	—	—

Note: Parameters at the stream (stream–component), section (section–component), length, removal, and fish are of dimension 1 (q), s (sq), sm, smr, $smrn_{h,i,j}$, respectively.

*Measured variable.

†Latent variable.

Fig. 1. Directed acyclic graphs of the abundance submodel. Frames indicate hierarchical levels: section ($h \in \{1, \ldots, s\}$), length class ($i \in \{1, \ldots, m\}$), removal ($j \in \{1, \ldots, r\}$), and mixture component ($k \in \{1, \ldots, q\}$). Variables outside frames are variables at the stream hierarchical level. Rectangles, deterministic nodes; ellipses, stochastic nodes; dark filled nodes, observed variables; light filled nodes, output variables ($\tau_k, \tau_{h,k}, n_{h,i}$) to the biomass submodel.
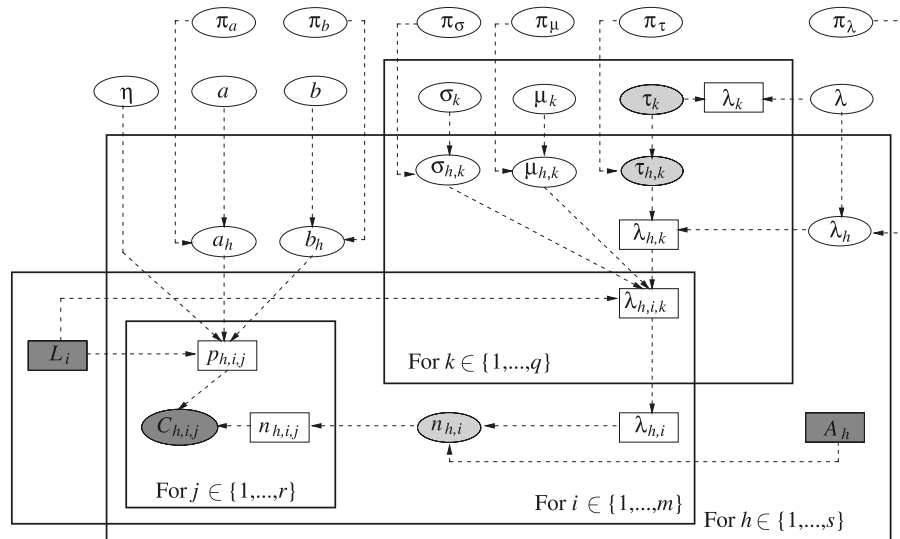
Removal sampling
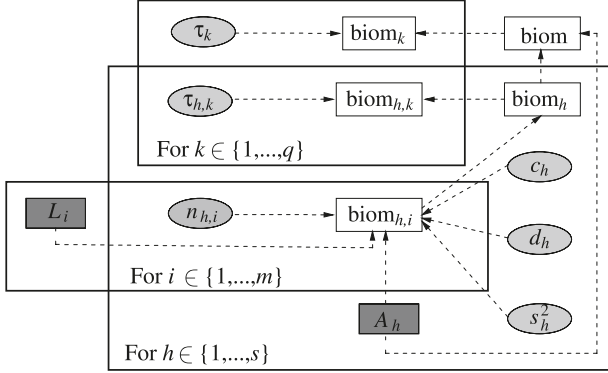
Let $n_{h,i}$ be the number of fish of length class i in section h, and $n_{h,i,j}$ be the number of fish of length class i remaining in section h before removal j, i.e., $n_{h,i,1} = n_{h,i}$ and $n_{h,i,j} = n_{h,i,j-1} - C_{h,ij-1}$ for $j \geq 2$. Let $p_{h,i,j}$ be the expectation of the catchability $p_{h,i,j,f}$ of the fish $f \in \{1, \ldots, n_{h,i,j}\}$ of length class

Fig. 2. Directed acyclic graphs of the biomass submodel. Hierarchical levels: section ($h \in \{1, \ldots, s\}$), length class ($i \in \{1, \ldots, m\}$), and mixture component ($k \in \{1, \ldots, q\}$). Light filled nodes, input variables (τ_k, $\tau_{h,k}$, $n_{h,i}$) from the abundance submodel and input variables (c_h, d_h, s_h) from the weight submodel.

i remaining in section h before removal j. Under several hypotheses (given later), the number of fish of length class i caught during removal j in section h is a binomial

$$(1) \qquad C_{h,i,j}|p_{h,i,j}, n_{h,i,j} \sim \text{Binomial}(p_{h,i,j}, n_{h,i,j}) \qquad \forall h, i, j$$

The modeling of the variability of the catchability $p_{h,i,j,f}$ (between length class, stream section, and residual) and of the population sizes $n_{h,i}$ (between length class and stream section) is detailed below.

Heterogeneity of the catchability between length class

We model variability of the catchability with fish length. To also model residual variability of the catchability between individuals of a same length class, we model variability with length of the expectation of the catchability during the first removal $p_{h,i,1}$. It is related to fish length by the logit regression model

$$(2) \qquad \text{logit}(p_{h,i,1}) = a_h L_i + b_h \qquad \forall h, i$$

The catchability $p_{h,i,1}$ increases with fish length if $a_h > 0$, decreases if $a_h < 0$, and does not vary with fish length if $a_h = 0$. The catchability $p_{h,i,1} = 1/2$ at the (algebraic) length $-b_h/a_h$.

Heterogeneity of the catchability within length class

We model residual individual variability of the catchability within a length class by assuming that the probabilities $p_{h,i,1,f}$ of capturing the fish $f \in \{1, \ldots, n_{h,i}\}$ (of length class i in section h) during the first removal are independent and beta distributed. We also assume that the probabilities of capturing any fish f during removal $j \geq 2$ given that the fish has not been captured earlier is equal to $p_{h,i,1,f}$. Under such hypotheses, Mäntyniemi et al. (2005) have shown that $p_{h,i,j,f}$ for $j \geq 2$ are still independent and beta distributed. As a result the number of caught fish $C_{h,i,j}$ is a binomial (hence eq. 1). Mäntyniemi et al. (2005) have shown that the expectation of $p_{h,i,j,f}$ is

$$(3) \qquad p_{h,i,j} = \frac{p_{h,i,1}}{1 + (j-1)/\eta_{h,i,1}} \qquad \forall h, i, j$$

where $p_{h,i,1} = \alpha_{h,i,1}/\eta_{h,i,1}$, $\eta_{h,i,1} = \alpha_{h,i,1} + \beta_{h,i,1}$, $\alpha_{h,i,1}$, and $\beta_{h,i,1}$ are

Table 2. Conditional distribution of the parameters at the section level.

Parameters
$\log(\lambda_h)|\lambda, \pi_\lambda \sim \text{Normal}[\log(\lambda), \pi_\lambda]$
$\text{logit}(\tau'_{h,k})|\tau_k, \pi_\tau \sim \text{Normal}[\text{logit}(\tau_k), \pi_\tau]$
$\text{logit}\left(\frac{\mu_{h,k}-\mu_{k-1}^{\lim}}{\mu_k^{\lim}-\mu_{k-1}^{\lim}}\right)|\mu_k, \pi_\mu \sim \text{Normal}\left[\text{logit}\left(\frac{\mu_k-\mu_{k-1}^{\lim}}{\mu_k^{\lim}-\mu_{k-1}^{\lim}}\right), \pi_\mu\right]$
$\log(\sigma_{h,k})|\sigma_k, \pi_\sigma \sim \text{Normal}[\log(\sigma_k), \pi_\sigma]$
$\log(a_h)|a, \pi_a \sim \text{Normal}[\log(a), \pi_a]$
$\log(b_h)|b, \pi_b \sim \text{Normal}[\log(b), \pi_b]$

Note: The mixture component proportions are $\tau_{h,k} = \tau'_{h,k}/\sum_{k'=1}^q \tau'_{hk'}$ ($h \in \{1, \ldots, s\}$, $k \in \{1, \ldots, q\}$).

the shape parameters of the beta distribution of $p_{h,i,1,f}$. For practical reasons, we parametrize the beta distribution of $p_{h,i,1,f}$ by using $p_{h,i,1}$ and $1/\eta_{h,i,1}$ instead of $\alpha_{h,i,1}$ and $\beta_{h,i,1}$. Indeed, $1/\eta_{h,i,1} = \delta_{h,i,1}/[p_{h,i,1}(1 - p_{h,i,1}) - \delta_{h,i,1}]$ where $\delta_{h,i,1}$ is the variance of $p_{h,i,1,f}$, $1/\eta_{h,i,1} = 0$ if there is no variability of the catchability between individuals. We assume in the following that the parameters $1/\eta_{h,i,1}$ are equal for all size classes and stream sections, hence $1/\eta_{h,i,1} = 1/\eta_h = 1/\eta$.

