Nonlinear models for stock-recruitment analysis from HBM for Ecological Data, chapter 4

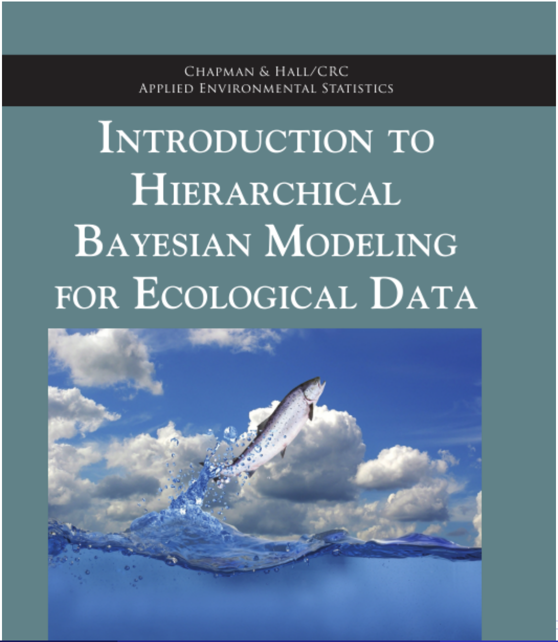
Éric Parent & Étienne Rivot
AgroParisTech, UMR518, Dpt of Maths, Paris, France
AgroCampus Ouest, UMR0985, Dpt of Ecology, Rennes, France
Eric.Parent@agroparistech.fr
Etienne.Rivot@agrocampus-ouest.fr

Monday, the 30th of June 2014

Main Ideas

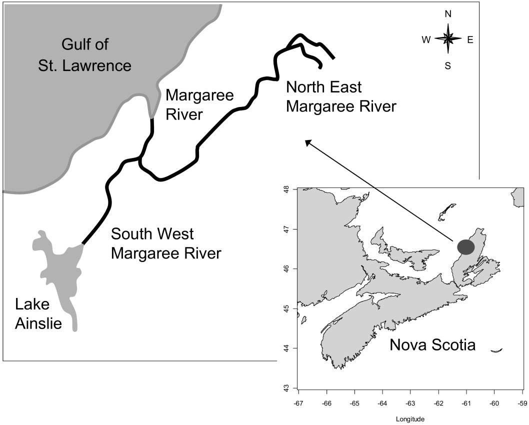
1. Stock-recruitment model for *A. Salmon* in the Margaree River (Nova Scotia, Canada) .
2. Quantify how unknowns, covariates and observations interact ? Hypotheses+Randomness.
3. Model=Crude simplification. Think Conditionally.
4. Which parameters the model should rely on ?
5. No *best* model. Visual inspection of model+ ecological knowledge+ good sense + quantitative criteria

This presentation is extracted from chapter 7



Margaree Motivating example

The Margaree recreational park is famous for Salmon fishing (fly fishing only). The Salmon fishing season begins June 1 and ends October 31. In 2007, nonresident anglers paid up to $133 for a seasonal license or $54 for a 7-day license. Fishing regulations such as bag limits are also enforced : 2 per day and 8 per season (only grilse up to 63 cm may be caught). The number of adult fish homing back to their native river is recorded. These fish will reproduce and die the same year in the Margaree River. Their eggs yield juveniles that will grow in the river and migrate down to the sea as smolts. After their marine journey (mostly one year), fish will ultimately swim back home to their natal river, reproduce and a new cycle will start.



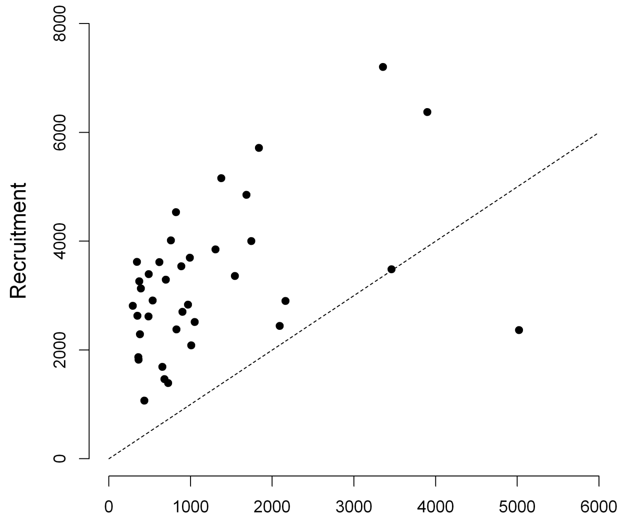
Data and Site

Fisheries scientists define the *stock* S_t as the number of spawners in year t, and the *recruitment* R_t as the number of spawners which are issued from the reproduction of the stock S_t. (measurements available for years 1947 to 1990)

