

Proportion of river- and sea-spawning whitefish in catches at the Åland Islands (Baltic Sea), estimated from gill raker counts

Andreas Lindén^{1)*‡}, Mikael Himberg²⁾, Mikael von Numers²⁾, Tom Wiklund³⁾, Christine Engblom³⁾, Piotr Hägerstrand⁴⁾, Jan-Olof Lill⁵⁾ and Henry Hägerstrand^{1)†3)‡}

¹⁾ *Novia University of Applied Sciences, Raseborgsvägen 9, FI-10600, Ekenäs, Finland*
(*corresponding author's e-mail: andreas.linden@iki.fi)

²⁾ *Environmental and Marine Biology, Faculty of Science and Engineering, Åbo Akademi University, BioCity, Artillerigatan 6, FI-20520 Åbo-Turku, Finland*

³⁾ *Laboratory of Aquatic Pathobiology, Environmental and Marine Biology, Faculty of Science and Engineering, Åbo Akademi University, BioCity, Artillerigatan 6, FI-20520 Åbo-Turku, Finland*

⁴⁾ *Skarpansvägen 19 B 17, AX-22100 Mariehamn, Åland Islands, Finland*

⁵⁾ *Accelerator Laboratory, Turku PET Centre, Åbo Akademi University, Porthansgatan 3, FI-20500, Åbo-Turku, Finland*

† Present address

‡ Contributed equally to this work

Received 19 Nov. 2018, final version received 26 Apr. 2019, accepted 21 Mar. 2019

Lindén A., Himberg M., von Numers M., Wiklund T., Engblom C., Hägerstrand P., Lill J.O. & Hägerstrand H. 2019: Proportion of river- and sea-spawning whitefish in catches at the Åland Islands (Baltic Sea), estimated from gill raker counts. *Boreal Env. Res.* 24: 101–113.

The Åland Islands are feeding grounds for river- and sea-spawning whitefish (*Coregonus lavaretus*), including also hatchery-reared released whitefish. Management of fisheries requires knowledge of whitefish origin from their catches, but such information is deficient at the Åland Islands. The numbers of gill rakers had unique (discretized) Gaussian frequency distributions in all three forms studied. This justified the usage of Gaussian finite mixture models on gill raker data for estimating the proportions of the forms in mixed catches during the non-spawning season. From 14 catches at the Åland Islands from 2012–2013, on average 54.8% were river-spawning whitefish, while the presence of hatchery-reared, sea-spawning whitefish was low but uncertain. Using a hierarchical variant of the model, we showed that the proportion of river-spawning whitefish decreased from 69.6% in June to 38.3% in August. Our study provides feasible methods for monitoring spatio-temporal variation in the proportions of river- and sea-spawning whitefish.

Introduction

Whitefish (*Coregonus lavaretus*) is an economically important fish species at the Gulf of Bothnia (Baltic Sea). The whitefish in this area consists of

two ecotypes: anadromous, river-spawning whitefish and sea-spawning whitefish (Himberg 1970, Lehtonen 1981). The two ecotypes coexist in the sea during non-spawning times and in areas containing both wild (i.e., naturally reproduc-

ing) and stocked whitefish. The river-spawning whitefish can undertake distant feeding migrations; for example between northern rivers and southern parts of the Gulf of Bothnia (Wikgren 1962, Lehtonen 1981, Lehtonen and Himberg 1992, Jokikokko and Huhmarniemi 2014). River-spawning whitefish abundantly occur at the Åland Islands and the adjacent Archipelago Sea, where they stay for several years (Lehtonen 1981). Following maturation, they return to their home rivers or stocking/imprinting sites to spawn. The sea-spawning whitefish spawn in coastal bays, e.g. at the eastern Archipelago Sea and at the Åland Islands. Due to the widely different ecology and behaviour of the two ecotypes, any possible management actions have to be tailored for the focal ecotype.