Size structure

The population sizes $n_{h,i}$ are taken to be independent Poisson random variables (Wyatt 2002)

$$(4) \qquad n_{h,i}|\lambda_{h,i}, A_h \sim \text{Poisson}(\lambda_{h,i} A_h)$$

where $\lambda_{h,i}$ and $\lambda_{h,i}A_h$ are the abundance and the expected number of fish of length class i in section h. We model variability of the abundance with fish length as a mixture of distributions (Pitcher 2002). We consider several distributions (Epanechnikov, biweight, triweight, and Gaussian) and denote K the respective kernels (Wand and Jones 1990). For instance, the Gaussian kernel is

$$(5) \qquad K(l) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{l^2}{2}\right) \qquad \forall l \in \mathbb{R}$$

The expected number of fish of length class i in section h is

$$(6) \qquad \lambda_{h,i} A_h = \lambda_h A_h \Delta_l \sum_{k=1}^q \frac{\tau_{h,k}}{\sigma_{h,k}} K\left(\frac{L_i - \mu_{h,k}}{\sigma_{h,k}}\right) \qquad \forall h, i$$

where λ_h and $\lambda_h A_h$ are the abundance and the expected number of fish in section h, q is the number of mixture components, $\mu_{h,k}$, $\sigma_{h,k}$, and $\tau_{h,k}$ ($k \in \{1, \ldots, q\}$) are the centers, the bandwiths, and the relative proportions of the mixture components, respectively. We also define $\lambda_{h,k} = \lambda_h \tau_{h,k}$ as the abundance of fish in section h of mixture component k. In the case of a Gaussian kernel, the centers and the bandwidths are expectations and standard deviations. The proportions are related by the constraint $\sum_{k=1}^q \tau_{h,k} = 1$ for all h. We sort the mixture components by increasing center, i.e., $\mu_{h,k} < \mu_{h,k+1}$ for all h and all $k \in \{1, \ldots, q-1\}$.

Variability among stream sections

We consider random variations among sections of the parameters related to the catchability (a_h, b_h) and the abundance (λ_h, $\tau_{h,k}$, $\mu_{h,k}$, $\sigma_{h,k}$). We provide additional expert

Table 3. Free parameters (stream level) are assigned vague priors.

Parameter	Prior	Marginal posterior
λ	Gamma(0.001, 0.001)	Lognormal(–0.35, 4)
τ_k'	Uniform(0,1)	Beta([27, 21, 11], [35, 39, 38])
μ_k	Uniform($\mu_{k-1}^{\lim}, \mu_k^{\lim}$)	Normal([53, 121, 188], [0.1, 0.3, 0.03])
σ_k^2	InvGamma(0.001, 0.001)	Lognormal([3.9, 5.4, 6.9], [37, 29, 15])
a	Normal(0, 0.001)	Normal(7.6, 0.14)
b	Normal(0, 0.001)	Normal(0.44, 6.6)
$1/\eta$	InvGamma(0.001, 0.001)	Beta(0.2, 1.4)
π_λ	Gamma(0.001, 0.001)	Gamma(0.9, 0.09)
π_τ	Gamma(0.001, 0.001)	Gamma(1.2, 0.25)
π_μ	Gamma(0.001, 0.001)	Gamma(2.0, 0.06)
π_σ	Gamma(0.001, 0.001)	Gamma(0.4, 0002)
π_a	Gamma(0.001, 0.001)	Gamma(0.1, 0.001)
π_b	Gamma(0.001, 0.001)	Gamma(0.3, 0.001)

Note: Parameters are shape and rate for gamma distributions, expectation and precision for normal and lognormal distributions, and shapes for β-distributions. The mixture component proportions are $\tau_k = \tau_k'/\sum_{k'=1}^{q} \tau_{k'}'$ ($k \in \{1, \ldots, q\}$). The last column provides the approximate distribution of the marginal posteriors of the free parameters (values within brakets refer to the first, second, and third mixture components, respectively) by using model alternative 1 with the Neste d'Oueil dataset.

knowledge that the definition sets of the mixture component centers do not overlap and have boundaries common for all sections. As a result, we add the constraints $\mu_{k-1}^{\lim} < \mu_{h,k} < \mu_k^{\lim}$ for all section h and mixture component k, with $\mu_0^{\lim} = 0$. The parameters at the section level are therefore bound to the constraints: $0 < \lambda_h$, $0 < \tau_{h,k} < 1$ with $\sum_{k=1}^{q} \tau_{h,k} = 1$, $\mu_{k-1}^{\lim} < \mu_{h,k} < \mu_k^{\lim}$, $0 < \sigma_{h,k}$. We log-transform, scale, and logit-transform the above parameters and assign normal distributions of expectations the respective values at the stream level (Table 2). We label π_λ, π_τ, π_μ, π_σ, π_a, π_b the precisions (i.e., the inverse of the variances) of the normal distributions. We assume that the precisions of the normal distributions related to $\tau_{h,k}$, $\mu_{h,k}$, and $\sigma_{h,k}$ are equal for all k.

Priors

The free parameters are λ, τ_k, μ_k, σ_k, a, b, η, and the precisions π_λ, π_τ, π_μ, π_σ, π_a, π_b. We apply the constraints $\mu_{k-1}^{\lim} < \mu_k < \mu_k^{\lim}$ on the mixture component centers. We assign vague priors to all free parameters (Table 3). The abundance of fish of mixture component k is $\lambda_k = \lambda \tau_k$.

Weight submodel

We relate the weight and the length of the fish $f \in \{1, \ldots, C_{h,j}\}$ caught during removal j in section h by the allometric relationship

$$(7) \qquad W_{h,j,f} = c_h L_{h,j,f}^{d_h} \exp(\epsilon_{h,j,f}) \quad \forall h, j, f$$

with $c_h|c$, $\pi_c \sim$ Normal(c, π_c) and $d_h|d$, $\pi_d \sim$ Normal(d, π_d), i.e., over- or under-weight of fish is allowed to vary between stream sections. The error in eq. 7 is lognormal, $\epsilon_{h,j,f}|s_h^2 \sim$ Normal(0, $1/s_h^2$). The error terms $c_h - c|\pi_c$, $d_h - d|\pi_d$, and $\epsilon_{h,j,f}|s_h^2$ are taken independent. Log-transformed, the weight submodel is a two-level linear mixed-effect model with a normal, heteroscedastic residual error (Pinheiro and Bates 2000).

The weight submodel that has just been described applies to a subset of C_h fish among the n_h fish present in section h. We assume that the fish used for model selection and parameter estimation are representative, in terms of length and weight, of the fish present in section h. In that case, the weight of the fish $f \in \{1, \ldots, n_h\}$ of section h are provided by the same weight submodel.