Year Cohort	Stock Spawners	Recruitment Returns
1947	1685	4852
1948	3358	7204
1949	1839	5716
1950	1744	4000
1951	2093	2440
...		
1985	1378	5156
1986	3461	3484
1987	3899	6375
1988	1545	3358
1989	2164	2900

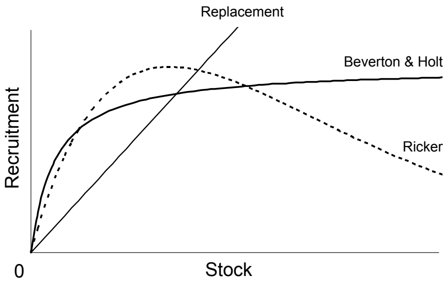
TABLE : Stock and recruitment data for the Margaree River from 1947 to 1990. (Data are reproduced by courtesy of the Department of Fisheries and Oceans from the Canadian Data Report of Fisheries and Aquatic Sciences No. 678.)

Data and Site



Searching for a *SR* model

Typical shape for Ricker and Beverton-Holt stock-recruitment relationships. If S and R are expressed in the same unit, then the SR relationship is directly comparable with the replacement line $S = R$.



Biologists will favor the conditional decomposition

$$[R, S] = [S] \times [R|S]$$

Assuming S is perfectly known (as a covariate), focus onto the modeling of $[R|S]$.

Searching for a R|S model

$$R = f(S, \epsilon)$$

Parsimony means Normal or LogNormal + parametrics

$$[R|S] = dnorm(R, f(S), \sigma(S))$$

Either additively

$$f(S, \epsilon) = f_1(S) + f_2(\varepsilon)$$

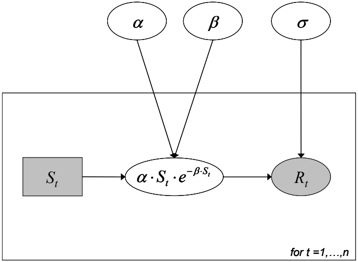
Either a multiplicative effect :

$$f(S, \epsilon) = f_1(S) \times f_2(\varepsilon)$$

Two density dependent widely used models for $f_1(S)$:

- Beverton-Holt taking into account a saturation effect. The replacement rate $\frac{R}{S}$ is slowly decreasing as the stock increases : $R = \alpha S/(1 + \beta S)$
- Ricker form with an exponentially decreasing replacement rate : $R = \alpha S e^{-\beta S}$

Directed Acyclic Graph for the Ricker model with natural parameters $\theta = (\alpha, \beta, \sigma)$



$$
\begin{cases}
R_t = f(\alpha, \beta, S_t) e^{\epsilon_t} \\
f(\alpha, \beta, S_t) = \alpha \cdot S_t e^{-\beta \cdot S_t} \\
\epsilon_t \overset{iid}{\sim} Normal(0, \sigma^2)
\end{cases}
$$

A Hidden regression

Noninformative priors for $log(\alpha)$,β, and the precision σ^{-2}

$$\begin{cases} log(\alpha) \sim Uniform(-10, 10) \\ \beta \sim Uniform(-10, 10) \\ \sigma^{-2} \sim Gamma(1, 0.25) \end{cases}$$

Bayesian analysis through a WinBUGS code

Parameters	Mean	Sd	2.5% pct	97.5% pct
α	6.07	0.64	4.91	7.42
β	0.00049	0.00007	0.00036	0.00062
R_{max}	4600	433	3874	5562
S_{max}	2084	298	1613	2777
σ	0.43	0.05	0.34	0.54

α = slope close to the origin $S = 0$, when the relation between S and R is nearly linear.
β^{-1} is the stock that produces the maximum recruitment (an indicator of the carrying capacity S_{max})
$R_{max} = f(S_{max}) = \frac{\alpha}{\beta} e^{-1}$.