River-spawning Baltic Sea whitefish is presently listed as endangered according to the Baltic Marine Environment Protection Commission — Helsinki Commission (<http://www.helcom.fi/Red%20List%20Species%20Information%20Sheet/HELCOM%20Red%20List%20Coregonus%20albula.pdf>). At the Gulf of Bothnia, a drastic decline in whitefish catches has occurred since the mid-1900s. The reproduction of the river-spawning whitefish has diminished mainly due to anthropogenic activities. In most major rivers at the Gulf of Bothnia, dams built for hydropower stations have restricted whitefish migration to their spawning grounds (Tuikkala and Pirttijärvi 1976, Säisä *et al.* 2008, Larsson *et al.* 2013). Eutrophication, dredging and sand and gravel prospecting have destroyed spawning grounds, particularly for sea-spawning whitefish (Himberg 1995, Vanhatalo *et al.* 2012, Hudd *et al.* 2013, Veneranta *et al.* 2013). Furthermore, at the Åland Islands, there is a distinct whitefish form (of the sea-spawning ecotype), which possibly originates from hatchery-reared fish. The proportion of this form in fishery catches during the non-spawning season is still unknown. The strong whitefish fishery pressure and the threatened status of the species are important arguments for the discussion about fishery restrictions. For stock assessment of whitefish, there is an interest both from an economical point of view and for preservation reasons to estimate the proportion of different ecotypes of whitefish in fishery catches at different locations and seasons.

At the Finnish west coast, river-spawning whitefish have on average 29–31 gill rakers, while sea-spawning whitefish have on average 26–28 gill rakers (Himberg 1970, Himberg *et al.* 2015). This difference has previously been utilised to approximate, in different manners, the proportion of river- and sea-spawning whitefish in catches from the Åland Islands (Himberg 1978, Lehtonen 1981) and the Archipelago Sea (Lehtonen and Böhling 1988). The gill raker number is also used to assure the ecotype identity of stocks used in hatcheries (Jokikokko and Huhmarniemi 1998). Notably, the sea-spawning whitefish that nowadays spawn in coastal bays at the Åland Islands, have an average gill raker number below 26; thus being a distinct form, differing from sea-spawning whitefish spawning in the eastern part of the Archipelago Sea. We still lack a coherent picture of the mix of different forms and the spatio-temporal distribution of the two ecotypes at the Åland Islands.

In this study, we analysed the composition of whitefish forms found in mixed catches sampled in the summer, during non-spawning times, and at non-spawning sites. We applied finite Gaussian mixture models (McLachlan and Peel 2000) to data on gill raker count (GRC) frequencies to statistically estimate the proportions of three different whitefish forms in mixed catches at the Åland Islands. To establish the basis for the identification of these forms, we used existing GRC data from nominally pure populations (each representing a distinct form) of whitefish sampled at spawning sites during spawning time. These data, hereafter called baseline samples, included: river-spawning populations from rivers at the Finnish west coast, sea-spawning whitefish populations from the Archipelago Sea and sea-spawning populations from the Åland Islands.

First, we investigated the assumption that each of the baseline samples of GRC data were normally distributed, validating the further use of Gaussian mixture models. Second, we assessed the proportion of three different forms in the complete summer data set, and used information theoretical model selection to reveal the most parsimonious mix of forms. Third, we used a hierarchical model to study the spatio-temporal variation of the proportion of the river- and sea-spawning ecotypes. We hypothesized that the proportion of



Fig. 1. Whitefish sampling sites used in this study. Spawning river-spawners were sampled at: **A** – Tornio River; **B** – Kokemäki River; **C** – Aura River; **D** – Kisko River. Spawning Archipelago Sea sea-spawners were sampled at: **E** – Bengtsår; **F** – Airisto; **G** – Rihntniemi; **H** – Pyhämaa. The sampling sites **I** (insert) – Åland Islands – were used for sampling mixed catches outside the spawning season: **1** – Mariehamn (Kobba Klintar), **2** – Lågskär, **3** – Eckerö, **4** – Geta, **5** – Saltvik, **6** – Kumlinge, and for spawning Åland Island sea-spawners: **1** – Mariehamn (Slemmern Bay), **4** – Geta, and **5** – Saltvik (Tengsöda Bay).

river-spawning whitefish decrease throughout the summer, as the mature individuals start migrating back toward their spawning grounds.

Materials and methods

Study area

The Gulf of Bothnia (Baltic Sea) extends ~725 km in the north–south direction (Fig. 1). The salinity of the surface water decreases from about 7‰ in the south to about 1‰ in the north. The Åland

Islands is a province located in the southernmost part of the Gulf of Bothnia. The Åland Islands' eastern archipelago is contiguous with the Archipelago Sea, which stretches to the southwest coast of Finland. The Archipelago Sea consists of thousands of islands, skerries, reefs and shallow waters abundant with food for whitefish.