Biomass submodel

The biomass $\text{biom}_{h,i}$ of fish of length class i in section h is

$$(8) \qquad \text{biom}_{h,i} = \sum_{j=1}^{r} \sum_{f=1}^{n_{h,i}} W_{h,j,f}/A_h \quad \forall h, i$$

which is approximately equal to eq. 9 (as shown in Appendix A and discussed later)

$$(9) \qquad \text{biom}_{h,i} = n_{h,i} c_h L_i^{d_h} \exp(s_h^2/2)/A_h \quad \forall h, i$$

The biomass of fish in section h is

$$(10) \qquad \text{biom}_h = \sum_{i=1}^{m} \text{biom}_{h,i} \quad \forall h$$

The average biomass of fish (stream level) is

$$(11) \qquad \text{biom} = \sum_{h=1}^{s} \text{biom}_h A_h / \sum_{h=1}^{s} A_h$$

and the biomass of fish in mixture component k is $\text{biom}_k = \tau_k \text{biom}$. The relationship between the biomass variables, the measured variables (A_h, L_i), and the input variables from the abundance submodel (τ_k, $\tau_{h,k}$, $n_{h,i}$) and from the weight submodel (c_h, d_h, s_h^2) are illustrated by a DAG (Fig. 2).

Computations

Computations of the three submodels can be carried out simultaneously by merging the submodels into a single hierarchical model. Computations of the three submodels can also be carried out successively. We chose to first carry out

Fig. 3. Number of caught fish $C_{h,i,j}$ per Δ_l = 10 mm length class (simulated dataset). The catch $C_{h,i,j}$ is the jth stacked sub-bar making up the bar of ith length class (x-axis) of the hth subplot. The true (broken line) and estimated (unbroken line) value of the expected population sizes $\lambda_{h,i}A_h\Delta_l$ are computed by using eq. 6 with the true value and the point estimates of the parameters λ_h, $\tau_{h,k}$, $\mu_{h,k}$, and $\sigma_{h,k}$ (Table 6).
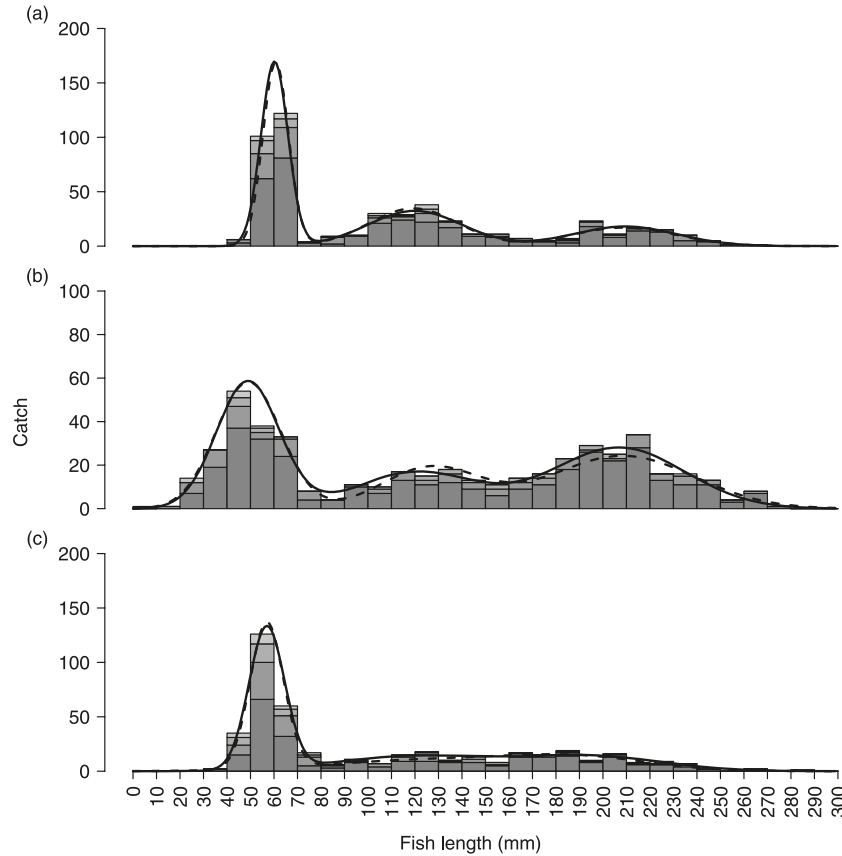
the computations related to the weight submodel, before jointly carring out the computations related to the abundance and biomass submodels (referred to in the following as the abundance–biomass model).

Models have been implemented by using OpenBUGS, open source version of WinBUGS (Lunn et al. 2009; Ntzoufras 2009). Computations of the weight submodel could, however, have been carried out by using R (Pinheiro and Bates 2000). Samples of the posterior are generated by using a Markov chain Monte Carlo (MCMC) method (Robert and Casella 2004). Samples are processed by using R (Crawley 2007). WinBUGS and R scripts are, together with data files, available online in the supplementary data[2], see Appendix S1. Reported point estimates of the parameters are posterior expectation estimates. Interval estimates are 2.5% and 97.5% quantile estimates of marginal posteriors. Convergence was investigated by using the ANOVA-type diagnostic described by Gelman and Rubin (1992) with three chains. Independent samples were obtained by thinning, guided through the examination of the autocorrelation functions of the posterior samples. One thousand independent posterior samples were generated for each model.

Alternative models are compared in terms of the Deviance Information Criterion, DIC $= \bar{D} + p_D$, where $\bar{D}$ is the posterior expectation of the deviance statistics and p_D is a count of effective parameters (Spiegelhalter et al. 2002). Weight

submodel alternatives are compared by using the DIC routine provided by OpenBUGS. The computation of the DIC is, however, not straightforward when using discrete variables. The conditional distribution of $C_{h,i,j}$ is discrete (binomial), for that reason we have not compared abundance–biomass models by using the DIC routine provided by OpenBUGS. We have computed approximate DIC estimates by approximating the count of effective parameters p_D by the true number of parameters.

Datasets

Simulated dataset

We simulate a removal sampling dataset ($s = 3$, $m = 299$, $r = 4$, $q = 3$ Gaussian components, $\Delta_l = 1$ mm). The areas of the stream sections are $A_1 = A_2 = A_3 = 500$ m². We set the values of the free parameters (stream level) to simulate parameter values at the section, length, and removal levels. The weight of the fish are simulated by using the weight submodel with parameter values found by using the field dataset (described below). The number of caught fish are $C_1 = 496$, $C_2 = 474$, and $C_3 = 434$. The number of caught fish $C_{h,i,j}$, as well as the true value of the expected population sizes $\lambda_{h,i}A_h\Delta_l$, are illustrated by a length–abundance plot (Fig. 3).

---

[2] Supplementary data for this article are available on the journal Web site (http://cjfas.nrc.ca).

Table 4. Abundance point estimates (section level) by using the simulated dataset with r ∈ {2, 3, 4} removals and length class widths $\Delta_l \in \{1, 2, 5, 10, 20\}$ mm.

	r = 2			r = 3			r = 4		
Δ_l	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
1	1.02	1.01	0.94	1.00	0.97	0.94	1.01	0.97	0.95
2	1.04	1.02	0.96	1.01	0.97	0.94	1.01	0.97	0.95
5	1.05	1.04	0.97	1.01	0.97	0.94	1.01	0.97	0.95
10	1.04	1.02	0.97	1.01	0.97	0.94	1.01	0.97	0.95
20	1.04	1.00	0.96	1.03	0.97	0.95	1.04	0.97	0.97

Note: True values are $\lambda_1 = 0.98$, $\lambda_2 = 0.97$, and $\lambda_3 = 0.87$ fish/m^2.