A Non Informative Regression results

These results show that :

- Good replacement ratio. S small, 1 adult will be replaced by 6 juveniles. One can bet the ratio stands between 5 and 8 with high confidence ;
- Carrying capacity $S_{\max}$ rather uncertain (95% posterior credible interval between 1628 and 2760), posterior mean = $2,100$ adults ;
- Environmental stochasticity quite high. $\sigma = 0.4$ for the log-value roughly equivalent to a 40% standard deviation for the relative error between the quantity of interest R and its phenomenological prediction !

Management parameters

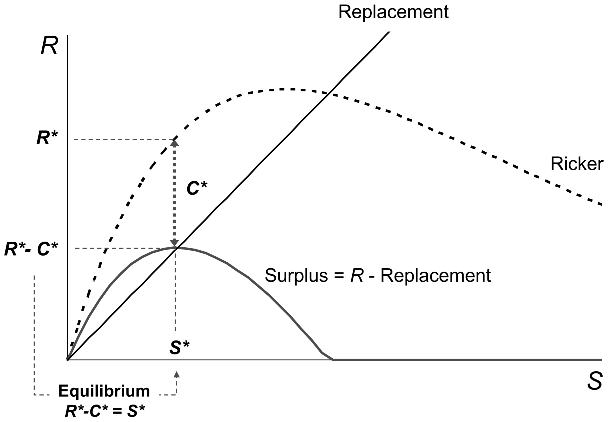


FIGURE : Management-related parameters for the Ricker stock-recruitment relationship (S and R expressed in the same unit).

Management parameters

Ricker and Beverton-Holt functions rewritten in terms of management-related parameters S^*, producing the maximum sustainable yield, C^*. Hypothesizing a sustainable population (*i.e.*, $\alpha > 1$), closed transformation (α, β) to reference points (S^*, h^*) or (S^*, C^*), where $h^* = \frac{C^*}{C^*+S^*}$ is the harvest rate at equilibrium.
Ricker model :

$$\begin{cases} \alpha = \dfrac{(S^* + C^*)}{S^*} \cdot e^{\frac{C^*}{S^*+C^*}} \\ \beta = \dfrac{C^*}{S^*(S^* + C^*)} \end{cases}$$

or equivalently

$$\begin{cases} \alpha = \dfrac{1}{1 - h^*} \cdot e^{h^*} \\ \beta = \dfrac{h^*}{S^*} \end{cases}$$

$((S^*, C^*))$ are solutions of the system :

$$\begin{cases} \textit{Equilibrium conditions}: \; R - C = S \\ \textit{Maximization of catches}: \; \dfrac{\partial(R - S)}{\partial S}|_{S=S^*} = 0 \end{cases}$$

Informative prior

Assume the biologist *a priori* (*i.e.*, without seeing the data) said :

- best guess for the Margaree River stock at maximum sustainable yield $S^* = 1000$ individuals,
- S^* lies between 700 and 1300 as a 70% credible set,
- the optimal sustainable exploitation rate h^* is likely to lie around 0.75.

Tentatively model prior knowledge by :

$$\begin{cases} h^* \sim Beta(3,1) \\ S^* \sim Normal(1000, 300^2) \end{cases}$$

Conjugate *Gamma*(p, q) for the precision σ^{-2} . Relative possible variation not far from 50% of the signal as a prior bet $p = 1; q = 0.25$.

Posterior pdf from informative prior on the management parameters

Param.	Mean	Std	2.5%	97.5%
C^*	2862	288	2330	3459
R^*	4209	351	3571	4948
α	6.2	0.62	5.08	7.50
S_{max}	1989	229	1596	2488
S^*	1347	124	1132	1621
h^*	0.68	0.02	0.63	0.72
σ	0.43	0.05	0.34	0.54

TABLE : Main features of marginal posterior distributions obtained with a Ricker-type recruitment function, informative priors on management parameters and logNormal random variations.

