# Prediction of biomass in Norwegian fish farms

Anders Løland[a,*], Magne Aldrin[a], Gunnhildur Högnadóttir Steinbakk[a], Ragnar Bang Huseby[a], Jon Arne Grøttum[b]

[a] *Norwegian Computing Center, Post Office Box 114 Blindern, NO-0314 Oslo, Norway*
[b] *Norwegian Seafood Federation, Post Office Box 1214 Pirsenteret, NO-7462 Trondheim, Norway*

[*] Corresponding author. Tel.: +47 22852500; fax: +47 22697660.
*Email address:* anders.loland@nr.no (Anders Løland)

*Preprint submitted to Can. J. Fish. Aquat. Sci.*

**Abstract**

We have constructed a statistical model to forecast, with uncertainty, the stock of Norwegian farmed Atlantic salmon (Salmo salar). The model provided good predictions of future biomass of Norwegian farmed salmon and can also be used to perform what-if analysis exploring the impact of varying scenarios for stocking and slaughtering. The model is based on the number of fish in each weight class (0-1 kg, 1-2 kg, ..., 10+ kg) and their average weight. The model, which is related to standard size-structured models, computes the number of fish growing into the next weight class the next month and the number of fish remaining in the same weight class. In addition, the number of new fish stocked, fish lost, slaughtered and wasted, as well as the sea temperature related to the growth, were modelled. All the model parameters were estimated based on monthly data from 2002–2007, and the model was validated statistically. Any animal production involving cycles may benefit from this forecasting tool.

*Key words:* farmed salmon, growth model, population dynamics, stochastic model, validation

## Introduction

Thanks to its extraordinary geographical characteristics, Norway became the first country to actively promote the development of salmon farming. Today, Norwegian interests play an important role in global salmon farming. The on-growing production is performed in sea based cages. Cage culture production of Atlantic salmon (Salmo salar) in the sea has expanded and intensified considerably over the years and in 2009 amounted to 865 000 tonnes. The production is run in about 700 sites along the coast, owned by approximately 160 different commercial companies of different sizes.

As one of the major producers of Atlantic salmon, the Norwegian fish farms industry meets variation in the global market demand for salmon, as well as increased international competition and price volatility. To minimise the fluctuations, it is important for the producers to have knowledge about the expected development of future production and demand. The demand can be influenced by timing or scaling of marketing campaigns. If information about expected production was available, the individual producers could adapt their production plans, for instance by changing the production regime, whereas the fish farming industry on a national level could develop improved production strategies, by changing the time for stocking or slaughtering.

The cycle for producing salmon is rather predetermined. After new (young) fish have been stocked, the fish are usually slaughtered after 1 1/2 to 3 years. The production cycle involves systematic fluctuations in biomass and quantity of slaughtered fish. When the new fish have been stocked, the biomass of slaughtered fish is to some degree predictable at a horizon of up to 2 years, since the growth of farmed fish to a large extent is systematic.

The production cyclicity is common amongst other farmed animals, such as cattle or hog production. This cyclicity and its relationship with the market have been studied for a long time (Russell, 1929) and are still debated (Crespi et al., 2010). Recent studies suggest that market timing influences the cattle cycle (Hamilton and Kastens, 2000). As (Allen, 1994, p. 81) points out, "there has been an overemphasis on explanation, and little interest in the predictive power of models" for agricultural production.

The market for farmed salmon is growing, but less mature than many other agricultural markets, and has experienced imperfections due to trade conflicts (Kinnucan and Myrland, 2005) and outbreaks of fish diseases. In any case, the increased production of farmed salmon has reduced the wild salmon prices (Guillotreau, 2004; Asche et al., 2005).

Until now, no reliable tool for forecasting future production on an aggregated level has been available. This is probably due to the lack of coherent data. Our aim is to present a prediction model for regional and national standing biomass. Such a model can also can be used to investigate consequences of changing production strategies.

Since 2002, production data from all Norwegian fish farms have been electronically collected. Monthly data, including the existing stock, stocking of new fish and slaughtering, are available on a regional level. The data are given as numbers and mean weight of fish in the standing stock within weight classes of 1 kg resolution (0-1 kg, 1-2 kg, ..., 9-10 kg, 10+ kg). Accurate age information is lacking.

Since the data are divided into weight classes, it is natural to model them by size-structured models (Hilborn and Mangel, 1997; Tuljapurkar and Caswell, 1997; Quinn and Deriso, 1999). Rizzo and Spagnolo (1996) describe optimal management in a single-farm model for sea bass, based on weight classes and biological sub-models. Results based on different sea temperatures, harvesting, feeding and stocking strategies are presented. Gangnery et al. (2001) investigate the production of oysters, where the oysters grow individually between weight classes, and mortality, harvesting and seeding are included. Halachmi et al. (2005) have studied how to optimise the management of a single (or a few) fish farm(s), and Bjørndal (1988) investigates optimal harvesting of farmed fish based on an adapted Beverton-Holt model for a year class of fish. In work related to farmed salmon, Kumbhakar and Tveterås (2003) have studied the risk preferences of Norwegian salmon farmers, while Bjørndal (1990) describes salmon aquaculture production and harvesting from an economical viewpoint.

Many of these models (Rizzo and Spagnolo, 1996; Gangnery et al., 2001) are stochastic on the individual level, but deterministic on the population level. Their parameters are either found from the literature, estimated on different data sets or determined more or less subjectively, as is also the case in Bjørndal (1988). There can be good reasons for this approach. For example, the data can be incomplete, or parameters like the natural mortality can be non-identifiable. Simulations from the models are presented, but the results are not compared with or validated against the actual data in a thorough analysis. Moreover, the uncertainty in parameters and predictions of the population is more or less ignored.

We propose a dynamic, statistical model that is related to, but not equal to, standard size-structured models. One difference is that the proposed model takes into account also the average weight in each size class in addition to the number of fish. Furthermore, it is constructed such that all unknown parameters can be estimated from one coherent dataset, and prediction uncertainty is an inherent part of the model.

The starting point of the model is the abundance and the average weight of the fish in each weight class. Based on this and amongst others a growth function, depending on weight class, sea temperature, seasonality and light conditions, the model computes per month how many fish that remain in their weight class and how many that grow into the next weight class the following month. In addition, we modelled the number of new fish stocked, lost (dead and escaped), slaughtered and wasted (downgraded slaughter fish), learning from the production strategies of earlier years. We demonstrate and justify the predictive ability of the model and describe how what-if scenarios can be made.

## Materials and methods

### Data

Production data from Norwegian fish farms have been collected into the database Havbruksdata since 2002. The data are monthly observations of abundance and standing biomass in each of 11 weight classes; $v = 0$: 0-1 kg, $v = 1$: 1-2 kg, ..., $v = 9$: 9-10 kg, $v = 10$: 10+ kg. When both the numbers and biomass are observed, the average weight in each weight class is implicitly given as the biomass divided by the number of fish. In addition, the monthly number of fish stocked (or inserted), lost (died and escaped), slaughtered and wasted (downgraded slaughter fish) as well as the gutted slaughtered biomass are known in each weight classes. All these data are aggregated into three regions; i) Southern, ii) Mid and iii) Northern Norway. The data are modelled on a regional basis, but results are aggregated further for Norway as a whole, and all data and results presented here are for Norway unless otherwise stated.

An overview of the various types of data is given (Table 1). For slaughtered fish, we distinguish between the full weight, simply denoted slaughter weight ($V_t^S$) and gutted weight ($V_t^G$). The slaughter weight, which is unobserved, is comparable to the weight of the standing stock, whereas the observable gutted weight is lower. The monthly distribution of the data (for Norway as a whole) over a year gives an impression of the seasonal variations

(Table 2). Note that the fish are usually stocked around May, and some in September and October, while the number of lost fish peaks during the summer. For an overview of the weight class distribution, a summary in absolute numbers per weight class is shown (Table 3). The average number of fish stocked each month is 19.4 million. The corresponding number of fish in weight class 0 is 93.2 million, which is far below 19.4 million times 12 months, since the fish grow into weight class 1 after a few months and some fish die each month (1.7% of the stock in weight class 0 according to the table). Time plots of some of the data described in the Results section complement the picture.

The number of daylight hours $D_t$ in month $t$ partly explains the growth in that month (Boeuf and Le Bail, 1999). We used equations from the National Oceanic and Atmospheric Administration, based on work by Meeus (1991), to compute the number of daylight hours for a mid point in the region for the 15th each month. In, e.g., Finnmark, in Northern Norway, there are no daylight hours in December and 24 daylight hours (midnight sun) in June.

The growth does also depend on the monthly sea temperature $ST_t$, averaged over each reporting fish farm in each region, and collected from the database Havbruksdata.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

### Method overview

The model presented below is first applied to each of the three regions separately. Results for Norway as a whole are then computed by aggregating the results over the regions.

The model contains five sub-models for monthly values of: 1) Standing stock distributed among weight classes, 2) stocked number of fish, 3) loss, 4) slaughter and waste, 5) sea temperature. The standing stock model is presented in detail below. The other models are outlined in a subsequent section, except for the model for sea temperature, which is described in Appendix A. Details for all the five sub-models are provided (Appendix A, C and D), as well as an overview (Figure 1).

[Figure 1 about here.]

Generally, the choice of models and the inclusion of linear, non-linear and seasonal effects were based on the available literature, analyses of the data and the resulting model's predictive performance.

### Model for standing stock

First, we go through the model for the standing stock. The average weight and the number of fish in each weight class are the key ingredients. For a given growth function, and based on the principle of balance of numbers and biomass, we compute how many fish that remain in their weight class and how many that grow into the next from one month to another. If the average weight of the fish in a weight class is close to the upper boundary of that weight class, many fish will grow into the next weight class. Similar computations are done for the weight in each weight class.

Since individual fish are not observed, we assume that all fish in a weight class $v$ has the same growth factor $f_{t,v}$ from time (month) $t$ to $t+1$, and that negative growth is impossible, i.e. $f_{t,v} \geq 0$. The growth is a function of i) weight class, ii) sea temperature, iii) the number of daylight hours and iv) season. The seasonal component accounts for additional seasonal variations that are not captured by sea temperature and daylight hours. The growth model is then

$$f_{t,v} = 1 + \exp(\eta_{t,v}),$$

where the linear predictor $\eta_{t,v}$ is a quadratic function of the sea temperature $ST_t$ and the number of daylight hours $D_t$. The quadratic terms were included to allow for non-linear effects, since the salmon does not grow linearly with $ST_t$ and $D_t$.