Fig. 4. Number of caught fish $C_{h,i,j}$ per $\Delta_l = 10$ mm length class at the Neste d'Oueil in July 2006 at Cires (*a*), Mayrègne (*b*), and Saint-Paul (*c*). The estimated expected population sizes (solid curve) is computed by using eq. 6 with the point estimates of the parameters λ_h, $\tau_{h,k}$, $\mu_{h,k}$, and $\sigma_{h,k}$ (Table 8).
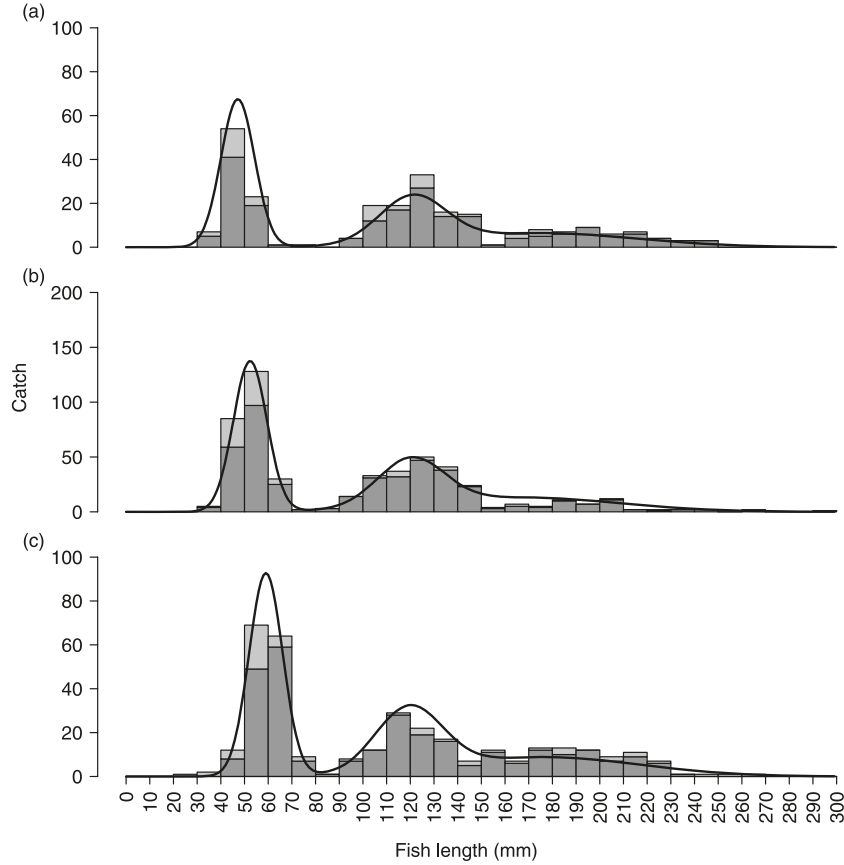
We investigate the consequences of degrading the information brought by the data by lowering the number of removals and enlarging the length class width. For that purpose, we reduce the original $\Delta_l = 1$ mm, 4-removal sampling dataset into 14 additionnal datasets ($r \in \{2, 3, 4\}$, $\Delta_l \in \{1, 2, 5, 10, 20\}$ mm, Table 4).

Field dataset

Data were collected in July 2006 at three locations on the Neste d'Oueil stream (Haute-Garonne, Pyrénées mountain range, France): Cires ($h = 1$), Mayrègne ($h = 2$), and Saint-Paul ($h = 3$). Stream sections are described elsewhere (Laplanche 2010). The stream has been electrofished in $r = 2$ removals. Only brown trout (*Salmo trutta fario*) are considered (Klemetsen et al. 2003). The number of caught fish per 10 mm length class are illustrated by a length–abundance plot (Fig. 4). The total number of fish caught are $C_1 = 246$, $C_2 = 509$, and $C_3 = 341$, respectively. We weighed only a fraction of the caught fish in July 2006. The weight dataset was enlarged by using fish caught in October 2006, leading to 133, 271, and 132 weight measurements at Cires, Mayrègne, and Saint-Paul, respectively.

Results

Length–weight relationship

Seven model reductions are considered, depending which constraints among $c_h = c$, $d_h = d$, and $s_h^2 = s^2$ are considered

Table 5. Weight submodel reductions and respective DIC.

Alternative	$c_h = c$	$d_h = d$	$s_h^2 = s^2$	DIC
1	–	–	–	2552
2	x	–	–	2551
3	–	x	–	2552
4	x	x	–	2619
5	–	–	x	2552
6	x	–	x	2551
7	–	x	x	2551
8	x	x	x	2617

Note: Alternative 1 is the model of eq. 7 with variability among stream sections of the parameters a_h, b_h, and s_h^2. Alternative 8 is the most reduced model, by assuming that the parameters c_h, d_h, and s_h^2 are equal for all sections; x, constrained; –, not constrained.

(Table 5). Burn-in (2500 iterations) and thinning (500) was investigated by using the most expanded model (alternative 1). The DIC estimates of the alternative weight submodels are also provided (Table 5). Alternative models can be classified into two groups: Models 4 and 8 of high DIC ∈ {2617, 2619} and remaining models (1–3, 5–7) of lower DIC ∈ {2551, 2552}. DIC results suggest that variability of c_h and (or) d_h should be considered, by rejecting models 4 and 8. Results also suggest that it is not relevant to include variability among sections of the residual variance in the weight submodel, in favor of models 5–7. Given that DIC results cannot discriminate among the remaining alternative models, we choose the most expanded model 5, i.e., model of eq. 7 with homoscedastistic residual error. Point (and interval) parameter estimates are $\widehat{c}_1 = 6.42$ (4.93, 8.08), $\widehat{c}_2 = 8.65$ (7.36, 10.19), $\widehat{c}_3 = 7.77$ (6.35, 9.34) µg/mm^3 and $\widehat{d}_1 = 3.088$ (3.041, 3.138), $\widehat{d}_2 = 3.030$ (2.997, 3.062), $\widehat{d}_3 = 3.063$ (3.025, 3.102) at Cires, Mayrègne, and Saint-Paul, respectively. The residual variance is $\widehat{s}^2 = 0.0047$ (0.0042, 0.0053).

Simulated dataset

Abundance is estimated by using the 4-removal, 3-removal, and 2-removal simulated dataset with $\Delta_l \in \{1, 2, 5, 10, 20\}$ mm (Table 4). It is not possible to use an index of complexity and fit to select the number of removals or the length class width, since the measured variables of models with different number of removals or class widths are distinct. Results show that, in the case of this dataset, abundance point estimates are not sensitive to the choice of the width of the length class, given that the width is small enough ($\Delta_l \leq 10$ mm). This result also applies to the remaining parameters as well as interval estimates (available online in Supplemental Appendix S1). Given that processing is faster (approximately 2, 5, 10, and 20× as fast) by using larger widths ($\Delta_l \in \{2, 5, 10, 20\}$ mm), following computations are carried out by using a $\Delta_l = 10$ mm class width. Results also show that the degradation of the information by using 3 removals is in this case negligible. Point and interval estimates of the parameters at the stream and section levels by using the 2-removal dataset with $\Delta_l = 10$ mm are provided (Burn-in: 50 000, thinning: 500) (Table 6). The estimated expected population sizes $\lambda_{h,i}A_h\Delta_l$ are illustrated (Fig. 3).

Field dataset

Fifteen abundance-biomass model alternatives are considered (Table 7). The first alternative (baseline model) is the abundance–biomass model that has been previously described with $q = 3$ Gaussian components ($\mu_1^{\lim} = 100$, $\mu_2^{\lim} = 150$, $\mu_3^{\lim} = 250$ mm). The second alternative (as well as the remaining alternatives) do not consider residual variability among fish of the catchability. In that case, $1/\eta = 0$ and eq. 3 simplifies into $p_{h,i,j} = p_{h,i,1}$, i.e., catchability is constant across removals. The third alternative describes the size structure as a mixture of triweight kernels, $K(l) = (35/32)(1 - l^2)^3$ for $l \in [-1, 1]$ and $K(l) = 0$ elsewhere. The fourth alternative describes the size structure as a mixture of $q = 4$ Gaussian components. In that case the boundaries of the mixture component centers are $\mu_1^{\lim} = 100$, $\mu_2^{\lim} = 150$, $\mu_3^{\lim} = 200$, and $\mu_4^{\lim} = 250$ mm. The remaining alternatives are model reductions that do not consider variability of the catchability with fish length (in that case $a_h = a = 0$ and eq. 2 becomes $\text{logit}(p_{h,i,1}) = b_h$) and (or) do not consider variability among stream sections of the parameters λ_h, $\tau_{h,k}$, $\mu_{h,k}$, $\sigma_{h,k}$, a_h, and b_h. Deviance, complexity, and DIC of model alternatives 2–15 are compared with respective values of the baseline model (Table 7). Burn-in (5000) and thinning (100) for alternative 2 were used for alternatives 3–15. Computations of the most expanded model (alternative 1) required larger burn-in (50 000) and thinning (500).