Catches used as baseline data

To establish the basis for the identification of forms, we used earlier-collected GRC data of

nominally pure river- and sea-spawning whitefish undergoing spawning in October–November. We restricted the baseline data to the most recent data available, and divided it into three groups (Table 1, Fig. 1): "River" — data on river-spawning whitefish ($n = 183$) from rivers along the Finnish west coast; "Åland" — data on sea-spawning whitefish ($n = 291$) sampled at coastal bays at the Åland Islands mainland, where fish from a local hatchery (Guttorp, Åland Islands) is annually stocked; "Archipelago" — data on sea-spawning whitefish ($n = 323$) sampled from coastal bays in the Archipelago Sea and adjacent areas at the Finnish west coast.

Mixed catches from the Åland Islands

Whitefish, of presumably several forms, were collected during non-spawning season in a total of 14 mixed catches ($n = 709$), from six sites around the Åland Islands, from June–August in 2012 and 2013 (Fig. 1, Table 2). The catches were divided into two spatial groups to facilitate comparison: north to east (NE) and south to west (SW) (Table 2). The sample size per catch was on average 50.6 individuals and the between-catch SD in sample size was 10.9 individuals.

The fish were caught with standing gill nets (45 mm knot distance, 1.8–3.0 m deep,

0.17 mm nylon tread diameter), which are the most common gear in whitefish fisheries at the Åland Islands and Archipelago Sea. Sampled whitefish were frozen at -20°C until examined.

Gill raker counting

Frozen whitefish were thawed and gills rinsed with tap water. The gills were cut out and mounted with pins on a backing. The number of gill rakers on the left outer-gill arch was counted with the aid of stereomicroscope at 5–10 times magnification. Minor rakers located at the ends of the arch were also counted.

Statistical analysis

Testing the normality of baseline samples

Although GRC data are discrete counts, the variable much resembles a continuous quantitative morphometric trait. GRC data have earlier been modelled applying Gaussian distributions and the choice has been justified with analyses of the empirical distributions of GRC in pure populations (Amundsen *et al.* 2004). To further justify the use of finite mixtures models with presumed distribution shapes for our data, we tested sepa-

Table 1. Summary of baseline data. The whitefish catches are divided into three groups according to their spawning behaviour (River – river-spawning; Åland – sea-spawning from the Åland Islands; Archipelago – sea-spawning from Archipelago Sea). For each catch, we reported the location (with associated symbols in Fig. 1), year, month, numbers of fish caught (n), sample mean and SD of the number of counted gill rakers.

Population	Location	Year	Month	n	Mean	SD
River	Torne River (A)	2013	10	47	30.19	1.84
	Kokemäki River (B)	2011	11	65	29.95	2.12
	Aura River (C)	1998	11	40	30.60	1.81
	Kisko River (D)	1994	11	31	29.84	2.08
Åland	Mariehamn-Slemmern Bay (I-1)	2013	11	57	25.67	2.87
	Geta (I-4)	2012	11	50	26.54	2.62
	Geta (I-4)	2013	11	36	26.42	2.12
	Saltvik-Tengsöda Bay (I-5)	2007	11	40	25.30	2.34
	Saltvik-Tengsöda Bay (I-5)	2012	11	52	24.65	2.20
	Saltvik-Tengsöda Bay (I-5)	2013	11	56	25.35	2.10
Archipelago	Bengtsår (E)	2009	11	29	27.90	1.90
	Airisto (F)	2012	10	50	27.60	2.54
	Rihtniemi (G)	2001	11	182	27.78	2.15
	Pyhämaa (H)	2012	10	62	27.60	2.66

rately for each of the baseline samples (whitefish forms) the null hypothesis that the GRC data were distributed according to a discretized normal distribution (Roy 2003). With a given mean (μ) and standard deviation (σ), this distribution is defined for all integer values, k , and is derived from the cumulative normal probability density function $\Phi(k | \mu, \sigma)$, as:

$$\text{DN}(k | \mu, \sigma) = \Phi(k + 0.5 | \mu, \sigma) - \Phi(k - 0.5 | \mu, \sigma). \quad (1)$$

For a given baseline sample (with sample size n), we first calculated the sample mean (\bar{x}) and standard deviation (α) of the GRC. We then determined the observed GRC frequencies of fish (Observed_k) for all integers (k) in the range $\bar{x} \pm 3\alpha$. Next, we determined the expected frequencies as:

$$\text{Expected}_k = n \text{DN}(k | \bar{x}, \alpha), \quad (2)$$

and calculated a χ^2 test statistic, to measure the amount of evidence against discrete normality, as:

$$\chi^2 = \sum (\text{Observed}_k - \text{Expected}_k) / \text{Expected}_k. \quad (3)$$

A suitable predefined range of studied integers — e.g., $2.5-3\alpha$ from the mean — should simultaneously avoid high rates of very small expected frequencies (causing noise in the test statistic; Eq. 3) and efficiently make use of data in the tails. In our case ($\pm 3\alpha$), 2 of 797 observations fell outside the range of integers.