The seasonal component is handled by a pair of sine and cosine functions with a period of one year. This gives

$$\begin{aligned} \eta_{t,v} = \beta^f_{0,v} & \\ &+ \beta^f_1 ST_t + \beta^f_2 ST_t^2 \\ &+ \beta^f_3 D_t + \beta^f_4 D_t^2 \\ &+ \beta^f_5 \sin\left(\frac{2\pi t}{12}\right) + \beta^f_6 \cos\left(\frac{2\pi t}{12}\right). \end{aligned} \tag{1}$$

Here, the coefficient $\beta^f_{0,v}$ is an intercept specific for each weight class $v$, whereas the other $\beta$s are regression coefficients common for all weight classes. To model seasonal variations, we also use the sine and cosine functions throughout the paper.

The regression coefficients are unknown, but estimated from historical data by minimising the prediction errors of monthly numbers and average weight in each weight class (see the next section). They are estimated simultaneously with other unknown parameters introduced later in this section. This is the model for $v \geq 1$. For weight class $v = 0$, the growth is in addition a function of the average weight $V_t(0)$ in weight class $v = 0$ at time $t$ (not shown).

Since all fish in weight class $v$ have the same growth, all fish with a weight above a threshold $v'_{t,v}$ at time $t$ will enter the next weight class (see Figure 2) at time $t+1$. The threshold $v'_{t,v}$ is given by $f_{t,v} \cdot v'_{t,v} = v_{\max}$, where $v_{\max} = v+1$ is the upper boundary of the weight class. This gives

$$v'_{t,v} = \frac{v_{\max}}{f_{t,v}}.$$

The number of fish remaining in weight class $N_{t+1}(v, v)$ and the number of fish switching in weight class $N_{t+1}(v+1, v)$ are given by

$$\begin{aligned} N_{t+1}(v, v) &= a_{t,v} \cdot N_t(v) \\ N_{t+1}(v+1, v) &= (1 - a_{t,v}) \cdot N_t(v). \end{aligned} \tag{2}$$

Here, $a_{t,v}$ is the proportion of fish remaining in the weight class $v$, and $N_t(v)$ is the number of fish in weight class $v$ in month $t$.

Assume that the weight of the individual fish in weight class $v$ is distributed according to a probability density $h_{t,v}(w)$ between the boundaries $v$ and $v+1$. The proportion $a_{t,v}$ is then given by

$$a_{t,v} = \int_{v}^{v'_{t,v}} h_{t,v}(w)dw. \tag{3}$$

The true distribution $h_{t,v}(w)$ is unknown, but the average weight $V_t(v)$ of the fish in each weight class $v$ is observed. We assume that the weight distribution $h_{t,v}(w)$ is a beta distribution between $v$ and $v+1$. The beta distribution has two parameters. We parameterise through the expectation $\mu_{v,t}$ and a dispersal parameter $\gamma_{v,t}$ in the spirit of generalised linear models (McCullagh and Nelder, 1989). The parameterisation (described in Appendix A) leads to the variance

$$\mathrm{VAR}(h_{v,t}(\cdot)) = \gamma_{v,t} \cdot \mu'_{v,t} \cdot (1 - \mu'_{v,t}), \tag{4}$$

where $\mu'_{v,t}$ is between 0 and 1 and given by $\mu_{v,t} = v + \mu'_{v,t}$, i.e. the part of the expectation that exceeds the lower limit $v$ of the weight class. Equation (4) implies that the standard deviation increases with the expected value to the middle of the weight class, where the maximum value is reached. The expectation $\mu_{v,t}$ is given by the mean weight $V_t(v)$, which is observed back in time and can be predicted ahead in time.

For $v \geq 1$, $\gamma_{v,t}$ is a constant $\beta^{\gamma}$ independent of time and common for all weight classes. For weight class 0, $\gamma_{v,t}$ varies seasonally according to

$$\gamma_{0,t} = g\left(\beta_0^{\gamma} + \beta_1^{\gamma} \sin\left(\frac{2\pi t}{12}\right) + \beta_2^{\gamma} \cos\left(\frac{2\pi t}{12}\right)\right), \tag{5}$$

where the $\beta$s are unknown regression coefficients. The term $\gamma_{0,t}$ must be between 0 and 1 (see Appendix A), which will be fulfilled by the function

$$g(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}. \tag{6}$$

The total of four unknown $\beta$-parameters are estimated from the data by minimising the prediction errors (see the next section).

We have chosen the beta distribution for $h(\cdot)$ since it is bounded and quite flexible. Note that the proportion $a_{t,v}$ is implicitly a function of both the average weight and the growth (see Figure 2 for an illustration; more details are found in Appendix A).

[Figure 2 about here.]

The number of fish in weight class $v$ in the next month $t+1$ is then given by the balance of numbers,

$$N_{t+1}(v) = N_{t+1}(v, v-1) + N_{t+1}(v, v) - N^R_{t+1}(v) + N^I_{t+1}(v), \tag{7}$$

which is the sum of the number of fish that switched from weight class $v-1$ ($N_{t+1}(v, v-1)$) and those that remained in weight class $v$ ($N_{t+1}(v, v)$), minus the number of fish removed ($N^R_{t+1}(v)$), plus fish stocked ($N^I_{t+1}(v)$). The latter is only non-zero for the first weight class $v = 0$, and the number $N^R_{t+1}(v)$ is the sum of fish lost, slaughtered and wasted.

There is a corresponding, but slightly more complicated, balance of biomass. First, consider those fish that is in weight class $v$ at time $t$ and will remain in the same weight class at time $t+1$. At time $t$, the average weight of these fish is given by

$$V^*_t(v, v) = \frac{\int_v^{v'_{t,v}} w \cdot h_{t,v}(w)dw}{\int_v^{v'_{t,v}} h_{t,v}(w)dw} = \frac{\int_v^{v'_{t,v}} w \cdot h_{t,v}(w)dw}{a_{t,v}}. \tag{8}$$

At time $t+1$, their average weight is increased by the growth factor $f_{t,v}$ to

$$V_{t+1}(v, v) = f_{t,v} \cdot V^*_t(v, v).$$

Correspondingly, consider those fish in weight class $v$ at time $t$, that grow into the next weight class at time $t+1$. At time $t$, the average weight of these fish is given by

$$V^*_t(v+1, v) = \frac{\int_{v'_{t,v}}^{v+1} w \cdot h_{t,v}(w)dw}{1 - a_{t,v}},$$

and at time $t+1$ their average weight has increased to

$$V_{t+1}(v+1, v) = f_{t,v} \cdot V^*_t(v+1, v).$$

To update the average weight, we also need to divide the number of fish removed, $N^R_{t+1}(v)$, into $N^R_{t+1}(v, v)$ and $N^R_{t+1}(v, v-1)$, i.e. into those remaining in the same and those coming from the weight class below at time $t$. This is done in a similar way as for $N_t(v)$ of the standing stock in Equation (2).

Furthermore, we assume that the fish removed and those not removed have the same average weight.

Finally, the average weight of fish in weight class $v$ in the next month $t+1$ is given by

$$\begin{aligned} V_{t+1}(v) =& \frac{\Big(N_{t+1}(v, v-1) - N^R_{t+1}(v, v-1)\Big) \cdot V_{t+1}(v, v-1)}{N_{t+1}(v)} + \\ & \frac{\Big(N_{t+1}(v, v) - N^R_{t+1}(v, v)\Big) \cdot V_{t+1}(v, v)}{N_{t+1}(v)}. \end{aligned} \tag{9}$$

Equation (9) above has to be modified for weight class $v = 0$: The quantity $N_{t+1}(v, v-1) - N^R_{t+1}(v, v-1)$ is replaced by the observed number $N_t^{I,\mathrm{obs}}$ of stocked fish and $V_{t+1}(v, v-1)$ is replaced by the average weight $V_t^I$ of stocked fish, such that

$$V_{t+1}(0) = \frac{N_t^{I,\mathrm{obs}} \cdot V_t^I + \Big(N_{t+1}(v, v) - N^R_{t+1}(v, v)\Big) \cdot V_{t+1}(v, v)}{N_{t+1}(v)}.$$

The average weight of stocked fish is unknown, but varies over the year. Similar to Equation (5), this average weight is therefore modelled as a seasonal function by

$$V_t^I = g\left(\beta_0^I + \beta_1^I \sin\left(\frac{2\pi t}{12}\right) + \beta_2^I \cos\left(\frac{2\pi t}{12}\right)\right), \tag{10}$$

where $g(\cdot)$ is given by Equation (6), and the $\beta$s are unknown parameters estimated from the data by minimising the prediction errors (see the next section).

As a summary, the parameters of the model for the standing stock are listed (Table 4).

[Table 4 about here.]

The proposed model can be compared with the framework of a size-structured or stage-structured model (Quinn and Deriso, 1999). In our case the weight classes are the stages, and the model is based on one-month transitions between the stages. We have imposed the constraint that fish may not grow beyond the next weight class in one month. The model for the number of fish described by (2) and (7) is therefore a special case known as an Usher model (Usher, 1996), which (disregarding fish lost, slaughtered, wasted and stocked) can also be written in matrix form,

$$\boldsymbol{N}_{t+1} = \boldsymbol{M}_t \boldsymbol{N}_t,$$

where $\boldsymbol{N}_t$ is a vector of the number of fish in each weight class. The growth transition matrix $\boldsymbol{M}_t$ is given as: $\boldsymbol{M}_t(v, v) = a_{t,v}$, $\boldsymbol{M}_t(v+1, v) = (1 - a_{t,v})$, where $a_{t,v}$ is given by Equation (3) for weight class $v$, and all other $\boldsymbol{M}_t(i, j) = 0$. Note that $\boldsymbol{M}_t$ is time-varying. Contrary to many of the stage-structured models, we need to model both the weight and the number of fish in each weight class, with a non-linear connection between the two, and based on the unknown weight distribution of fish within each weight class. Our weight model (9) cannot easily be written in matrix form.

**Estimation of model for the standing stock**

The unknown parameters from the model for the standing stock are estimated from the historical data by minimising the prediction errors of monthly numbers and average weight in each weight class. It is essential that the estimated model gives sensible predictions for both abundance and weight of fish, for different weight classes as well as on the short and long term. How these various goals should be balanced is, however, not obvious. Furthermore, it is not obvious which (joint) statistical distribution the numbers and weight of fish follow. We have therefore chosen not to use Maximum Likelihood or any model based fitting criterion, butinstead to minimise a least squares type criterion, which turns out to be relatively robust against modelling error and numerical optimisation problems.

This has the unfortunate consequence that measures of precision for the parameters in Table 4 are not readily available. (The sub-model parameters are accomanied by precision estimates.) However, our primary concern has been to build a proper forecasting model for noisy data, and the precision parameter of the in sample fit is not crucial in pursuing the best forecasting model.