Results suggest that, in this case study, residual variability among fish of the catchability is not significant (by comparing alternatives 1, 2), the use of the Gaussian or triweight kernel are comparable (2, 3), and the use of 4 Gaussian components is ill-advised (2, 4). Results suggest that it is not relevant to consider variability among locations of a_h, b_h, $\tau_{h,k}$, $\sigma_{h,k}$, (2, 7, 9, 10). Results strongly suggest that variability with fish length of the catchability and variability among locations of λ_h and $\mu_{h,k}$ should be considered (2, 5, 6, 8). Remaining results (11, 12, 13, 14, 15) are in favor of model alternative 12, i.e., the abundance submodel described earlier with $q = 3$ Gaussian components without variability among locations of a_h, b_h, $\tau_{h,k}$, $\sigma_{h,k}$ nor residual variability of the catchability.

Point and interval estimates of the parameters at the stream and section levels are computed by using model alternative 12 (Table 8). The estimated abundances at Cires, Mayrègne, Saint-Paul (section level) and the Neste d'Oueil (stream level) are $\widehat{\lambda}_1 = 0.53$, $\widehat{\lambda}_2 = 0.98$, $\widehat{\lambda}_3 = 0.62$, and $\widehat{\lambda} = 0.83$ fish/m^2, respectively. The estimated biomasses are 11.50, 19.88, 14.76, and 15.47 g/m^2, respectively. Estimated expected population sizes are represented together with $C_{h,i,j}$ (Fig. 4).

Discussion

Point (and interval) estimates of the abundance of 0+ trout parrs at Cires, Mayrègne, and Saint-Paul are $\widehat{\lambda}_{1,1} = 0.24$ (0.21,0.28), $\widehat{\lambda}_{2,1} = 0.45$ (0.40,0.50), $\widehat{\lambda}_{3,1} = 0.29$ (0.25,0.32) fish/m^2. These results are similar to those found by Laplanche (2010): $\widehat{\lambda}_{1,1} = 0.18$ (0.15,0.24), $\widehat{\lambda}_{2,1} = 0.49$ (0.45,0.58), $\widehat{\lambda}_{3,1} = 0.29$ (0.26,0.36) fish/m^2. Laplanche (2010) presented a hierarchical model accounting for ran-

Table 6. True value and parameter estimates (stream and section levels) by using the simulated dataset with $r = 2$ removals and a $\Delta_l = 10$ mm class width.

Unit	Parameter	$q_{2.5\%}$	Truth	Estimate	$q_{97.5\%}$
—	$1/\eta$	0.00	0.10	0.11	0.66
—	π_a	0.0	10	122.5	1068.0
—	π_b	0.2	10	79.3	707.3
—	π_λ	4.5	50	350.4	1746.0
—	π_μ	1.9	10	16.6	60.2
—	π_σ	1.8	10	10.0	26.5
—	π_τ	0.5	10	4.2	13.7
g/m^2	biom	29.53	31.63	31.34	37.74
Fish/m^2	λ	0.80	1	1.02	1.46
—	τ_1	0.32	0.50	0.45	0.58
—	τ_2	0.16	0.30	0.27	0.40
—	τ_3	0.16	0.20	0.28	0.41
mm	μ_1	44.7	50	55.2	64.9
mm	μ_2	113.5	125	120.4	127.6
mm	μ_3	191.3	200	203.6	214.0
mm	σ_1	5.2	10	8.5	13.1
mm	σ_2	15.6	20	26.6	43.8
mm	σ_3	16.1	30	27.3	42.9
—	b	−1.25	0	−0.17	0.74
Per metre	$10^3 a$	2.8	5	6.7	10.8
g/m^2	biom$_1$	25.1	26.8	26.9	32.8
Fish/m^2	λ_1	0.90	0.98	1.04	1.47
—	$\tau_{1,1}$	0.43	0.44	0.48	0.55
—	$\tau_{1,2}$	0.27	0.32	0.32	0.37
—	$\tau_{1,3}$	0.15	0.18	0.20	0.24
mm	$\mu_{1,1}$	59.4	60.7	60.2	61.0
mm	$\mu_{1,2}$	116.4	119.4	120.0	123.6
mm	$\mu_{1,3}$	203.2	209.0	209.3	214.5
mm	$\sigma_{1,1}$	5.1	5.4	5.7	6.4
mm	$\sigma_{1,2}$	17.0	18.9	20.1	23.8
mm	$\sigma_{1,3}$	18.1	22.4	22.3	27.4
—	b_1	−1.06	0.01	−0.07	0.56
Per metre	$10^3 a_1$	2.5	4.9	6.4	10.2
g/m^2	biom$_2$	41.33	43.83	43.59	51.35
Fish/m^2	λ_2	0.89	0.97	1.02	1.43
—	$\tau_{2,1}$	0.33	0.43	0.40	0.47
—	$\tau_{2,2}$	0.15	0.21	0.22	0.31
—	$\tau_{2,3}$	0.30	0.39	0.39	0.46
mm	$\mu_{2,1}$	46.5	49.1	48.9	51.3
mm	$\mu_{2,2}$	113.7	126.8	121.1	129.0
mm	$\mu_{2,3}$	200.3	208.4	206.7	212.7
mm	$\sigma_{2,1}$	11.7	13.8	13.3	15.3
mm	$\sigma_{2,2}$	17.6	20.8	26.6	43.8
mm	$\sigma_{2,3}$	23.9	30.5	28.1	32.9
—	b_2	−1.02	0.44	0.01	0.70
Per metre	$10^3 a_2$	3.3	4.7	6.7	10.3
g/m^2	biom$_3$	21.2	24.3	23.5	30.1
Fish/m^2	λ_3	0.80	0.87	0.97	1.38
—	$\tau_{3,1}$	0.47	0.45	0.54	0.61
—	$\tau_{3,2}$	0.15	0.20	0.25	0.35
—	$\tau_{3,3}$	0.14	0.17	0.22	0.30
mm	$\mu_{3,1}$	55.8	56.9	56.9	58.2
mm	$\mu_{3,2}$	110.2	125.4	119.9	129.6
mm	$\mu_{3,3}$	182.5	189.9	194.9	205.9
mm	$\sigma_{3,1}$	6.5	7.1	7.6	8.7
mm	$\sigma_{3,2}$	20.0	42.0	33.3	48.2
mm	$\sigma_{3,3}$	24.3	30.4	30.4	38.4
—	b_3	−1.31	−0.11	−0.45	0.20
Per metre	$10^3 a_3$	3.4	4.8	7.1	12.0

Note: —, no unit.

Table 7. Abundance submodel alternatives, estimated deviances, and differences of complexity and DIC to the baseline model (alternative 1) by using the field dataset.