To relax the assumption of a strictly χ^2 distributed test statistic under the null hypothesis, we assessed the statistical significance with a Monte Carlo test, where we repeatedly applied the procedure described above (Eq. 2 and Eq. 3) to 10 000 simulated data sets, generated under the null hypothesis of discrete normality. Each simulation event was done by first drawing n random numbers from a normal distribution with mean \bar{x} and standard deviation s_{sim} , then rounding them to the nearest integer. In the simulation we used:

$$s_{\text{sim}} = 2\alpha - \text{SD}_{\text{sim}}, \quad (4)$$

instead of α , because the rounding causes a slight bias on the standard deviation (Roy 2003). We

obtained SD_{sim} as the sample standard deviation of 1 000 000 rounded, random normal numbers with mean \bar{x} and standard deviation α . We assessed normality graphically in relation to the simulations, and using one-tailed tests, where the statistical significance is the proportion of simulated χ^2 test statistics being larger than the empirical one.

The finite mixture model

Mixture models are commonly applied in many fields of science to identify and analyse the presence and proportions of two or several subpopulations or clusters in a larger sample, without the need to identify the group membership of the individual observations. Finite Gaussian mixture models assume that the samples concerned are a mixture of a finite number of subpopulations, each characterized by a univariate or multivariate Gaussian distribution (McLachlan and Peel 2000). Here, we applied finite mixture models to univariate GRC data, assuming a discretized version of the normal distribution to match the discrete character of the data (Roy 2003; see Eq. 1).

For the full summer data set, we assumed that the evaluated GRC data, $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$ are independent observations from a mixture of three normal distributions having different means

Table 2. Summary of mixed catches. The year, month and numbers of whitefish caught (n), spatial grouping (Area: SW – south to west, NE – north to east) and the location (with associated symbols in Fig. 1) were reported.

Location	Year	Month	n	Area
Mariehamn-Kobba Klintar (I-1)	2012	6	34	SW
		7	61	SW
		8	34	SW
	2013	6	73	SW
		8	50	SW
Lågskår (I-2)	2013	7	50	SW
Eckerö (I-3)	2013	6	50	SW
		7	50	SW
Geta (I-4)	2013	6	53	NE
		7	52	NE
Saltvik-Tengsöda Bay (I-5)	2013	6	51	NE
		8	50	NE
Kumlinge (I-6)	2013	6	37	NE
		8	64	NE

(μ_1, μ_2, μ_3) and standard deviations $(\sigma_1, \sigma_2, \sigma_3)$, representing the "River", "Åland" and "Archipelago" forms, respectively. The contributions of the forms are described using two parameters: p_1 = the proportion of the river-spawning ecotype; and p_3 = the proportion of the Archipelago form among all sea-spawning whitefish $(1 - p_1)$. Given a vector of parameters $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, p_1, p_3)$, the marginal probability of each data point (i) in the mixed catch data is defined as:

$$P(x_i | \theta) = p_1 \text{DN}(x_i | \mu_1, \sigma_1) + (1 - p_1)(1 - p_3) \text{DN}(x_i | \mu_2, \sigma_2) + (1 - p_1)p_3 \text{DN}(x_i | \mu_3, \sigma_3). \quad (5)$$

The baseline data sets with presumably known origins are simply thought to be distributed according to the discrete normal distribution specified for each particular form: **Br** = river-spawning; **Bs** = sea-spawning at Åland; and **Ba** = sea-spawning at the Archipelago Sea:

$$P(\mathbf{Br} | \theta) = \text{DN}(\mathbf{Br} | \mu_1, \sigma_1) \quad (6)$$

$$P(\mathbf{Bs} | \theta) = \text{DN}(\mathbf{Bs} | \mu_2, \sigma_2) \quad (7)$$

$$P(\mathbf{Ba} | \theta) = \text{DN}(\mathbf{Ba} | \mu_3, \sigma_3) \quad (8)$$

The model was fitted with a direct maximum likelihood search by searching the combination of parameters (θ), which minimizes the negative log-likelihood function:

$$-\ln L(\theta) = -\sum \ln P(x | \theta) - \sum \ln P(\mathbf{Br} | \theta) - \sum \ln P(\mathbf{Bs} | \theta) - \sum \ln P(\mathbf{Ba} | \theta) \quad (9)$$