First, let $N^{\text{obs}}_{t+k}(v)$ and $V^{\text{obs}}_{t+k}(v)$ denote the observed number and average weight of fish in weight class $v$ at time $t + k$. Furthermore, let $\widehat{N}_{t+k|t}(v)$ and $\widehat{V}_{t+k|t}(v)$ denote the corresponding conditional $k$-step-ahead predictions, conditioned on all observed quantities at time $t$, and on the observed number of removed (lost, slaughtered, wasted) and stocked fish as well as on the sea temperature and the number of daylight hours at times between $t+1$ and $t+k$. The following fitting criterion takes into account prediction errors for numbers and average weights for the first seven weight classes and for prediction horizons from one month to one year:

$$\sum_{k=1}^{12}\left(\sqrt{\sum_{t=1}^{T}\sum_{v=0}^{6}\Big(w_N(\widehat{N}_{t+k|t}(v)-N^{\text{obs}}_{t+k}(v))\Big)^2}\right.$$
$$\left.+\sqrt{\sum_{t=1}^{T}\sum_{v=0}^{6}w_V(v)\left(\widehat{V}_{t+k|t}(v)-V^{\text{obs}}_{t+k}(v)\right)^2}\right), \tag{11}$$

where $t=1,2,\dots,T$ are months over the data period 2002–2007 and $w_N$ and $w_V(v)$ are weights chosen to give a reasonable balance between weight classes and between numbers and average weights of fish. The weights for numbers are given by $w_N = 1/\overline{N}^{\text{obs}}$, where $\overline{N}^{\text{obs}}$ is the sample mean of the observed number of fish over the entire data period. The first term of the criterion (11) then consists of a sum of relative prediction errors for the number of fish. The weights for average weights put more emphasis on the dominating weight classes, defined as $w_V(v) = N_t(v)^{\text{obs}}/N^{\text{obs}}_t$. Since data in the upper weight classes ($v = 7-10$) are scarce and very variable, these weight class are ignored in the estimation.

**Model for number of fish stocked and other sub-models**

Typically, a lot of fish are stocked in some months (late spring and early autumn) and very few are stocked in other months. There is, in principle, no upper limit on the number of stocked fish, but a lower limit at 0. The number of fish stocked at month $t$ is modelled as a gamma distribution;

$$N^I_t \sim \text{Gamma}(\lambda_t, \sigma_t). \tag{12}$$

Here, $\lambda_t$ is the expected value at month $t$ given by

$$\lambda_t = s(t) + \beta^\lambda_1 \sin\left(\frac{2\pi t}{12}\right) + \beta^\lambda_2 \cos\left(\frac{2\pi t}{12}\right),$$

where $s(t)$ is a smooth trend over the data period with four unknown parameters (see Appendix F for details) and the $\beta$s are unknown regression coefficients expressing seasonal variation. The corresponding standard deviation $\sigma_t$ is a function of the expected value, given by $\sigma_t = \sigma_0(\lambda_t)^\delta$, where

$\delta$ and $\sigma_0$ are unknown parameters. In terms of a standard gamma distribution parameterisation, $\text{shape}_t = \lambda_t^{2(1-\delta)}/\sigma_0^2$ and $\text{scale}_t = \lambda_t/\text{shape}_t$. This distribution allows for a seasonally varying relative variance (or coefficient of variation).

The models for number of fish lost, slaughtered and wasted (Appendix C) are more or less similar. Furthermore, there is a model for the gutted weight of slaughtered fish (Appendix D), and a model for the sea temperature (Appendix E).

All the sub-models, except for the model for the standing stock, are estimated separately by Maximum Likelihood (Pawitan, 2001), by numerically maximising the log likelihood or by using standard software for generalised linear models. For each of the sub-models (and for each weight class), the parameters are estimated conditional on all other data. We assume, for example, that the number of fish $N_{t-1}(v)$ is known when estimating the loss model (C.3) for $N_t^L(v)$.

Our sub-models are richer than what is common in the literature (Rizzo and Spagnolo, 1996; Ferreira et al., 1998; Gangnery et al., 2001). The sub-models are not deterministic, to better represent the actual possible variation between months and years in stocking, loss, slaughtering and waste. Assuming deterministic sub-models for (e.g.) stocking would underestimate the uncertainty in future loss and standing biomass due to uncertain and time-varying loss, and hence produce worse predictions.

**Predictions with uncertainty**

The fitting criterion (Equation (11)) we used to estimate the standing stock model involved in the conditional $k$-step-ahead predictions $\widehat{N}_{t+k|t}(v)$ of numbers and $\widehat{V}_{t+k|t}(v)$ of average weights, where we conditioned not only on data known at time $t$, but also on observed values of stocked and removed fish and sea temperature between time $t+1$ and $t+k$. However, in an ordinary prediction situation, these additional variables are unknown and must be predicted as well, by their respective sub-models. To take into account their prediction uncertainties, Monte Carlo simulated values from the sub-models are used.

Even if the number of stocked and removed fish and the sea temperatures between time $t+1$ and $t+k$ were known, there would be a discrepancy between predicted and observed values, due to model and observational error. To take this uncertainty into account, we compute empirical prediction errors from the historical data, and add samples from these to the predicted values to get a predictive distribution (Appendix B).

Our procedure may be summarised as followed: To predict future states, we i) condition on the present data known at time $t$, ii) sample future random numbers between time $t+1$ and $t+k$ for every term in the model that follows a distribution (e.g. the gamma distribution for the number of fish stocked), iii) compute the number and weight of fish the next $k$ months and iv) add a sample of empirical prediction errors for numbers and average weight. This procedure is repeated $B$ times. The prediction is equal to the sample mean of the $B$ simulations for each future point in time. To distinguish these unconditional predictions from the conditional ones, we use the notation $\widetilde{N}_{t+k|t}(v)$ etcetera. A 90% prediction interval is correspondingly found from the 5% and 95% quantiles of the simulations.

**Validation**

To demonstrate that the model provides sensible and robust predictions, we validate the predictions by out-of-sample validation. Starting with two years of historical data (2002 and 2003), we estimated parameters based on these data and predict the next year. For each prediction month, we compared the prediction with the actual data and computed the prediction error. Next, we used two years plus one month of historical data, two years plus two months of historical data, and so on. Based on this, we computed the mean relative prediction error $k$ months ahead, given the observation $Y_{t+k}$ of one of the quantities of interest (for instance $N_{t+k}$), and the corresponding prediction $\widetilde{Y}_{t+k|t}$ at time $t$;

$$\frac{1}{T-k-t_1+1}\sum_{t=t_1}^{T-k}\frac{|\widetilde{Y}_{t+k|t}-Y_{t+k}|}{Y_{t+k}} \tag{13}$$

for $k = 1 - 12$ months, where $t_1$ is December 2003.

For comparison, the same criterion (13) was computed for a naive predictor, which was chosen to be the last observed value. For one month ahead prediction, this is the value of the current month.

The naive predictor is expected to perform worse if the seasonality is strong, and therefore the comparison may be unfair, yet it is simple and easy to understand. Since the data are quite stationary, a simple regression model with a linear time trend and seasonal variation could be an alternative and improved naive predictor. However, even if such an empirical predictor can give reasonably good predictions within the present historical data, it will neither be useful in the future if important aspects, such as the stocking strategy, changed, nor in different what-if analysis. Therefore, an improved naive predictor is not a realistic alternative to our more causal model, which satisfies the balance of numbers and mass.

**Scenarios for stocked fish or slaughtering strategy**

The model can also be used for investigating what-if scenarios. This means that we replace simulations from one or more of the sub-models with certain scenarios for those quantities. For instance, one may be interested in developing a production strategy that gives less seasonal variations than today. This can be investigated by changing stocking or slaughtering strategies according to given scenarios and study the corresponding changes in seasonal patterns for standing stock or gutted slaughtered biomass.

Scenarios can also be useful if the approximate number of smolt to be set out is known some months in advance. Then, using the known number of future stocked fish instead of the predicted numbers from our stocking model may improve the precision of the predictions of the standing stock considerably.

## Results

### Parameter estimation

[Figure 3 about here.]

The various sub-models were estimated for each of the three regions (South, Mid, Northern Norway) separately on the data from 2002 to 2007. The estimated growth as a function of the sea temperature for four weight classes in Mid Norway is displayed (Figure 3). The relative growth decreases along with increasing weight, and has a maximum when the sea temperature is around 11-12°C.

There are plenty of parameters in the model (but not more than what is justifiable from the amount of data). Many of the parameters are not directly interpretable. Estimates for the most important parameters (the parameters in the model for the standing stock) are supplied for the three regions (Table 5). The intercepts ($\beta^f_{0,0}$, $\beta^f_{0,1}$, $\ldots$) of the growth growth model (1) generally decrease with increasing weight class (in agreement with Figure 3). For equal sea temperature and number of daylight hours, the growth is generally lower in the North. Both the sea temperature and the number of daylight hours growth parameters imply a concave growth function ($\beta_1^f$ and $\beta_3^f$ are positive, $\beta_2^f$ and $\beta_4^f$ are negative) for all three regions. The variance intercept terms ($\beta^\gamma$ and $\beta_0^\gamma$) and the weight of stocked fish intercept terms ($\beta_0^I$) are all quite similar across the regions (after the transformation (6)). Of the corresponding seasonal fluctuations, the absolute values of the parameters for the seasonally varying weight of stocked fish ($\beta_1^I$ and $\beta_2^I$) for the Northern region stand out as high.

[Table 5 about here.]

**Predictions with uncertainty**

We display two examples of predictions with uncertainty, based on aggregation of regional results to make predictions for Norway (the predictions in Figures 4–5 are based on $B = 1\ 000$ simulations). The results are shown for weight classes 0–5, since the upper ones are of marginal interest.

The predicted biomass in each weight class (Figure 4) is increasing, in line with the historical trend, and approximately proportional to the predicted number of fish (not shown), because the biomass is the product of the number of fish and their average weight. The latter is quite stable over time and is predicted with rather high precision (results not shown).

The predicted slaughtered biomass of gutted fish (Figure 5) displays a more diverse picture, also because the slaughtered biomass historically has been more variable than the current stock quantities. The slaughtered biomass is quite uncertain due to uncertainty in both numbers and weight, as well as historically changing slaughtering strategies.

Generally, the relative uncertainty is increasing with prediction horizon and weight class, as expected.

[Figure 4 about here.]

[Figure 5 about here.]

**Model validation**

To investigate the goodness of the model fit, we performed out-of-sample validation of the predictions on the data for the period 2004–2007 (at least two years are used for estimating the model). The mean relative prediction error (13) for the biomass is shown (Figure 6). As expected, the prediction error generally increases with the prediction horizon and the weight class. The model based predictions are better than the naive predictor in the short run, and approximately equal to the naive predictor twelve months ahead.

The results for the numbers of fish (not shown) were very similar. As explained above, this is because the average weight per fish per weight class is both stable over time and predicted rather precisely.