Alternative	Kernel	q	$1/\eta = 0$	$a_h = a = 0$	$\lambda_h = \lambda$	$\tau_{h,k} = \tau_k$	$\mu_{h,k} = \mu_k$	$\sigma_{h,k} = \sigma_k$	$a_h = a,$ $b_h = b$	$\bar{D}$	Δp	ΔDIC
1	Gaussian	3	–	–	–	–	–	–	–	343.3	0	0
2	Gaussian	3	x	–	–	–	–	–	–	321.9	–1	–22.4
3	Triweight	3	x	–	–	–	–	–	–	323.2	–1	–21.1
4	Gaussian	4	x	–	–	–	–	–	–	321.2	+8	–14.1
5	Gaussian	3	x	x	–	–	–	–	–	354.5	–4	+7.2
6	Gaussian	3	x	–	x	–	–	–	–	367.5	–3	+21.2
7	Gaussian	3	x	–	–	x	–	–	–	323.8	–5	–24.5
8	Gaussian	3	x	–	–	–	x	–	–	337.5	–7	–12.8
9	Gaussian	3	x	–	–	–	–	x	–	322.1	–7	–28.2
10	Gaussian	3	x	–	–	–	–	x	x	320.7	–5	–27.6
11	Gaussian	3	x	–	–	x	–	–	x	322.1	–9	–30.2
12	Gaussian	3	x	–	–	x	–	x	x	322.6	–15	–35.7
13	Gaussian	3	x	x	–	x	–	x	x	357.4	–16	–1.9
14	Gaussian	3	x	–	x	x	–	x	x	332.2	–17	–28.1
15	Gaussian	3	x	–	–	x	x	x	x	344.3	–21	–20.0

Note: The baseline model is the most expanded model presented in the text with $q = 3$ Gaussian components. x, constrained; –, not constrained.

Table 8. Parameter estimates at the stream level (Neste d'Oueil) and section level [Cires ($h = 1$), Mayrègne ($h = 2$), and Saint-Paul ($h = 3$)] by using model alternative 12 (Table 7).

Parameter	$q_{2.5\%}$	Estimate	$q_{97.5\%}$
$10^3 a$	3.9	7.7	11.5
b	0.17	0.59	1.03
π_λ	0.3	9.6	37.6
π_μ	1.8	22.3	76.3
σ_1	6.5	7.1	7.7
σ_2	12.2	14.1	16.5
σ_3	28.2	40.0	48.7
τ_1	0.43	0.46	0.49
τ_2	0.23	0.29	0.36
τ_3	0.18	0.25	0.32
biom	13.95	15.47	17.07
λ	0.31	0.83	1.50
μ_1	43.1	52.8	61.9
μ_2	115.4	120.4	125.4
μ_3	159.1	174.5	194.3
biom_1	9.68	11.50	13.58
λ_1	0.46	0.53	0.59
$\mu_{1,1}$	45.6	47.2	48.7
$\mu_{1,2}$	117.6	121	124.3
$\mu_{1,3}$	160.5	176.6	195.7
biom_2	17.22	19.88	22.71
λ_2	0.90	0.98	1.07
$\mu_{2,1}$	51.4	52.4	53.3
$\mu_{2,2}$	117.6	120.3	123.0
$\mu_{2,3}$	155.0	169.8	192.3
biom_3	12.69	14.76	17.31
λ_3	0.56	0.62	0.69
$\mu_{3,1}$	57.9	59.0	60.2
$\mu_{3,2}$	115.8	119.5	123.3
$\mu_{3,3}$	161.1	177.0	193.7

dom spatial variability of the abundance, random spatial variability of the catchability, and residual variability of the catchability with fish. By using an index of complexity of fit, Laplanche (2010) suggested that a reduction with no variability of the catchability (among stream sections and residual) should be preferred. We have attained similar conclusions ($a_h = a$, $b_h = b$, and $1/\eta = 0$). Results show, however, that heterogeneity of the catchability between individuals is significant, and would be explained by heterogeneity of the catchability due to fish length. Once heterogeneity of the catchability due to fish length is accounted for, we find no significant decrease of the catchability across removals due to (residual) between individual variability of the catchability. Such a result does not hold the possibility of different sources of alteration of the catchability across removals, e.g., owing to adjustments of the behavior of the fish or of the electrofisher to depletion (Mahon 1980; Schnute 1983; Peterson et al. 2004). The extension of the current model together with model selection could help in evaluating the value of the consideration of additional sources of variability of the catchability.

Model selection suggests that variability between stream sections of the abundances (λ_h), the component centers ($\mu_{h,k}$), and the allometric parameters (c_h and (or) d_h) are sig-

nificant. Biomass estimates are at Cires, Mayrègne, and Saint-Paul: 11.50 (9.68, 13.58), 19.88 (17.22, 22.71), and 14.76 (12.69, 17.31) g/m^2, respectively. Differences of total abundance (Bayley and Dowling 1993), growth (Ebersole et al. 2009), and as a result biomass (Bowlby and Roff 1986) are likely due to habitat variability. We believe that extending the current model by relating the above variables to key physico-chemical habitat covariates is an interesting perspective. This issue is discussed more thoroughly later by introducing several extensions of the current model.

The modeling of the length–abundance relationship as a mixture of Gaussian distributions may be inappropriate since the Gaussian probability density function is defined for all lengths. In this case study, expected population sizes for negative or large (>300 mm) length values are negligible. The use of a Gaussian kernel might be in that case satisfactory. We have also investigated the use of bounded, bell-shaped kernels. The comparison of model fit by using such a kernel (e.g., triweight) and the Gaussian kernel also suggests that the Gaussian kernel is acceptable. The use of a Gaussian kernel may, however, be problematic when mixing components of large variance. The use of different kernels has been suggested, for instance gamma kernels (Pitcher 2002). The current model can accustom any kernel given that it is possible to provide to WinBUGS a closed form of the probability density function of the kernel. We do not believe, however, that using a non-Gaussian kernel is the most promising solution to handle this issue. We rather suggest two different approaches. The first would be to extend the current model by constraining the component bandwiths to the centers, e.g., bandwiths proportionnal to centers. A second approach would be to separate components of large bandwiths into cohorts (of lesser bandwidths) by extending the current model with a growth submodel (which is discussed more thoroughly later).

The length–abundance plots of the catch at the Neste d'Oueil suggest that the use of three components would be satisfactory. Model comparison suggests that, indeed, a description of the abundance as a mixture of three components is fine. In this case study, the two first components unambiguously represent cohorts (0+ and 1+). The third component is not a cohort (brown trout in the Neste d'Oueil living longer than 3 years) but the mixture of several cohorts (2+ and older). Index of fit alone (deviance) suggests that the uses of three or four cohorts are comparable, index of complexity and fit (DIC) suggests that the use of three cohorts is preferable, prior knowledge suggests that the use of four or more cohorts is more realistic. One approach could be to set the number (q) and the definition sets of the centers ($\mu_c^{\lim}$) of the cohorts by using prior knowledge and not use an index of complexity and fit. Another approach could be (once again) to extend the current model with a growth submodel and let the model separate the cohorts.

The error resulting on biom$_h$ of approximating fish weights $W_{h,i,f}$ by expectations $c_h L_i^{d_h} \exp(s_h^2/2)$ is negligible (as demonstrated in Appendix A). This holds if the population sizes n_h are sufficiently large. With lower population sizes, the consequences of the above approximation on the biomass estimate should be considered more closely. For that purpose, we suggest two solutions. The first option is to include the fish level within the biomass submodel and simulate the fish weights $W_{h,i,f}$ instead of the expected fish weights $E_f(W_{h,i,f})$ to compute biom$_{h,i}$. As a result, the posterior distribution of the biomass variables (hence interval estimates) takes into account inter-individual variability of the weight. The second option is to use an a posteriori Monte Carlo run: (i) run the model to compute an estimate of $\lambda_{h,i}$, (ii) simulate $n_{h,i}$, (iii) compute biom$_{h,i}$ by using eq. 9, simulate the fish weights $W_{h,i,f}$, and compute biom$_{h,i}$ by using eq. 8, (iv) compute both alternatives of biom$_h$ by using eq. 10 and their difference Δbiom$_h$, (v) repeat steps (ii–iv), (vi) estimate the standard deviations of Δbiom$_h$ ($h \in \{1, \ldots, s\}$). By using the dataset collected in the field, we find that the standard deviations of the error on biom$_1$, biom$_2$, and biom$_3$ by approximating the fish weights $W_{h,i,f}$ by the expectations $c_h L_i^{d_h} \exp(s_h^2/2)$ are 0.09, 0.10, and 0.10 g/m^2, respectively.