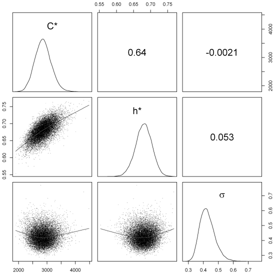
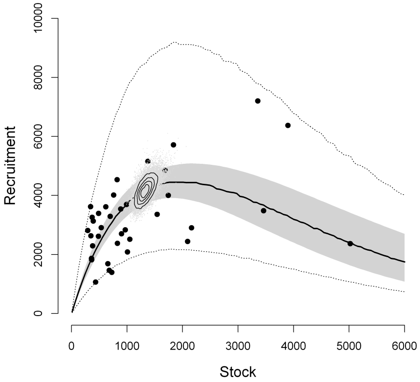


FIGURE : Posterior distributions of key parameters (C^*, h^*, σ) obtained using the Ricker-type SR model with logNormal random variations. Marginal distributions are shown in the diagonal. Joint MCMC draws are shown in the lower part. The upper part shows linear correlations between the MCMC draws.

Ricker model fitted with logNormal noise and informative prior on the management parameters



Compared to non informative prior, S_{max} tends to be smaller and the environmental noise level σ remains in the same range. The gray zone shows the 90% credible interval for the model (uncertainty around parameters (S^*, h^*) only), the upper and lower lines (dotted) gives a 90% posterior predictive interval for the data (including the environmental noise σ.) The joint posterior distribution of management parameters appears as a cloud showing the uncertainty about (S^*, R^*).
There is an unexplained hump of S in the neighborhood of S^* and no data at all is laying out of the 95% predictive range (although it should concern approximately 5% of the sample).

Changing the error term from logNormal to Gamma

Explore the alternative hypothesis $\sigma^2(R) = \delta S$. Gamma distribution with parameters a and b

$$\begin{cases} \mu = \dfrac{a}{b} \\ \sigma^2 = \dfrac{a}{b^2} = \dfrac{\mu}{b} \end{cases}$$

$$\begin{cases} \mu(R) = \dfrac{a(S)}{b(S)} = f(S) = \alpha \cdot S \cdot e^{-\beta \cdot S} \\ \sigma^2(R) = \dfrac{\mu(R)^2}{a(S)} = \delta \cdot S \end{cases}$$

so that :

$$\begin{cases} R \sim Gamma(a(S), b(S)) \\ with \\ a(S) = \dfrac{\alpha^2 \cdot S \cdot e^{-2\beta \cdot S}}{\delta} \\ b(S) = \dfrac{\alpha \cdot e^{-\beta \cdot S}}{\delta} \end{cases}$$

Changing the error term from logNormal to Gamma

Parameters	Mean	Sd	2.5%	97.5%
$R^* - S^*$	2972	301	2426	3623
R^*	4295	411	3570	5151
α	6.54	0.66	5.33	7.93
S_{max}	1916	285	1408	2525
S^*	1323	164	1018	1660
h^*	0.69	0.02	0.64	0.74
σ	0.86	0.12	0.64	1.09

TABLE : Main features of marginal posterior distributions obtained with a Ricker-type recruitment function, informative priors on management parameters and a Gamma random variations assuming the variance is proportional to the stock.

Changing the error term from logNormal to Gamma

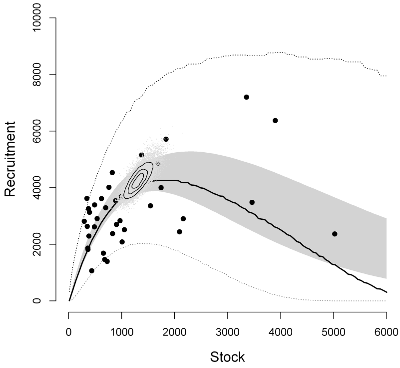


FIGURE : Ricker model fitted with Gamma random variations. The gray zone shows the 90% credible interval for the model (uncertainty around parameters (S^*, h^*) only), the upper and lower lines (dotted) give a 90% posterior predictive interval for the data (including the environmental noise σ.) The joint posterior distribution of management parameters appears as a cloud showing the uncertainty about (S^*, R^*).