The mean relative prediction error for the slaughtered biomass is presented (Figure 7). The results for weight classes 0–2 are of marginal interest, and therefore not shown here, since most of the slaughtering takes place in weight classes 3–5. Here, the results based on the model are slightly better than the naive predictor, but the prediction error is large (20–40%) for both predictors, and at its lowest (20%) in weight class 4, where most of the biomass is being slaughtered.

We also estimated how often the prediction intervals were violated. The results (not shown here) indicate that the uncertainty is somewhat underestimated.

[Figure 6 about here.]

[Figure 7 about here.]

**What-if scenarios for stocking strategy and slaughtering strategy**

We performed a scenario analysis on stocking strategy to demonstrate how the model can be used for what-if analysis to investigate consequences of potential changes in production strategy.

One problem in the production has been the strong seasonality, which to some extent forces the amount of slaughtered fish to be driven by the supply and not by the demand of fish. Therefore, it is of interest to investigate if other production strategies can give a more stable production over time.

Traditionally, most fish have been stocked during spring and autumn (see Table 2). As a hypothetical alternative to the current practice, we assumed a scenario where all the fish were stocked during the spring. This is an extreme stocking strategy, but provides a good illustration of the model. We compared the results with a scenario where the current stocking strategy was followed as predicted by our stocking sub-model (Equation (12)). The slaughtering strategy was the same in the two scenarios, given by our slaughter model to represent the current strategy. Predictions for the two scenarios are displayed (Figure 8), conditioned on the situation in December 2007 (the prediction intervals depicted in Figure 4 are not shown here). Since the predictions for the first months depend highly on the initial conditions, we produced forecasts up to four years ahead, to see the long term effect of the stocking strategy. For weight class 0, the stocking strategy resulted in a more pronounced seasonality. For weight classes 2 and 3, the seasonal variation was reduced a great deal. For weight classes 4 and 5, however, the reduction is less, due to the high number of fish being slaughtered during the autumn.

[Figure 8 about here.]

Slaughtering is not only governed by changing prices. Sometimes there is a need to slaughter to avoid too many fish in the upper weight classes. Our second scenario therefore combines the stocking scenario (Figure 8) with a slaughtering strategy where (as far as possible) an equal amount of fish is slaughtered each month in weight classes 3–5 (Figure 9). There is little slaughtering in weight classes 0–2 (Table 3), and we use the sub-models for slaughtering from these weight classes. The result (compared to Figure 8) is quite few changes in weight classes 4 and 5. This demonstrates that, allowing for a stocking strategy, a flat slaughtering strategy is possible with quite few consequences for the standing stock.

[Figure 9 about here.]

## Discussion

We have described a statistical model to predict the stock of Norwegian farmed salmon. The model was designed for our unique data set, where production data from all Norwegian fish farms have been electronically collected into the database Havbruksdata since 2002. The model starts with the number and the weight of fish in a weight class. We computed how many fish that remain in a weight class and how many that grow into the next weight class. In addition, we modelled the number of new fish stocked, lost, slaughtered and wasted. Based on the statistical validation, we concluded that the model provides sensible predictions with uncertainty of future production of Norwegian farmed salmon. In addition to prediction, the model can be used for what-if scenarios, particularly for testing the effect of changed stocking and slaughtering strategies.

The growth rate of individual farmed salmon (from the same age class) depends on their weight, temperature in the sea, light conditions given by geographical location and amount of feed (Gardeur et al., 2001; Mørkøre and Rørvik, 2001). We have chosen to disregard feed consumption to explain growth, since it is difficult to separate the effect of feed consumption from the effects of sea temperature and seasonality. Other factors may impact production, such as a suboptimal environment, reduced fish health and technological challenges. However, the influence of these factors are not systematic and therefore almost impossible to model and predict. Usually these factors will influence on a smaller part of the production and will therefore not have a major influence on the estimated values.

Our model can be compared with the framework of a size-structured or stage-structured model (Quinn and Deriso, 1999). However, we go beyond the standard stage-structured model, since both the number of fish and the average weight of fish in each weight class (stage) are modelled, and we estimated the unknown weight distribution of fish in each weight class to accomplish this.

The parameters of our model were estimated from and validated on one coherent data set. In contrast to similar models in the literature, which usually are deterministic, our sub-models are stochastic to better represent the actual variation in stocking, loss and slaughtering.

One may wonder why the regional approach, where data are aggregated over multiple farms, was used instead of a set of single-farm models. This was not a feasible option for two reasons: i) The data are far from error free. Even though all farmers are obliged to report each month, data from a few farms were missing each month. The proportion of missing farms was known, and the data were scaled up to avoid bias. However, even with all farms reporting their numbers, the numbers themselves are not error free. The impact of these errors was reduced when aggregating over multiple farms. ii) Individual farm data are protected and can not be used without a written consent.

Forecasting production fluctuations may be an important tool for optimising the production planning for Norwegian fish farms, especially since large companies represent the majority of the Norwegian production. There is a range of interesting what-if scenarios to investigate. We have already studied a stocking and a slaughtering strategy. One could also study the effect of changing framework conditions, for example: i) High prices lead to excess slaughtering in weight class 3, resulting in fewer fish in weight classes 4 and 5 later on. ii) Fish diseases lead to obligatory slaughtering of half the fish in weight class 2 in a particular month. iii) The number of stocked fish is halved one year due to smolt production problems. iv) Increased spring slaughtering is needed to cope with increased demand during the spring.

Any animal production involving cycles may benefit from this tool, since it takes quite some time between the time a producer decides to expand the production, and the time those animals have reached their slaughter weight. Both for livestock and farmed fish, the long and short term impact of market prices can be built into the model.

The model is, however, a first attempt at predicting the production, and can be improved in many ways. Since the balances of number and mass are not always correct, there are inconsistencies in the data. An improved model, where inconsistent data were weighted down, could provide even better predictions.

To achieve that goal, and avoid our somewhat ad-hoc approach, we advocate the use of Bayesian hierarchical models (Gelman et al., 2004) within a state-space framework (Durbin and Koopman, 2001). This allows for seperate modelling of the system process and the data process. The system process is essentially the theoretical model for the standing stock and the sub-models. The data process describes the observable quantities, measurement errors and how the data are connected to the system process. Still, how to properly balance the numbers and weight of fish, as well as the short and long term performance of the model, continues to be a challenge.

**Acknowledgments**

We thank Peder A. Jansen for useful suggestions on drafts of this paper. This research has been funded by The Fishery and Aquaculture Industry Research Fund (FHF), project number 263044, and the Research Council of Norway project “Modern Application driven Statistical Challenges (186951/I30)”.

## References


Allen, P. 1994. Economic-forecasting in agriculture. Int. J. Forecasting **10**(1): 81–135.

Asche, F., Guttormsen, A., Sebulonsen, T., and Sissener, E. 2005. Competition between farmed and wild salmon: the Japanese salmon market. Agr. Econ. **33**(3): 333–340.

Bjørndal, T. 1988. Optimal harvesting of farmed fish. Mar. Resour. Econ. **5**: 139–159.

Bjørndal, T. 1990. The Economics of Salmon Aquaculture. Blackwell Scientific Publications, Oxford, United Kingdom.

Boeuf, G., and Le Bail, P. 1999. Does light have an influence on fish growth? Aquaculture **177**(1-4): 129–152, 3rd International Symposium on Research for Aquaculture - Fundamental and Applied Aspects, Barcelona, Spain, Aug 24-27, 1997.

Crespi, J. M., Xia, T., and Jones, R. 2010. Market power and the cattle cycle. Amer. J. Agr. Econ. **92**(3): 685–697.

Durbin, J., and Koopman, S. J. 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford, United Kingdom.

Ferreira, J. G., Duarte, P., and Ball, B. 1998. Trophic capacity of Carlingford Lough for oyster culture – analysis by ecological modelling. Aquat. Ecol. **31**: 361–378.

Gangnery, A., Bacher, C., and Buestel, D. May 2001. Assessing the production and the impact of cultivated oysters in the Thau lagoon (Mediterranee, France) with a population dynamics model. Can. J. Fish. Aquat. Sci. **58**(5): 1012–1020.

Gardeur, J., Paspatis, M., Gelineau, A., and Boujard, T. 2001. Biostatistical implications of individual variability in growth in rainbow trout and Atlantic salmon. Aquaculture **195**: 51–59.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. Bayesian Data Analysis, 2nd Edition. Chapman & Hall/CRC, Boca Raton, Florida.

Guillotreau, P. 2004. How does the European seafood industry stand after the revolution of salmon farming: An economic analysis of fish prices. Mar. Policy **28**(3): 227–233.

Halachmi, I., Simon, Y., Guetta, R., and Hallerman, E. 2005. A novel computer simulation model for design and management of re-circulating aquaculture systems. Aquacult. Eng. **32**: 443–464.

Hamilton, S., and Kastens, T. 2000. Does market timing contribute to the cattle cycle? Amer. J. Agr. Econ. **82**(1): 82–96.

Hilborn, R., and Mangel, M. 1997. The Ecological Detective: Confronting Models with Data. Monographs in population biology. Princeton, Princeton, New Jersey.

Kinnucan, H., and Myrland, O. 2005. Effects of income growth and tariffs on the world salmon market. Appl. Econ. **37**(17): 1967–1978, 50th Workshop on Global Aquaculture, Bergen, Norway, June 11, 2003.

Kumbhakar, S., and Tveterås, R. 2003. Risk preferences, production risk and firm heterogeneity. Scand. J. Econ. **105**: 275–293.

McCullagh, P., and Nelder, J. A. 1989. Generalized Linear Models, 2nd Edition. Chapman & Hall, London, United Kingdom.

Meeus, J. 1991. Astronomical Algorithms. Willmann-Bell, Richmond, Virginia.

Mørkøre, T., and Rørvik, K. 2001. Seasonal variations in growth, feed utilisation and product quality of farmed Atlantic salmon (Salmo salar) transferred to seawater as 0+smolts or 1+smolts. Aquaculture **199**: 145–147.

Pawitan, Y. 2001. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press, Oxford, United Kingdom.

Quinn, T. J., and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press, New York, New York.

Rizzo, G., and Spagnolo, M. 1996. A Model for the Optimal Management of Sea Bass Dicentrarchus Labrax Aquaculture. Mar. Resour. Econ. **11**: 267–286.

Russell, S. W. 1929. Forecasting hog production and marketing. J. Amer. Statistical Assoc. **24**(165): pp. 225–233.

Tuljapurkar, S., and Caswell, H. 1997. Structured-Population Models in Marine, Terrestrial, and Freshwater Systems. Vol. **18** of Population and Community Biology Series. Springer, New York, New York.

Usher, M. B. 1996. A matrix approach to the management of renewable resources, with special reference to selection forests. J. Appl. Ecol. **3**: 355–367.