We have assigned vague priors to all free parameters. Laplanche (2010) conducted an informal sensitivity analysis to evaluate the consequences of different selections of vague priors. The author showed, in his case study, that only estimates at the stream level are sensitive to the choices of vague priors. The reason is that the author considered (as we have done here) a hierarchical structure to model the variability between three stream sections. The amount of information to provide estimates of parameters at the stream level is limited. By extrapolating his results, in the case that only interpretation of the data is inquired with minimum influence of prior knowledge, we would advise to use the abundance submodel with vague priors and report parameter estimates at the section level. In the aim of providing estimates at the stream level with minimum influence of prior knowledge, a larger number of sampling sections would be required. In the prospect of using prior knowledge, we provide the (approximate) parametrized marginal posteriors of the free parameters of the baseline abundance submodel. Marginal posterior parameters that are reported are specific to this case study. We could use the marginal posteriors reported as priors to further study the Neste d'Oueil, by using data collected in July of another year, for instance. One could also adjust the parametrization of the posterior distributions reported with expert knowledge to provide priors to a different study area.

In the aim of using the approach presented by Marin et al. (2005) to model abundance as a distribution mixture, we first regarded fish length as a continuous variable. On the other hand, to allow for random localization of the fish within a stream section (eq. 4), to incorporate the removal sampling model (eq. 1), and to compute biomass, we needed to consider fish length as a discrete variable. By regrouping fish within $\Delta_l = 1$ mm length classes (equal to our measurement accuracy of the fish length), we could attain both objectives. In view of the computation time required to simulate the model with a $\Delta_l = 1$ mm length class, we adapted the model to consider larger class widths. Regrouping fish counts within larger length classes results, however, in a loss of information. Using smaller class widths increases the number of measured variables but the complexity of the parametrization of the distribution mixture ($3q$ per stream section) remains unchanged. In other words, using larger class widths does not make the distribution mixture model less complex and consequently more robust. Never-

theless, as reported, in the case of the simulated dataset, point and interval estimates of the parameters are not sensitive to the choice of the width of the length classes, given that the width is small enough. The reason is that when using datasets with large catches, the relative loss of information by grouping catches within larger size classes is negligible. With poorer datasets, the relative loss of information by grouping may not be negligible. Grouping catch within larger size classes also diminishes the capability of the model to detect variability of the catchability with length. By grouping, the amount of variation in catchability to be explained by length is reduced (since the amount of variation to be explained by length within a size class is null), consequently inflating the residual variability $1/\eta$. For such reasons, we recommend the use of a minimal class width when possible. With large datasets, grouping within larger size classes is conceivable.

We highlight some extensions to the current abundance-biomass model. One could relate parameters at the section level, either connected to catchability (η_h, a_h, b_h) or abundance (λ_h, $\tau_{h,k}$, $\mu_{h,k}$, $\sigma_{h,k}$), to environmental variables. With an appreciable number of sampling sections, parameters could be empirically related to habitat covariates (Rivot et al. 2008; Ebersole et al. 2009). Observations could be taken as independent, or parameters could be spatially related by using an auto-correlation model (Webster et al. 2008) completed by a Geographic Information System (Wyatt 2003). Higher hierarchical levels could be incorporated into the model to deal with a structured sampling strategy, like a larger spatial scale (e.g., watershed) or by integrating a time scale as done by Rivot et al. (2008). We believe that connecting the parameters of the distribution mixture through the development of a discrete population dynamics submodel and a growth submodel are interesting perspectives. The advantage of extending the abundance submodel with population dynamics and growth submodels is to derive variables (such as mortality, recruitment, and growth rates) more indicative of fish populations than component mixture parameters. The extensions suggested earlier (the use of environmental covariates, GIS, structured sampling) would in that case apply to the newly derived variables. As an illustration, we have extended the current abundance submodel with a growth submodel by relating growth rate and time of emergence to water temperature. An expanded hierarchical model would provide a unified framework where interval estimates of several variables indicative of fish populations (abundance, biomass, growth, mortality, recruitment, production, overweight, etc.) at a large spatial scale could be computed and used to study the biological health of freshwater sytems.

Acknowledgements

References

Anderson, C. 1995. Measuring and correcting for size selection in electrofishing mark-recapture experiments. Trans. Am. Fish. Soc. **124**(5): 663–676. doi:10.1577/1548-8659(1995)124<0663:MACFSS>2.3.CO;2.

Bayley, P., and Dowling, D. 1993. The effects of habitat in biasing fish abundance and species richness estimates when using various sampling method in streams. Pol. Arch. Hydrobiol. **40**(1): 5–14.

Bedrick, E. 1994. Maximum-likelihood-estimation for the removal method. Can. J. Stat. **22**(2): 285–293. doi:10.2307/3315590.

Bhattacharya, C.G. 1967. A simple method of resolution of a distribution into gaussian components. Biometrics, **23**(1): 115–135. doi:10.2307/2528285. PMID:6050463.

Bowlby, J., and Roff, J. 1986. Trout biomass and habitat relationships in southern Ontario streams. Trans. Am. Fish. Soc. **115**(4): 503–514. doi:10.1577/1548-8659(1986)115<503:TBAHRI>2.0.CO;2.

Buisson, L., Thuiller, W., Lek, S., Lim, P., and Grenouillet, G. 2008. Climate change hastens the turnover of stream fish assemblages. Glob. Change Biol. **14**(10): 2232–2248. doi:10.1111/j.1365-2486.2008.01657.x.

Carle, F., and Strub, M. 1978. A new method for estimating population size from removal data. Biometrics, **34**(4): 621–630. doi:10.2307/2530381.

Cone, R. 1989. The need to reconsider the use of condition indexes in fishery science. Trans. Am. Fish. Soc. **118**(5): 510–514. doi:10.1577/1548-8659(1989)118<0511:TNTRTU>2.3.CO;2.

Congdon, P. 2006. Bayesian statistical modelling. Wiley series in probability and statistics. 2nd ed. John Wiley and Sons, Ltd., Chichester, UK.

Cowx, I. 1983. Review of the methods for estimating fish population size from survey removal data. Fish. Manage. **14**(2): 67–82.

Crawley, M. 2007. The R book. John Wiley and Sons, Ltd., Chichester, UK.

Dorazio, R.M., Jelks, H.L., and Jordan, F. 2005. Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. Biometrics, **61**(4): 1093–1101. doi:10.1111/j.1541-0420.2005.00360.x. PMID:16401283.

Dorazio, R.M., Mukherjee, B., Zhang, L., Ghosh, M., Jelks, H.L., and Jordan, F. 2008. Modeling unobserved sources of heterogeneity in animal abundance using a Dirichlet process prior. Biometrics, **64**(2): 635–644. doi:10.1111/j.1541-0420.2007.00873.x. PMID:17680831.

Ebersole, J., Colvin, M., Wigington, P., Jr., Leibowitz, S., Baker, J., Church, M., Compton, J., and Cairns, M. 2009. Hierarchical modeling of late-summer weight and summer abundance of juvenile coho salmon across a stream network. Trans. Am. Fish. Soc. **138**(5): 1138–1156. doi:10.1577/T07-245.1.

Gelman, A., and Rubin, D. 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. **7**(4): 457–472. doi:10.1214/ss/1177011136.

Gouraud, V., Baglinière, J.L., Baran, P., Sabaton, C., Lim, P., and Ombredane, D. 2001. Factors regulating brown trout populations in two French rivers: application of a dynamic population model. Regul. Rivers Res. Manage. **17**(4–5): 557–569. doi:10.1002/rrr.655.