In practice, we applied the Nelder–Mead simplex algorithm for multidimensional optimization, using the function, *optim()*, in the programming environment R ver. 3.4.4 (R Core Team 2018). The numerical search over the means and standard deviations was applied without transformations, but for the parameters p_1 and p_3 – both constrained to the interval $\{0, 1\}$, the search was done on the logit-scale (i.e., the parameters were back-transformed when evaluated, at each iteration). The maximum number of iterations was set to 9999 and we applied a relative tolerance of 10^{-20} to define convergence. In this type of numerical search, we have to define a set of starting values for the parameters. For parameters μ and σ , we set the starting values to the sample means and

sample standard deviations, respectively, from each baseline data set. For the parameters p_1 and p_3 we used the starting value 0.5 (i.e., $\text{logit}(p_1) = 0$ and $\text{logit}(p_3) = 0$).

We assessed the precision of obtained sets of parameter estimates $\theta_{\text{estimated}}$ by simulating their sampling distribution using parametric bootstrapping, with $B = 2000$ repetitions. In other words, we randomized all the stochastic components of the models, given the estimated parameters and their presumed distributions, to simulate B resampled data sets, to each of which we refitted the same model. From the B simulated sets of potential parameter estimates, and under the scenario of $\theta_{\text{estimated}}$ being true, we calculated the standard errors (SE) by taking the sample standard deviation of the resampled estimates. Further, we calculated for all parameters, 95% confidence intervals (CI), by taking for each parameter the 2.5th and 97.5th percentiles of the resampled estimates. Under repeated resampling, the CI can be expected to contain the true parameter value in 95% of the cases.

An R-function was written to perform the analyses described above. The code performs the full model fitting procedure for one sample, which is a mixture of one, two or three subpopulations. The code also has an option for parametric bootstrapping.

Model selection

For detailed analysis of the proportions of whitefish forms based on GRC data in mixed samples, (e.g., variation in time and space), we needed to choose a relevant and parsimonious composition of the three potential forms to be considered in the mixed samples. All candidate forms are not necessarily identifiable using our model, or even present in relevant numbers in the mixed samples, so we can see this step as a data compression problem. Using information theoretic model selection, we evaluated the support of seven candidate models, representing different mixtures of the three forms. Models were evaluated and ranked with regard to their parsimony, according to the Akaike Information Criterion (AIC), which is a statistic derived from the Kullback–Leibler divergence (see Burnham and Anderson 2002).

All candidate models were special cases of the most complex model #7 (see Table 3), which allows a mixture of all three forms (river spawning, Archipelago sea- and Åland sea-spawning whitefish). The full model was described previously in the "The finite mixture model" section, under "Materials and methods", and its parameters and CI were evaluated regardless of the model selection result. In other words, the results from model #7 were only to view the general role and presence of all the different whitefish forms.

The models were compared in terms of differences in AIC (hereafter ΔAIC); i.e., units of AIC from the best-ranked model (with the lowest AIC). Models with $\Delta\text{AIC} < 3$ were considered competitive and relevant to consider. An exception occurred if any model with $\Delta\text{AIC} < 3$ was essentially the highest ranked model and in addition, an extra, so-called uninformative parameter, which does not improve the model (Arnold 2010). Such models were disregarded because the largest possible drop in AIC when adding one more parameter is 2.