## Appendix A The weight distribution in one weight class

We have assumed that the weight $w$ of fish in weight class $v$ is beta distributed between $v$ and $v+1$. If we write $w = v + x$, where $x$ is the weight that exceeds the lower limit in the weight class, $x$ is beta distributed between 0 and 1 with density

$$h'(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1} = x^{p-1}(1-x)^{q-1}/B(p,q), \qquad \text{(A.1)}$$

where we use $h'(\cdot)$ to distinguish it from the shifted distribution $h(\cdot)$ between $v$ and $v+1$. Here $p > 0$ and $q > 0$ are parameters, $\Gamma(\cdot)$ is the gamma function and $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$. The relation between $p$ and $q$ and the parameters $\mu$ and $\gamma$ we used in the model for the standing stock, are found from:

$$\text{Expected value} = \mu = (V_t^{\text{obs}} - v) = \frac{p}{p+q},$$

$$\begin{aligned} \text{Variance} &= \frac{pq}{(p+q)^2(p+q+1)} = \gamma(\mu(1-\mu)) \\ &\Downarrow \\ \gamma &= \frac{1}{p+q+1}. \end{aligned} \qquad \text{(A.2)}$$

The parameter $\gamma$ is between 0 and 1, since both $p$ and $q$ are greater than 0.

Using the definition of the beta density (A.1), we may write

$$\begin{aligned} x \cdot h'(x; p, q) &= x \cdot x^{p-1}(1-x)^{q-1}/B(p,q) \\ &= \frac{B(p+1,q)}{B(p,q)} x^p (1-x)^{q-1}/B(p+1,q) \\ &= \frac{B(p+1,q)}{B(p,q)} h'(x; p+1, q). \end{aligned}$$

The average weight $V_t^*(v,v)$ of those fish that will remain in their weight class $v$, given in Equation (8), can be computed by

$$
\begin{aligned}
V_t^*(v,v) &= \frac{\int_v^{v'} w \cdot h(w;p,q)dw}{a} \\
&= \frac{\int_0^{v'-v}(x+v)h'(x;p,q)dx}{a} \\
&= \frac{\int_0^{v'-v} xh'(x;p,q)dx}{a} + v\frac{\int_0^{v'-v} h'(x;p,q)dx}{a} \\
&= \frac{B(p+1,q)}{B(p,q)} \cdot \frac{\int_0^{v'-v} h'(x;p+1,q)dx}{a} + v.
\end{aligned}
$$

Here, $a$ is the proportion of fish remaining in the weight class $v$ in month $t$, defined in Equation (3). To simplify the notation, the indices $v$ and $t$ are here omitted from $h$, $h'$ and $a$.

## Appendix B Empirical prediction errors

The model for the standing stock, conditioned on the values of removed and stocked fish and sea temperatures, is essentially deterministic. The reason is that we require a model that is causal in structure and strictly fulfils the principles of balance of numbers and biomass. This is easier to achieve in a deterministic model. However, the data contain measurement errors and the model is not perfect (model error). To take this uncertainty into account when calculating the prediction uncertainty, we calculate empirical residuals. Then we add samples from these empirical residuals to the predicted values.

The empirical $k$-step ahead prediction errors are the differences between predicted and observed values,

$$\begin{aligned}\widehat{e}^{N}_{t+k} &= \widehat{N}_{t+k|t}(v) - N^{\text{obs}}_{t+k}(v)\\ \widehat{e}^{V}_{t+k} &= \widehat{V}_{t+k|t}(v) - V^{\text{obs}}_{t+k}(v).\end{aligned}$$

For each time point $t$, a consecutive series of empirical prediction errors is calculated for $k = 1$ up to usually $k = 24$ for both numbers and weight, to keep the correlation structure between prediction errors of numbers and weight and between different prediction horizons. When predicting, we first predict for the entire prediction horizon, and then add a sample of the correlated prediction errors afterwards. When predicting 24 months ahead, this sample consists of 24 consecutive prediction errors from one historical 24 month prediction. This is repeated several times, giving a distribution of the prediction errors.

# Appendix C Sub-models for number of fish lost, slaughtered and wasted

When updating the number of fish in each weight class from one time point to the next, the order of the different steps influences the result. The fish removed at each time step is the sum of fish lost, slaughtered and wasted. First, we update the number of fish lost, then the number of fish slaughtered and wasted and finally the growth of the fish remaining, where some fish grow into the next weight class. In addition, the stocked fish are added in weight class 0.

Therefore, the number of fish lost at time $t$ in weight class $v$ can at most be the number of fish in that weight class at time $t-1$. For a given probability $p^L_{t,v}$, the number of fish lost $N^L_t(v)$ is binomially distributed $(p^L_{t,v}, N^L_t(v))$. To allow for over-dispersion, we further assume that the probability $p^L_{t,v}$ is stochastic and follows a beta distribution with expectation $\pi^L_{t,v}$ and an additional parameter $a^L_v$, where $a^L_v$ corresponds to the parameter $p$ in Equation (A.1). The expectation $\pi^L_{t,v}$ depends on the time and the weight class, while $a^L_v$ depends on the weight class only. The number of fish lost is therefore beta-binomially distributed (Agresti, 2002),

$$N^L_t(v) \sim \text{Beta-binomial}\Big(\pi^L_{t,v}, a^L_v, N_{t-1}(v)\Big). \tag{C.3}$$

The expected proportion or probability $\pi^L_{t,v}$ is modelled as a smoothly varying trend, using a logit link where the linear predictor is a smooth function of time (Appendix F) plus a seasonal adjustment with one parameter per month.

After fish lost have been removed, there are $N_{t-1}(v) - N^L_t(v)$ left. Let $N^{SW}_t(v) = N^S_t(v) + N^W_t(v)$ be the sum of slaughtered and wasted fish. Similar to fish lost ($N^L_t(v)$), $N^{SW}_t$ is modelled as a proportion of remaining fish,

$$N^{SW}_t(v) \sim \text{Beta-binomial}\Big(\pi^{SW}_{t,v}, a^{SW}_v, N_{t-1}(v) - N^L_t(v)\Big).$$

## Appendix D Model for weight of slaughtered fish

Monthly data are available for mean gutted weight of slaughtered fish in each weight class, which is lower than the weight of live fish. The gutted weight of slaughtered fish is assumed to be proportional to the weight per fish in the same weight class in the same month, and follows a gamma distribution;

$$V_t^G(v) \sim \text{Gamma}\Big(\beta_{v,t}^G \cdot V_t(v), a_v^G\Big),$$

where the product $\beta_{v,t}^G \cdot V_t(v)$ is the expected value per month and $a_v^G$ is the shape parameter. In addition, we let the proportionality factor $\beta_{v,t}^G$ vary smoothly with time (Appendix F). This model, as opposed to Equation (12), falls into the generalised linear modelling framework (McCullagh and Nelder, 1989).

## Appendix E Model for sea temperature

The sea temperature affects the growth. Hence, to predict and simulate future possible growth, we have to predict and simulate the sea temperature. We assumed that the temperature varies randomly around a seasonal term, and that the temperature may not deviate too much from this trend, where the deviations (residuals) are modelled as an AR(1) process:

$$\begin{aligned} ST_t &= \Lambda_t + \varepsilon_t, \\ \varepsilon_t &= \alpha\varepsilon_{t-1} + w_t, \end{aligned}$$

where $w_t \sim \mathcal{N}(0, \sigma^2_{\mathrm{ST},})$. Here, $\Lambda_t$ is a seasonal term described by seven parameters:

$$\begin{aligned} \Lambda_t = \beta_0^{ST} &+ \beta_1^{ST} \sin\left(\frac{2\pi t}{12}\right) + \beta_2^{ST} \cos\left(\frac{2\pi t}{12}\right) \\ &+ \beta_3^{ST} \sin\left(\frac{4\pi t}{12}\right) + \beta_4^{ST} \cos\left(\frac{4\pi t}{12}\right) \\ &+ \beta_5^{ST} \sin\left(\frac{8\pi t}{12}\right) + \beta_6^{ST} \cos\left(\frac{8\pi t}{12}\right). \end{aligned}$$

## Appendix F Time trend

In the models for the number of stocked fish, the number of fish lost, slaughtered and wasted and weight for slaughtered fish, we use a smooth time trend to handle changes in the overall level over time. Such changes may occur due to improved farming in some sense (fewer deaths) or changed routines, like gradually slaughtering heavier fish. The time trend is modelled by B-splines (Eilers and Marx, 1996). A B-spline of degree $q$ consists of $q+1$ polynomial pieces, each of degree $q$, on $n'$ intervals. Let $t = 1, 2, \ldots, T$ be the historical times (number of months since the first observation). Then the time level is given by the linear combination

$$s(t) = \sum_{l=1}^{r} \alpha_l b_l(t), \tag{F.4}$$

where $b_l(t)$ denotes the $l$th B-spline with regression coefficient $\alpha_l$ and $r$ is the number of equidistant B-splines covering the interval $[1, T]$. We chose B-splines for their flexibility, and at the same time they fit in the linear regression modelling framework. A further improvement would have been to include penalties to facilitate further flexibility, while still controlling the effective number of parameters (Eilers and Marx, 1996).

When adding extra parameters, as we did for the time trend, we ran the risk of over-fitting. Partly by considering Akaike's Information Criteria (Pawitan, 2001), we saw that the time trend was justified, with $q = 2$ and $r = 4$ in (F.4), at least for the major weight classes (0-5), which are the important ones. This criterion has also been used for selecting the sea temperature model (Appendix E).

When predicting, we used $s(T)$ for all future time points. Hence, we assumed a constant level for future time points, since we did not assume that we can predict future time level changes.

## References

Agresti, A. 2002. Categorical data analysis. Probability and Statistics. Wiley, Hoboken, New Jersey.

Eilers, P. H. C., and Marx, B. D. 1996. Flexible Smoothing with B-splines and Penalties (with discussion). Stat. Sci. **11**(2): 89–121.

McCullagh, P., and Nelder, J. A. 1989. Generalized Linear Models, 2nd Edition. Chapman & Hall, London, United Kingdom.

Pawitan, Y. 2001. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press, Oxford, United Kingdom.