- No major discrepancies between expert judgment and information conveyed by the data
- A much better fit of environmental noise
- The hump around S^* has vanished and the environmental noise increases with the number of spawners
- Highly skewed

Changing the deterministic explanation term from Ricker to Beverton-Holt

$$\begin{cases} R = \dfrac{\alpha}{1+\beta\cdot S}\cdot e^{\varepsilon} \\ \varepsilon \sim Normal(0,\sigma^2) \end{cases}$$

Management parameters :

$$\begin{cases} \alpha = \dfrac{R^{*2}}{S^{*2}} \\ \beta = \dfrac{R^{*}-S^{*}}{S^{*2}} \end{cases}$$

Parameters	Mean	Sd	2.5% pct	97.5% pct
$C^* = R^* - S^*$	2294	206	1918	2719
R^*	3191	319	2622	3883
α	13.75	5.02	8.11	26
R_{max}	4453	596	3416	5772
S^*	897	185	559	1292
h^*	0.72	0.04	0.65	0.80
σ	0.39	.05	0.31	0.49

TABLE : Main features of marginal posterior distributions obtained with a Beverton-Holt recruitment function, informative priors on management parameters and logNormal random variations.

Beverton-Holt with LogNormal noise

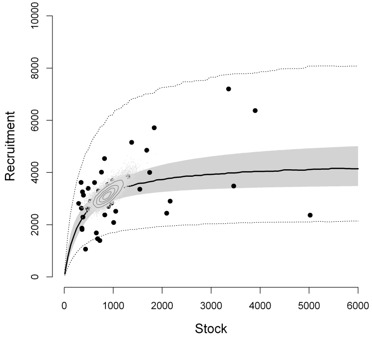


FIGURE : Beverton-Holt model fitted with logNormal random variations. The gray zone shows the 90% credible interval for the model, the upper and lower lines gives a 90% posterior predictive interval for the data. The posterior of management parameters appears as a cloud showing the uncertainty about (S^*, R^*).

Beverton-Holt model differs much from the Ricker structure

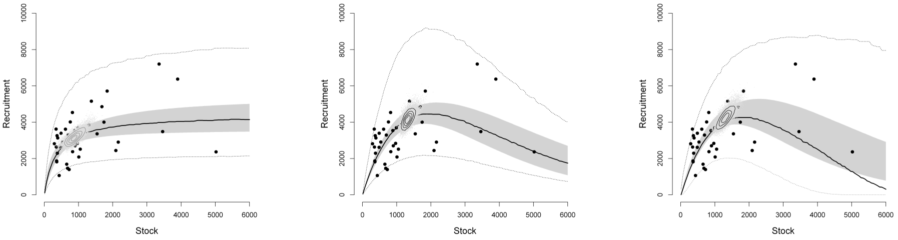
- asymptotic number of recruits around 4000 individuals (with a rather large posterior uncertainty)
- the predictive confidence interval increases with S (and R)
- the 95% predictive quantile might be far from the data and rather overpessimistic.

Model choice with informative prior

Three proposals :

- Beverton-Holt-type with logNormal random variations (M_1),
- Ricker-type and logNormal random variations (M_2),
- Ricker-type and Gamma random variations (M_3).

$$B_{ij} = \frac{[M_i|\mathbf{y}]}{[M_j|\mathbf{y}]}\frac{[M_j]}{[M_i]} = \frac{\int [\theta|M_i][\mathbf{y}|\theta, M_i]d\theta}{\int [\theta|M_j][\mathbf{y}|\theta, M_j]d\theta}$$



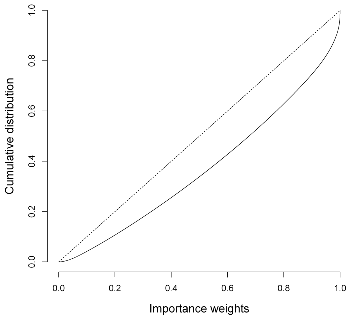
Importance sampling to compute Bayes factors

Importance sampling :

1. Approximate $[\theta|M_i, \mathbf{y}]$ by the multi-Normal distribution $\pi(\theta)$;
2. Let the importance distribution π appears by rewriting $[\mathbf{y}|M_i] = \int_\theta \left(\frac{[\theta|M_i][\mathbf{y}|\theta,M_i]}{\pi(\theta)}\right) \pi(\theta)d\theta$;
3. Generate a $G-$ sample drawn from π, $(\theta^{(g)})_{g=1,\ldots,G}$;
4. Compute the weighted sum $\widehat{[\mathbf{y}|M_i]} = \sum_{g=1}^{G} \omega(\theta^{(g)})[\mathbf{y}|\theta^{(g)}, M_i]d\theta$

$$\omega(\theta^{(g)}) = \frac{\frac{[\theta^{(g)}|M_i]}{\pi(\theta^{(g)})}}{\frac{1}{G}\sum_{g=1}^{G} \frac{[\theta^{(g)}|M_i]}{\pi(\theta^{(g)})}}$$