A spatio-temporal extension of the model

For assessing spatio-temporal variation in the proportion of river-spawning whitefish (q_j) between the 14 mixed samples (each denoted with index j), we developed a hierarchical version of the model described above, for two forms, river- and Archipelago sea-spawning, instead of three (or in other words, p_3 fixed to one). Depending on the results from the previous analysis (model selection) and the estimate of p_3 , the composition of the sea-spawning baseline sample (Åland and Archipelago whitefish) can be adjusted, if necessary. Assuming a logit-linear relationship, the proportion of the river-spawning ecotype was modelled using numbers indicating the summer month (June = 0, July = 1; August = 2) and spatial grouping (Area; NE = 0, SW = 1) as explanatory variables, as:

$$\text{logit}(q_j) = a + b_1 \text{Month}_j + b_2 \text{Area}_j + \varepsilon_j, \quad (10)$$

$$\varepsilon_j \sim \text{Normal}(0, \sigma_{\text{smp}}). \quad (11)$$

Here, a is the intercept term, b_1 is the coefficient describing the monthly change, and b_2

is the coefficient describing the spatial contrast along the NE–SW-axis. The unexplained spatio-temporal variation in the proportion (q_j) is modelled as a random effect (ε_j) of the mixed sample ID, being normally distributed with zero mean and standard deviation σ_{smp} . Given the parameter q_j (which is defined by the higher level parameters of Eq. 10 and Eq. 11, $\boldsymbol{\beta} = (a, b_1, b_2, \sigma_{\text{smp}})$) and the lower level parameters $\boldsymbol{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2)$, the marginal probability distribution of each data point (i) from mixed sample (j) is:

$$P(x_{ij} | q_j, \boldsymbol{\theta}) = q_j \text{DN}(x_{ij} | \mu_1, \sigma_1) + (1 - q_j) \text{DN}(x_{ij} | \mu_2, \sigma_2). \quad (12)$$

The negative log-likelihood function is obtained by integrating over the random effects (ε) from minus infinity to plus infinity (for the mixed samples), and similarly to Eq. 9, combining with the baseline sample likelihoods as:

$$-\ln L(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\sum \ln \int P(\mathbf{x} | \varepsilon, \boldsymbol{\beta}, \boldsymbol{\theta}) P(\varepsilon | \boldsymbol{\beta}) d\varepsilon - \sum \ln P(\mathbf{B}\mathbf{r} | \boldsymbol{\theta}) - \sum \ln P(\mathbf{B}\mathbf{s} | \boldsymbol{\theta}). \quad (13)$$

Table 3. Description and model selection results of the seven models evaluated for the complete summer data set using Akaike Information Criterion. The criterion was used to determine the most relevant mix of three populations (River – river-spawning; Åland – sea-spawning from the Åland Islands; Archipelago – sea-spawning from Archipelago Sea). All models are special cases of model #7 – the mix of three populations. K is the number of estimated parameters. For p_1 and p_3 , the values are either estimated (+), fixed to a specific number (0 or 1), or redundant, but in practice fixed to 0 (–). ΔAIC is the difference in Akaike information criterion of the fitted model compared with the most parsimonious one (model #5).

Model	Description	K	p_1	p_3	ΔAIC
#1	Only River	6	1	–	37.80
#2	Only Åland	6	0	0	334.30
#3	Only Archipelago	6	0	1	61.47
#4	Mixture of River and Åland	7	+	0	19.74
#5	Mixture of River and Archipelago	7	+	1	0
#6	Mixture of Åland and Archipelago	7	0	+	63.47
#7	Mixture of all three populations	8	+	+	2.0

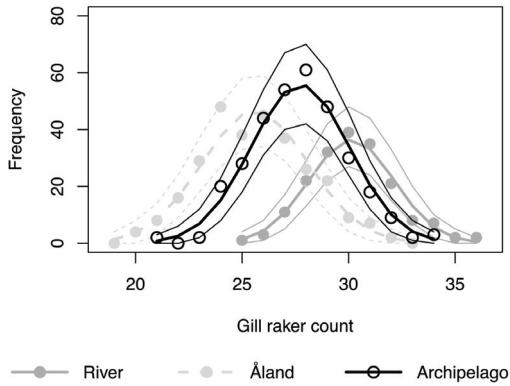


Fig. 2. In all three baseline samples, the observed gill raker count frequencies (filled and open circles), are well in accordance with the assumption of discretized normal distributions (expectation with thick lines; 95 % confidence bands with thin lines).

Here, $P(\mathbf{x} \mid \varepsilon, \boldsymbol{\beta}, \boldsymbol{\theta})$ is a short notation for applying Eq. 12, and the relationship between ε_j and q_j described in Eq. 10 to vector \mathbf{x} , while $P(\varepsilon \mid \boldsymbol{\beta})$ refers to the normal distribution presented in Eq. 11. For numerical integration, we used adaptive Gauss–Hermite quadrature, with 25 node positions, applying the R-package fastGHQuad (Blocker 2014).