## List of Figures

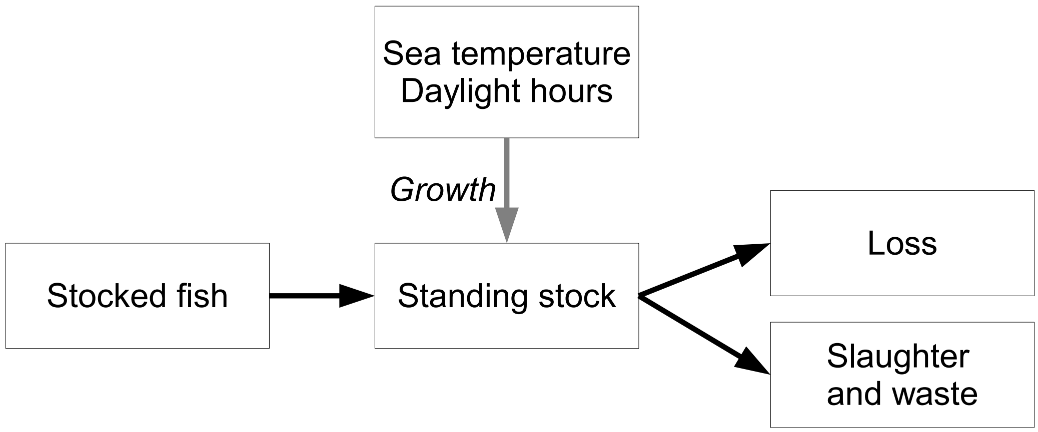

Sea temperature
Daylight hours
Growth
Stocked fish
Standing stock
Loss
Slaughter
and waste

Figure 1:

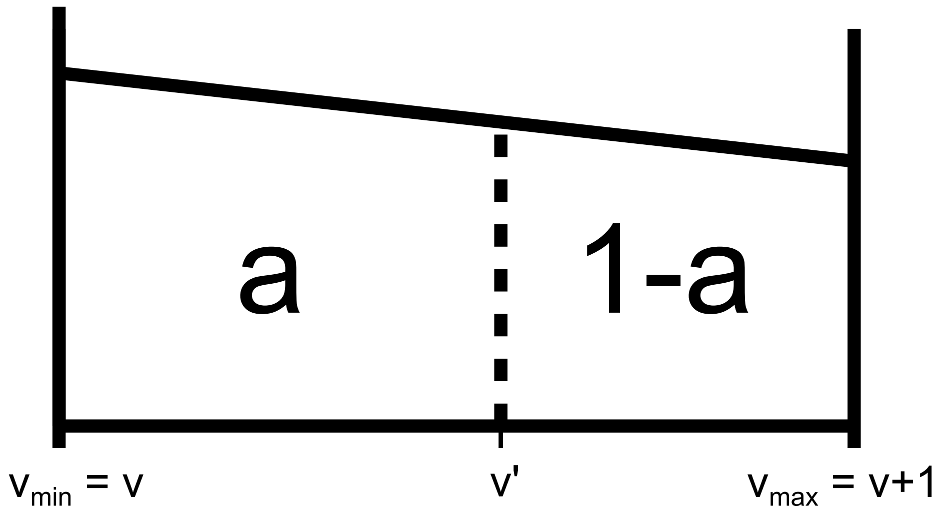



Figure 2:

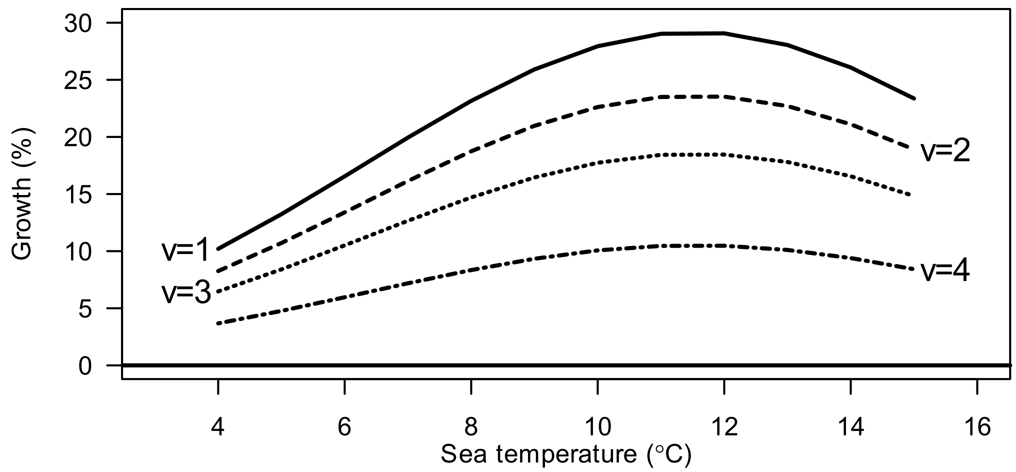
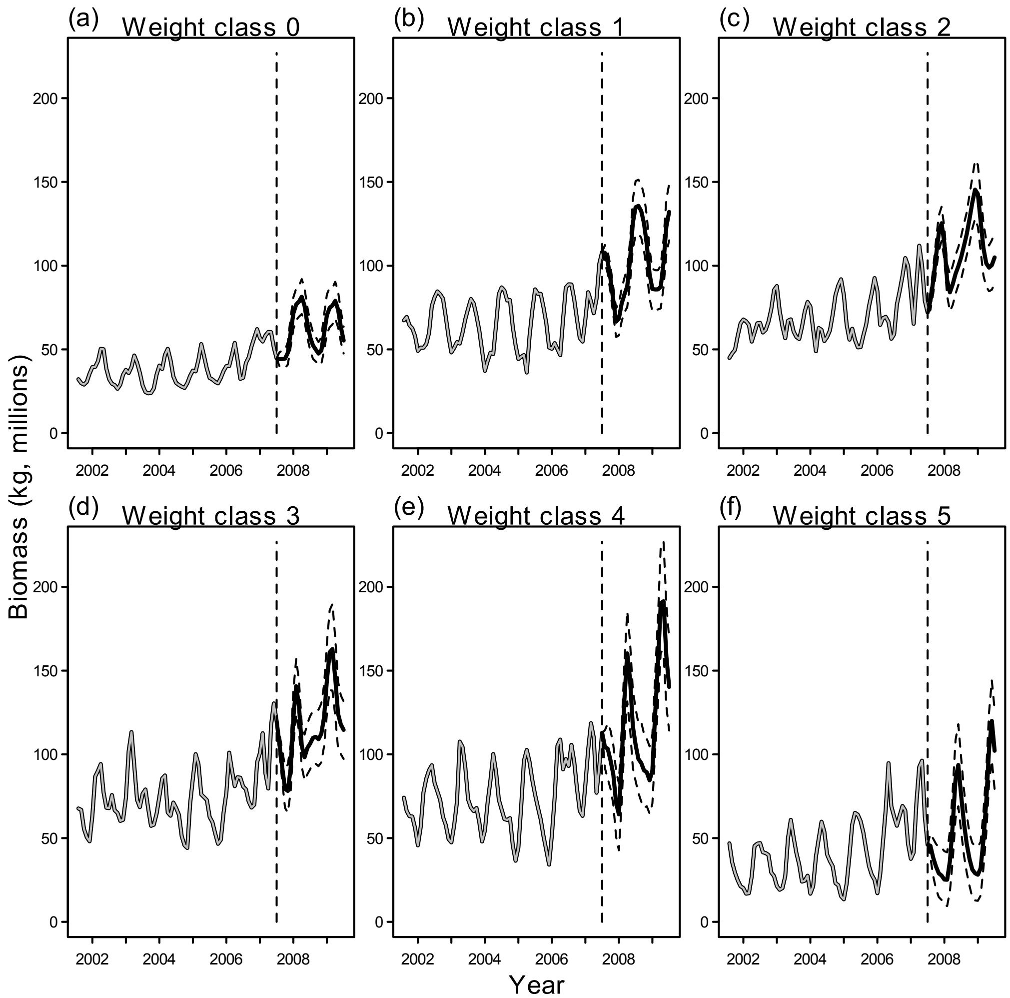

30
25
20
15
10
5
0
Growth (%)
v=1
v=2
v=3
v=4
4
6
8
10
12
14
16
Sea temperature (°C)

Figure 3:



(a) Weight class 0
(b) Weight class 1
(c) Weight class 2
(d) Weight class 3
(e) Weight class 4
(f) Weight class 5
Biomass (kg, millions)
Year
0
50
100
150
200
2002
2004
2006
2008


Figure 4:

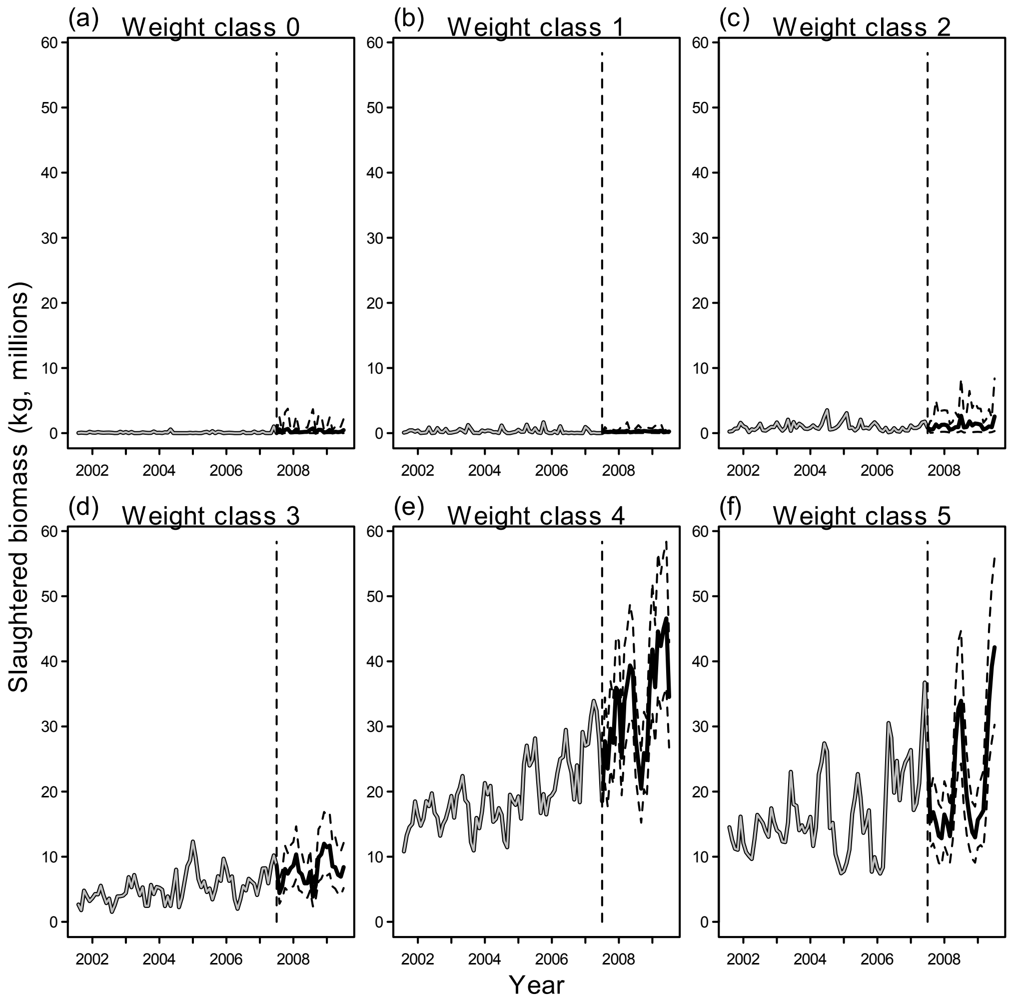

(a) Weight class 0
(b) Weight class 1
(c) Weight class 2
(d) Weight class 3
(e) Weight class 4
(f) Weight class 5
Slaughtered biomass (kg, millions)
Year
0
10
20
30
40
50
60
2002
2004
2006
2008