Model	Rank	$Log(\widehat{[\mathbf{y}\vert M_i]})$	BF_{1vsi}
Beverton-Holt + LogNormal noise	**1**	-335.75	1
Ricker + LogNormal noise	2	-340.23	88
Ricker + Gamma noise	3	-340.56	123



Checking the IS computation for the Ricker model with logNormal noise. Cumulative distribution of the importance weights (solid line) compared with cumulative distribution of a Uniform distribution (dotted line)

Conclusions and perspectives : ecological models and environmental stochasticity.

Noise is a convenient all-in-one-bag concept for unexplained variations.

1. Stochasticity in survival of eggs that may vary due to environmental factors (temperature, low riverflows, floods..) ;
2. Experimental errors

Beware :

- The *SR* Ricker model with logNormal noise is a linear regression model in disguise !
- Many ecological time series data sets are shorter than the respectable 38 years of the Margaree data collection !

Important sources of bias put under the carpet :

1. *The errors in variable problem*. Stock values, although explanatory variables, are not known without error.
2. *The time series bias*. The ecological dynamics has been cut off.

References stock-recruitment & Salmon

G. Chaput, J. Allard, F. Caron, J.B. Dempson, C.C. Mullins, and M.F. O'Connell.
River-specific target spawning requirements for Atlantic salmon (salmo salar) based on a generalized smolt production model.
Canadian Journal of Fisheries and Aquatic Sciences, 55 :246–261, 1998.

R. Hilborn and C.J. Walters.
Quantitative Fisheries Stocks Assessment : Choice, Dynamics & Uncertainty.
Col. Natural Resources. Chapman & Hall, New York, 1992.

R.E. Kass and A.E. Raftery.
Bayes factors.
Journal of the American Statistical Association, 90(430) :773–795, 1995.

S.B. Munch, A. Kottas, and M. Mangel.
Bayesian nonparametric analysis of stock-recruitment relationships.
Canadian Journal of Fisheries and Aquatic Sciences, 62 :1808–1821, 2005.

C.L. Needle.
Recruitment models : diagnosis and prognosis.
Reviews in Fish Biology and Fisheries, 11 :95–111, 2002.

E.C.E. Potter, J.C. MacLean, R.J. Wyatt, and R.N.B. Campbell.
Managing the exploitation of migratory salmonids.
Fisheries Research, 62 :127–142, 2003.

E. Prévost, G. Chaput(Eds INRA Editions, Paris, 2001).
Stock, Recruitment and Reference Points Assessment and Management of Atlantic salmon.

E. Prévost, E. Parent, W. Crozier, I. Davidson, J. Dumas, G. Gudbergsson, K. Hindar, P. McGinnity, J. MacLean, and L.M. Sattem.
Setting biological reference points for Atlantic salmon stocks : Transfer of information from data-rich to sparse-data situations by Bayesian hierarchical modelling.
ICES Journal of Marine Science, 60 :1177–1193, 2003.

J.J. Quinn, I and R.B. Deriso.
Quantitative Fish Dynamics, chapter 3 : Stock and Recruitment.
Biological resource management. Oxford University Press, Oxford, New York, 1999.

E. Rivot, E. Prévost, and E. Parent.
How robust are Bayesian posterior inferences based on a Ricker model with regards to measurement errors and prior assumptions about parameters ?
Canadian Journal of Fisheries and Aquatic Sciences, 58 :2284–2297, 2001.

J.T. Schnute, A. Cass, and L. Richards.
A Bayesian decision analysis to set escapment goals for Frazer river Sockeye salmon (onchorhynchus nerka) fishery.
Canadian Journal of Fisheries and Aquatic Sciences, 57 :962–979, 2000.

J.T. Schnute and A.R. Kronlund.
A management oriented approach to stock recruitment analysis.
Canadian Journal of Fisheries and Aquatic Sciences, 53 :1281–1293, 1996.

J.T. Schnute and A.R. Kronlund.
Estimating salmon stock-recruitement relationships from catch and escapement data.
Canadian Journal of Fisheries and Aquatic Sciences, 59 :433–449, 2002.