For model fitting, we again used the Nelder–Mead simplex algorithm for multidimensional optimization, now over parameters $\boldsymbol{\beta} = (a, b_1, b_2, \sigma_{\text{smp}})$ and $\boldsymbol{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2)$, with the target to minimize the expression in Eq. 13.

The maximum number of iterations was set to 9999 and the relative tolerance was set to 10^{-12} . Finally, after having estimated the parameters ($\boldsymbol{\beta}_{\text{estimated}}, \boldsymbol{\theta}_{\text{estimated}}$), we estimated the random effect of each mixed sample (ε_j) separately and hence, also the corresponding proportions of river-spawning whitefish (q_j), by minimizing the expression:

$$-\ln L(\varepsilon_j) = -\sum \ln [P(x_{ij} \mid \varepsilon_j, \boldsymbol{\beta}_{\text{estimated}}, \boldsymbol{\theta}_{\text{estimated}}) \times P(\varepsilon_j \mid \boldsymbol{\beta}_{\text{estimated}})] \tag{14}$$

Again, the sampling distribution of the parameters was simulated using parametric bootstrapping and $B = 2000$ resampling events.

Results

Normality of the number of gill rakers

All investigated baseline samples appeared to be close to a discretized normal distribution (parameter estimates shown in Table 4), with deviations within the range of what can be expected from randomness (Fig. 2). According to the Monte Carlo χ^2 -tests, we could not reject the null hypothesis of discrete normality for any of the baseline samples studied (River: $\chi^2 = 8.03$, mean simulated $\chi^2 = 8.86$, p -value = 0.520; Åland: $\chi^2 = 8.87$, mean simulated $\chi^2 = 11.96$, p -value = 0.715; Archipelago: $\chi^2 = 13.93$, mean simulated $\chi^2 = 10.96$, p -value = 0.231).

Table 4. Parameters and their estimates according to the most parsimonious model (#5), where the pooled summer data (June–August; $n = 709$) represent the mixed catch. Uncertainties of the estimates are shown as standard errors (SE) and 95% confidence intervals ($CI_{\text{low}}, CI_{\text{high}}$), based on a parametric bootstrapping procedure with 2000 repetitions. For comparison, raw sample means and SDs from the baseline data sets are also shown (Baseline).

Notation	Explanation	Baseline	Estimate	SE	CI_{low}	CI_{high}
μ_1	Mean, river	30.14	30.08	0.132	29.82	30.35
σ_1	SD, river	1.98	2.04	0.077	1.88	2.18
μ_2	Mean, Åland	25.62	25.62	0.142	25.33	25.90
σ_2	SD, Åland	2.48	2.46	0.105	2.25	2.67
μ_3	Mean, Archipelago	27.73	27.77	0.120	27.54	28.01
σ_3	SD, Archipelago	2.29	2.24	0.075	2.09	2.38
p_1	Proportion river vs. sea	–	0.55	0.056	0.44	0.67
p_3	Proportion Åland vs. Archipelago	–	1.00	–	–	–

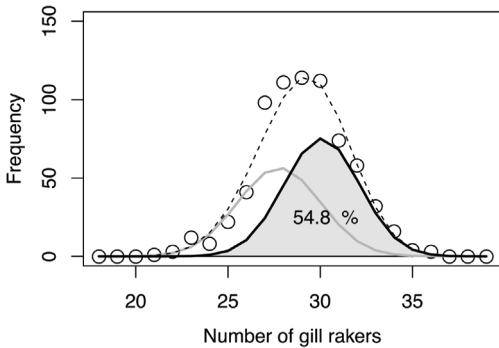


Fig. 3. Observed frequencies of whitefish gill raker counts from June–August (open circles), the fitted contributions of river-spawning fish (shaded area under black solid line), fish spawning in the Archipelago Sea (grey solid line) and the mixture distribution (black dashed line). The estimated proportion of river-spawning whitefish is 54.8%.

Model selection and estimated proportions in the pooled summer data

In the complete summer data set ($n = 709$), model #5 — a mixture of river-spawning and sea-spawning whitefish of the Archipelago type — was the most parsimonious model in terms of AIC (Table 3). According to this model, the estimated proportion of river-spawning whitefish (p_1) was 54.8% (CI = 43.5–65.6%) (Fig. 3; Table 4). The final estimated parameters (μ_1, μ_2, μ_3 , and $\sigma_1, \sigma_2, \sigma_3$) describing the distributions of gill rakers in the baseline sample, differed fairly little from the raw point estimates calculated directly from the samples (Table 4). Hence, as expected, these parameters were dominated by the baseline samples.