Figure 5:

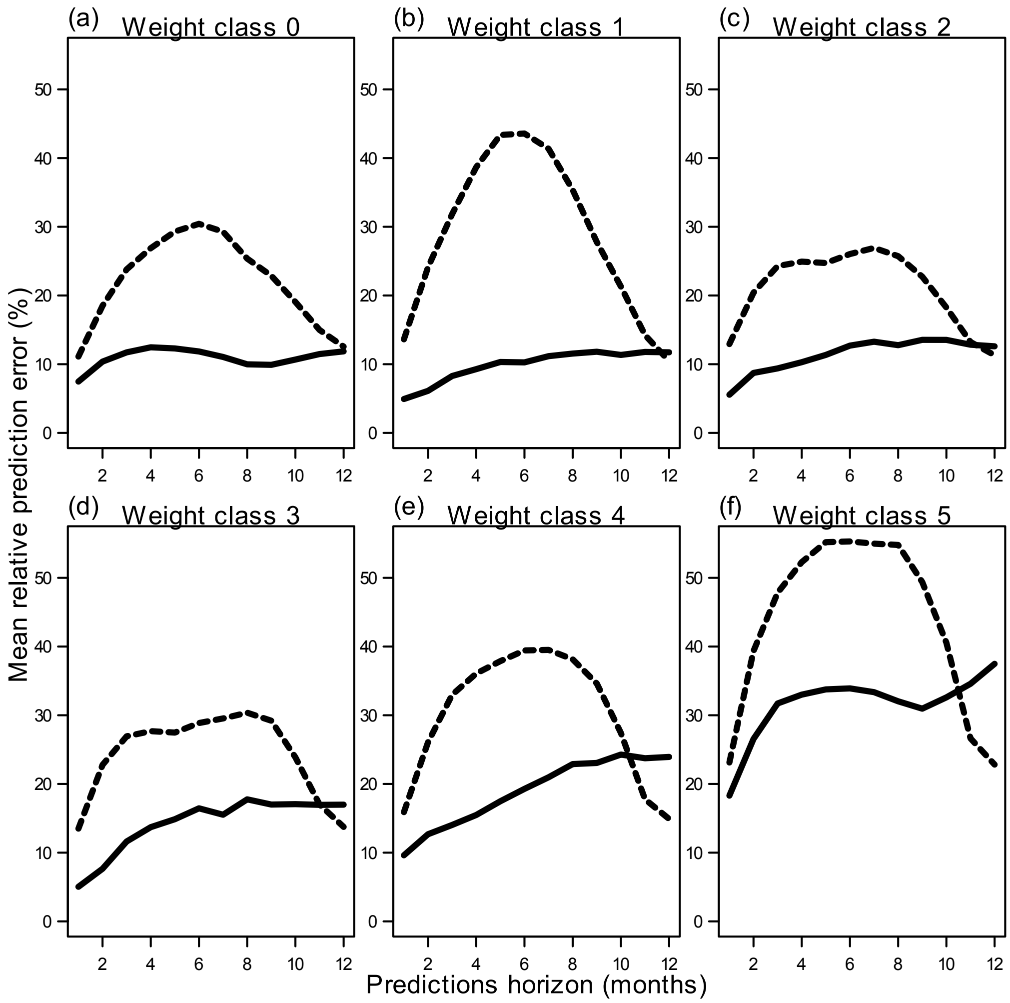

(a) Weight class 0
(b) Weight class 1
(c) Weight class 2
(d) Weight class 3
(e) Weight class 4
(f) Weight class 5
Mean relative prediction error (%)
Predictions horizon (months)
0
10
20
30
40
50
2
4
6
8
10
12

Figure 6:

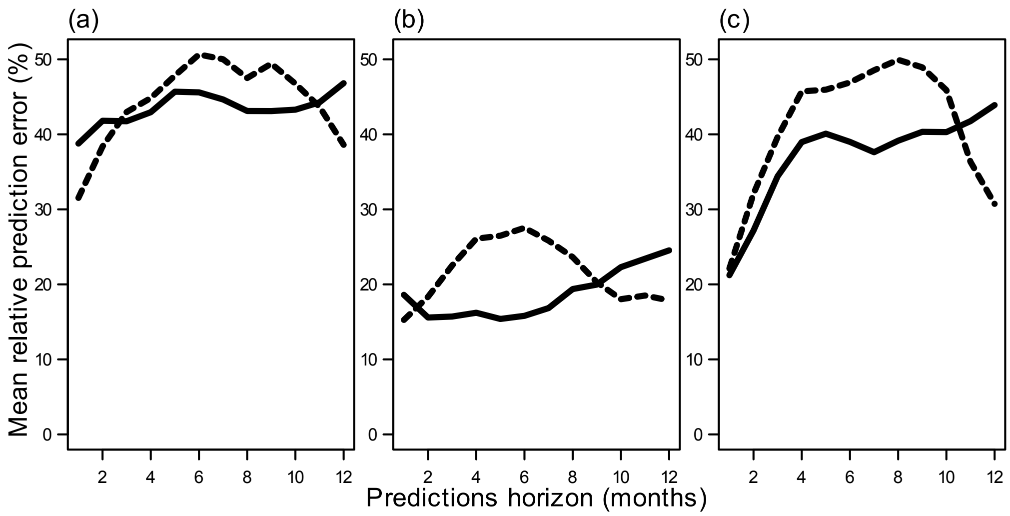
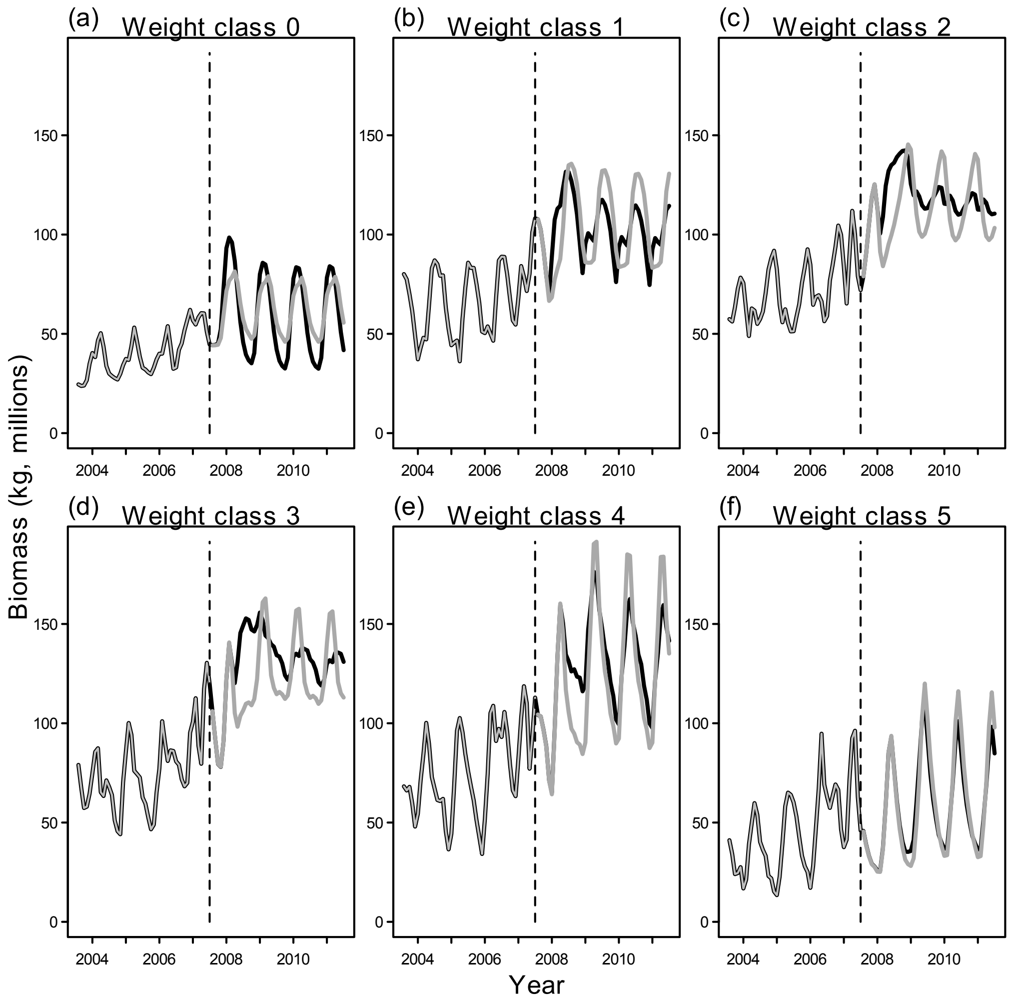

(a)
(b)
(c)
Mean relative prediction error (%)
0
10
20
30
40
50
2
4
6
8
10
12
Predictions horizon (months)

Figure 7:



(a) Weight class 0
(b) Weight class 1
(c) Weight class 2
(d) Weight class 3
(e) Weight class 4
(f) Weight class 5
Biomass (kg, millions)
Year
0
50
100
150
2004
2006
2008
2010


Figure 8:

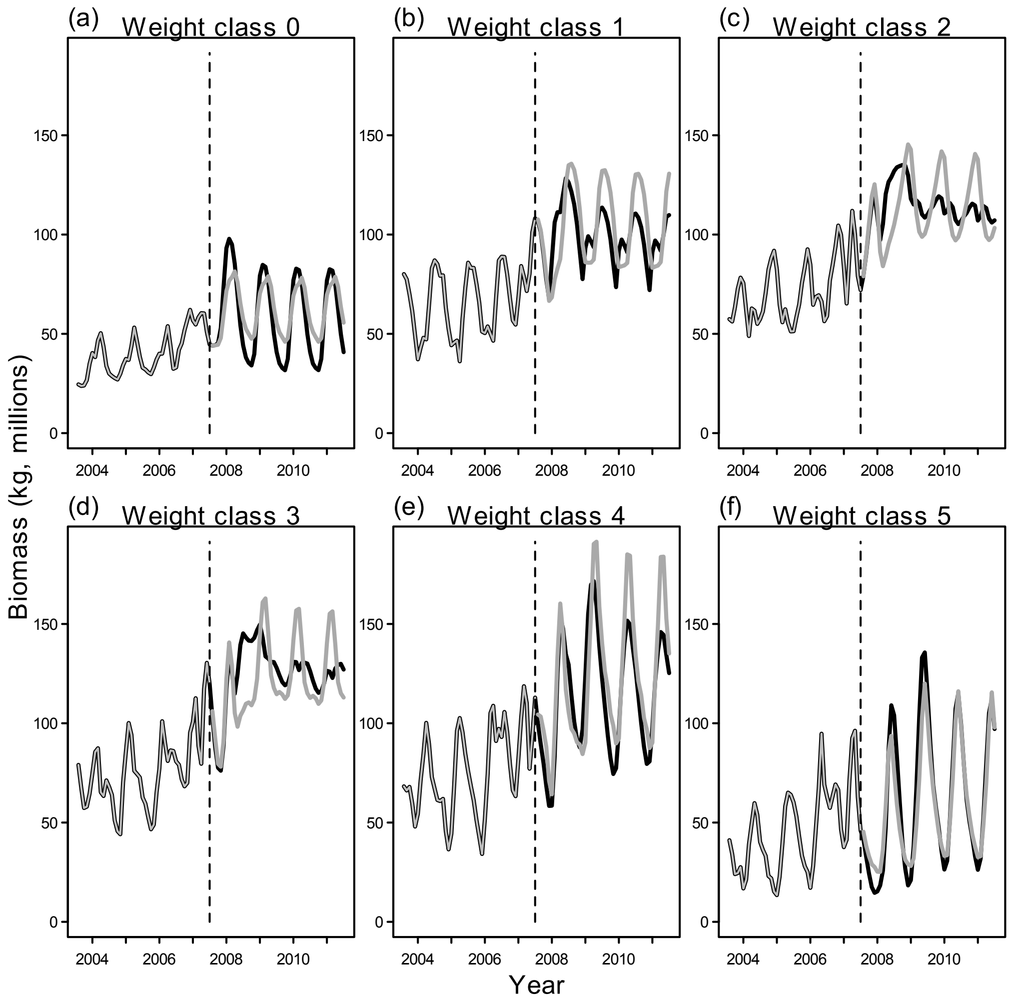

(a) Weight class 0
(b) Weight class 1
(c) Weight class 2
(d) Weight class 3
(e) Weight class 4
(f) Weight class 5
Biomass (kg, millions)
Year
0
50
100
150
2004
2006
2008
2010

Figure 9:

Table 1: Overview of data and notation. (#) denotes observable quantities.

(a) Total numbers and numbers per weight class and month $t$. Note that $B_t = \sum_{v=0}^{10} B_t(v)$, $N_t = \sum_{v=0}^{10} N_t(v)$, $V_t^B = B_t/N_t$, etcetera.

| Biomass | Number of fish | Weight per fish | Description |
|---|---|---|---|
| (#) $B_t$ | (#) $N_t$ | (#) $V_t^B$ | Standing stock |
| $B_t^L$ | (#) $N_t^L$ | $V_t^L$ | Lost |
| $B_t^S$ | (#) $N_t^S$ | $V_t^S$ | Slaughtered |
| (#) $B_t^G$ | as above | (#) $V_t^G$ | Gutted (slaughtered) |
| $B_t^W$ | (#) $N_t^W$ | $V_t^W$ | Wasted |
| $B_t^R$ | (#) $N_t^R$ | $V_t^R$ | Lost, slaughtered and wasted |
| $B_t^I$ | (#) $N_t^I$ | $V_t^I$ | Stocked (in weight class 0) |
| (#) $B_t(v)$ | (#) $N_t(v)$ | (#) $V_t(v)$ | At time $t$ in weight class $v$ |
| | $N_t(v, v-1)$ | $V_t(v, v-1)$ | At time $t$ in weight class $v$, who were in weight class $v-1$ at time $t$ |

(b) Other quantities.

| | Description |
|---|---|
| (#) $ST_t$ | Sea temperature |
| (#) $D_t$ | Number of daylight hours |
| $f_{t,v}$ | Growth function |

Table 2: Distribution of Norwegian farmed salmon data per month (in percentage of yearly quantities), computed as sample averages over the data period 2002–2007; number of fish in the standing stock ($N_t$), standing biomass ($B_t$), number of fish lost ($N_t^L$) and slaughtered ($N_t^S$), gutted slaughtered biomass ($B_t^G$) and number of stocked fish ($N_t^I$).

| | $N_t$ | $B_t$ | $N_t^L$ | $N_t^S$ | $B_t^G$ | $N_t^I$ |
|---|---|---|---|---|---|---|
| Jan. | 7.7 % | 8.5 % | 6.3 % | 7.3 % | 7.6 % | 0.4 % |
| Feb. | 7.3 % | 8.1 % | 5.2 % | 6.7 % | 7.0 % | 0.4 % |
| Mar. | 7.0 % | 7.6 % | 5.2 % | 8.0 % | 8.2 % | 2.0 % |
| Apr. | 7.3 % | 7.3 % | 6.2 % | 7.5 % | 7.3 % | 12.0 % |
| May. | 8.7 % | 7.2 % | 9.0 % | 8.2 % | 8.1 % | 29.5 % |
| June | 9.0 % | 7.2 % | 13.3 % | 8.4 % | 8.2 % | 13.4 % |
| July | 8.7 % | 7.8 % | 11.2 % | 7.6 % | 7.3 % | 3.7 % |
| Aug. | 8.6 % | 8.6 % | 10.1 % | 7.9 % | 7.6 % | 4.5 % |
| Sept. | 8.9 % | 9.2 % | 9.0 % | 8.8 % | 8.6 % | 13.5 % |
| Oct. | 9.3 % | 9.6 % | 8.5 % | 9.5 % | 9.6 % | 14.3 % |
| Nov. | 9.1 % | 9.6 % | 8.5 % | 10.2 % | 10.6 % | 5.1 % |
| Dec. | 8.6 % | 9.2 % | 7.6 % | 10.0 % | 10.0 % | 1.1 % |
| Total | 100 % | 100 % | 100 % | 100 % | 100 % | 100 % |

Table 3: Summary of Norwegian farmed salmon data per weight class (average quantities per month, computed as sample averages over the data period 2002–2007). Number of fish lost ($N_t^L$) and slaughtered ($N_t^S$) is given in percentage of the number of fish in the standing stock ($N_t$) in each weight class. Biomass ($B_t$ and $B_t^G$) is given in 1 000 tonnes, number of fish ($N_t$ and $N_t^I$) is given in 1 000 000.

| | $N_t$ | $B_t$ | $N_t^L$ | $N_t^S$ | $B_t^G$ | $N_t^I$ |
|---|---|---|---|---|---|---|
| Weight class 0 | 93.2 | 38.2 | 1.7 % | 0.2 % | 0.1 | 19.4 |
| Weight class 1 | 44.5 | 64.9 | 0.8 % | 0.4 % | 0.3 | – |
| Weight class 2 | 27.6 | 68.0 | 0.5 % | 1.2 % | 0.9 | – |
| Weight class 3 | 21.3 | 74.2 | 0.6 % | 5.9 % | 5.0 | – |
| Weight class 4 | 16.6 | 74.0 | 0.7 % | 24.5 % | 19.3 | – |
| Weight class 5 | 7.7 | 41.6 | 0.6 % | 38.8 % | 16.5 | – |
| Weight class 6 | 1.9 | 12.2 | 0.5 % | 38.6 % | 4.7 | – |
| Weight class 7 | 0.4 | 2.7 | 0.6 % | 38.0 % | 1.0 | – |
| Weight class 8 | 0.1 | 1.1 | 0.9 % | 30.7 % | 0.3 | – |
| Weight class 9 | 0.1 | 0.5 | 1.3 % | 16.8 % | 0.1 | – |
| Weight class 10 | 0.0 | 0.3 | 1.7 % | 21.6 % | 0.1 | – |

Table 4: Overview of unknown parameters in the model for the standing stock.

| Parameter | Parameter group | Eq. | Description |
|---|---|---|---|
| $\beta^f_{0,v}$ | Growth | (1) | Intercept growth, weight class $v$ |
| $\beta^f_1$, $\beta^f_2$ | Growth | (1) | Sea temperature driven growth |
| $\beta^f_3$, $\beta^f_4$ | Growth | (1) | Daylight hours driven growth |
| $\beta^f_5$, $\beta^f_6$ | Growth | (1) | Seasonal growth |
| $\beta^\gamma$ | Weight distribution within a weight class | | Intercept dispersal, weight class $v \geq 1$ |
| $\beta^\gamma_0$ | Weight distribution within weight class 0 | (5) | Intercept dispersal, $v = 0$ |
| $\beta^\gamma_1$, $\beta^\gamma_2$ | Distribution within weight class 0 | (5) | Seasonal dispersal, $v = 0$ |
| $\beta^I_0$ | Weight of stocked fish | (10) | Intercept weight |
| $\beta^I_1$, $\beta^I_2$ | Weight of stocked fish | (10) | Seasonal weight |

Table 5: Values for of the estimated parameters in the model for the standing stock for each of the three regions (confer with Table 4 for a description of the parameters). Some parameters are listed pairwise. For example, for Mid Norway, $\beta^f_{0,0} = -3.9$ and $\beta^f_{0,1} = -4.1$.

| | | Regions of Norway | | |
|---|---|---|---|---|
| Parameter | Parameter group | Southern | Mid | Northern |
| $\beta^f_{0,0}$, $\beta^f_{0,1}$ | Growth | –3.9, –3.8 | –3.9, –4.1 | –4.6, –5.2 |
| $\beta^f_{0,2}$, $\beta^f_{0,3}$ | Growth | –4.2, –4.5 | –4.4, –4.6 | –5.3, –5.6 |
| $\beta^f_{0,4}$, $\beta^f_{0,5}$ | Growth | –5.3, –4.8 | –5.2, –5.6 | –6.0, –6.1 |
| $\beta^f_1$, $\beta^f_2$ | Growth | 0.22, –0.0077 | 0.43, –0.019 | 0.51, –0.021 |
| $\beta^f_3$, $\beta^f_4$ | Growth | 0.16, –0.0062 | 0.040, –0.00092 | 0.15, –0.0049 |
| $\beta^f_5$, $\beta^f_6$ | Growth | –0.034, –0.19 | 0.024, –0.060 | 0.32, –0.026 |
| $\beta^\gamma$ | Weight distribution | –0.70 | –0.61 | –0.61 |
| $\beta^\gamma_0$ | Weight distribution | –0.79 | –0.65 | –0.81 |
| $\beta^\gamma_1$, $\beta^\gamma_2$ | Weight distribution | –0.11, 0.16 | –0.10, 0.079 | 0.0019, 0.22 |
| $\beta^I_0$ | Weight of stocked fish | –2.0 | –1.7 | –2.2 |
| $\beta^I_1$, $\beta^I_2$ | Weight of stocked fish | –0.10, –0.66 | –1.3, –0.55 | –2.4, 0.44 |