Heimbuch, D., Wilson, H., Weisberg, S., Vølstad, J., and Kazyak, P. 1997. Estimating fish abundance in stream surveys by using double-pass removal sampling. Trans. Am. Fish. Soc. **126**(5): 795–803. doi:10.1577/1548-8659(1997)126<0795:EFAISS>2.3.CO;2.

Jonsson, B., and Jonsson, N. 2009. A review of the likely effects of climate change on anadromous Atlantic salmon *Salmo salar* and

brown trout *Salmo trutta*, with particular reference to water temperature and flow. J. Fish Biol. **75**(10): 2381–2447. doi:10.1111/j.1095-8649.2009.02380.x. PMID:20738500.

Klemetsen, A., Amundsen, P.-A., Dempson, J., Jonsson, B., Jonsson, N., O'Connell, M., and Mortensen, E. 2003. Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L., and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecol. Freshwat. Fish, **12**(1): 1–59. doi:10.1034/j.1600-0633.2003.00010.x.

Kwak, J., and Waters, T. 1997. Trout production dynamics and water quality in Minnesota streams. Trans. Am. Fish. Soc. **126**(1): 35–48. doi:10.1577/1548-8659(1997)126<0035:TPDAWQ>2.3.CO;2.

Laplanche, C. 2010. A hierarchical model to estimate fish abundance in alpine streams by using removal sampling data from multiple locations. Biometrical J. **52**(2): 209–221.

Lobón-Cerviá, J. 1991. Dinámica de poblaciones de peces en ríos: pesca eléctrica y métodos de capturas sucesivas en la estima de abundancias. Monografías del Museo Nacional de Ciencias Naturales, Madrid, Spain.

Lobón-Cerviá, J. 2009. Recruitment as a driver of production dynamics in stream-resident brown trout (*Salmo trutta*). Freshw. Biol. **54**(8): 1692–1704. doi:10.1111/j.1365-2427.2009.02218.x.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. 2009. The BUGS project: Evolution, critique and future directions. Stat. Med. **28**(25): 3049–3067. doi:10.1002/sim.3680. PMID:19630097.

Mahon, R. 1980. Accuracy of catch-effort methods for estimating fish density and biomass in streams. Environ. Biol. Fishes, **5**(4): 343–363. doi:10.1007/BF00005188.

Mäntyniemi, S., Romakkaniemi, A., and Arjas, E. 2005. Bayesian removal estimation of a population size under unequal catchability. Can. J. Fish. Aquat. Sci. **62**(2): 291–300. doi:10.1139/f04-195.

Marin, J.-M., Mengersen, K., and Robert, C. 2005. Bayesian modelling and inference on mixtures of distributions, vol. 25 of Handbook of Statistics. Springer-Verlag, New York, N.Y.

Moran, P. 1951. A mathematical theory of animal trapping. Biometrika, **38**(3–4): 307–311.

Ntzoufras, I. 2009. Bayesian modeling using WinBUGS. Wiley series in computational statistics. John Wiley and Sons, Inc., Hoboken, N.J.

Pauly, D., and Moreau, J. 1997. Méthodes pour l'évaluation des ressources halieutiques. Cépaduès-Editions, Toulouse, France.

Penczak, T., Zalewski, M., Suszycka, E., and Molinski, M. 1981. Estimation of the density, biomass and growth rate of fish populations in two small lowland rivers.. Ekol. Pol. **29**(2): 233–255.

Peterson, J., Thurow, R., and Guzevich, J. 2004. An evaluation of multipass electrofishing for estimating the abundance of stream-dwelling salmonids. Trans. Am. Fish. Soc. **133**(2): 462–475. doi:10.1577/03-044.

Petts, G. 1984. Impounded rivers. Perspectives for ecological management. John Wiley and Sons, Ltd., Chichester, UK.

Petts, G., and Gurnell, A. 2005. Dams and geomorphology: Research progress and future directions. Geomorphology, **71**(1–2): 27–47. doi:10.1016/j.geomorph.2004.02.015.

Pinheiro, J., and Bates, D. 2000. Mixed-effects models in S and S-Plus. Springer-Verlag, New York, N.Y.

Pitcher, T. 2002. A bumpy old road: Sized-based methods in fisheries assessment, vol. 2 of Handbook of fish biology and fisheries. Blackwell Science Ltd., Oxford, UK. pp. 189–210.

Reiss, M. 1989. The allometry of growth and reproduction. Cambridge University Press, Cambridge, UK.

Riley, S., and Fausch, K. 1992. Underestimation of trout population size by maximum likelihood removal estimates in small streams. N. Am. J. Fish. Manage. **12**(4): 768–776. doi:10.1577/1548-8675(1992)012<0768:UOTPSB>2.3.CO;2.

Rivot, E., Prevost, E., Cuzol, A., Baglinière, J.-L., and Parent, E. 2008. Hierarchical Bayesian modelling with habitat and time covariates for estimating riverine fish population size by successive removal method. Can. J. Fish. Aquat. Sci. **65**(1): 117–133. doi:10.1139/F07-153.

Robert, C., and Casella, G. 2004. Monte Carlo statistical methods. Springer Texts in Statistics. 2nd ed. Springer-Verlag, New York, N.Y.

Schnute, J. 1983. A new approach to estimating populations by the removal method. Can. J. Fish. Aquat. Sci. **40**(12): 2153–2169. doi:10.1139/f83-250.

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc., B, **64**(4): 583–639. doi:10.1111/1467-9868.00353.

The World Commission on Dams. 2000. Dams and development. A new framework for decision-making. Earthscan Publications Ltd., London and Sterling, Va.

Wand, M., and Jones, M. 1990. Kernel smoothing. Monographs on statistics and applied probability.Vol. 60. Chapman and Hall, London, UK.

Webster, R.A., Pollock, K.H., Ghosh, S.K., and Hankin, D.G. 2008. Bayesian spatial modeling of data from unit-count surveys of fish in streams. Trans. Am. Fish. Soc. **137**(2): 438–453. doi:10.1577/T06-138.1.

Wyatt, R. 2002. Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model. Can. J. Fish. Aquat. Sci. **59**(4): 695–706. doi:10.1139/f02-041.

Wyatt, R. 2003. Mapping the abundance of riverine fish populations: integrating hierarchical Bayesian models with a geographic information system (GIS). Can. J. Fish. Aquat. Sci. **60**(8): 997–1006. doi:10.1139/f03-085.

Xenopoulos, M., Lodge, D., Alcamo, J., Märker, M., Schulze, K., and van Vuuren, D. 2005. Scenarios of freshwater fish extinctions from climate change and water withdrawal. Glob. Change Biol. **11**(10): 1557–1564. doi:10.1111/j.1365-2486.2005.001008.x.

Appendix A

The weight of a fish of length class i in section h is lognormally distributed of expectation $E_f(W_{h,i,f}) = c_h L_i^{d_h} \exp(s_h^2/2)$ and variance $\text{var}_f(W_{h,i,f}) = (c_h L_i^{d_h})^2 \exp(s_h^2)[\exp(s_h^2) - 1]$. By using central limit theorem, the variables $\text{biom}_{h,i}$ are approximately normally distributed of expectation $n_{h,i} E_f(W_{h,i,f})/A_h$ and variance $n_{h,i} \text{var}_f(W_{h,i,f})/A_h^2$. Since the $\text{biom}_{h,i}$ are independent, the biomass $\text{biom}_h = \sum_{i=1}^{m} \text{biom}_{h,i}$ of fish in section h is approximately normally distributed of expectation $\sum_{i=1}^{m} n_{h,i} E_f(W_{h,i,f})/A_h$ and variance $\sum_{i=1}^{m} n_{h,i} \text{var}_f(W_{h,i,f})/A_h^2$. With $n_h = \sum_{i=1}^{m} n_{h,i}$ sufficiently large, the coefficient of variation of biom_h is approximately null.