Superficially, the second best model was model #7 — the scenario with all three forms present in the mixed catches — where the AIC-score was ca. 2 units higher. Since estimation of the additional parameter p_3 (compared to model #5, where p_3 was fixed to 1) made the model less parsimonious, and the largest possible difference was $\Delta\text{AIC} = 2$, model #7 was better not regarded as competitive (Arnold 2010). We consulted model #7 only to view the estimated proportion of Archipelago type whitefish among sea-spawning ones (p_3), which was 99.6%

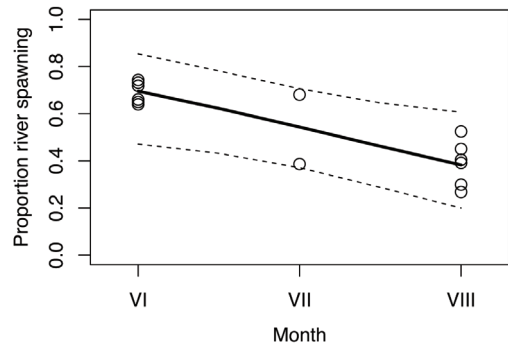


Fig. 4. The hierarchical version of the Gaussian mixture model, with temporal variation in the proportions of the 14 mixed samples. A decreasing trend in the share of river-spawning whitefish throughout the summer months was observed (VI–VIII). The thick black line is the model prediction, while the dashed lines are 95% CIs of the prediction. The open circles are estimated proportions of each mixed sample, combining the fixed and random effects.

(CI = 85.1–100.0%). Hence, the Åland type whitefish was estimated to be virtually absent from the samples, but in the light of the CI, its representation might be up to 15% of the sea-spawning part.

None of the other candidate models had any relevant support in relation to these two models. The third lowest AIC-score was model #4, i.e., a mixture of river-spawning and Åland-type sea-spawning whitefish, with $\Delta\text{AIC} = 19.74$ (Table 3). We therefore restricted all further investigations to only the mixture represented in model #5.

Spatio-temporal comparison of the proportion of river-spawning whitefish

For the mixed catches collected in summer, we fitted a hierarchical model to reveal any temporal or spatial trends in the proportion of river-spawning whitefish. As baseline data for the sea-spawning whitefish, we used GRC data from the Archipelago Sea alone (see the results from the previous section). The proportion of river-spawning whitefish decreased throughout the summer (estimated effect of month: $b_1 = -0.669 \pm 0.260$ SE; 95% CI = $-1.252, -0.242$), but the spatial differ-

ence between the areas did not differ from zero (estimated effect of SW: $b_2 = 0.702 \pm 0.453$ SE; 95% CI = $-0.136, 1.629$).

Given only 14 mixed samples and minimizing the risk for overparameterization, we refitted the model without the spatial parameter (setting $b_2 = 0$). The temporal effect was still clear (estimate of $b_1 = -0.652 \pm 0.277$ SE; 95% CI = $-1.265, -0.201$); Table 5), showing a decrease in the proportion of river-spawning whitefish, ranging from 69.6% in June (CI = 47.1–85.4%) to 38.3% in August (CI = 20.0–60.6%) (Fig. 4). All model parameters are presented in Table 5.

Discussion

We studied the composition of whitefish in mixed catches at the Åland Islands, applying finite mixture models on GRC data. The method assumed that fish with a particular origin shows (discretized) normally distributed numbers of gill rakers. Data sets from the nominally pure populations of river-spawning whitefish and the two forms of sea-spawning whitefish were used as baseline data to inform the model about the parameters of these distributions. Our results confirmed for all three baseline samples that GRC data can be successfully modelled using discretized normal distributions, validating the model assumptions. Hence, similar to Amundsen *et al.* (2004), our study provides support for approximating the frequency distribution of GRC data for whitefish populations with normal

distributions, or as in our study, their discrete counterparts.

The analyses of mixed catches showed that the river-spawning ecotype of whitefish, as one group, and sea-spawning whitefish from the Archipelago Sea, as another group, together dominated the mixed catches. The model, assuming a mix of these two forms, and disregarding the sea-spawning whitefish from the Åland Islands as the third group, was the most parsimonious, suggesting that the model is a simple and robust candidate for further analysis on the composition of mixed catches. For model #7, where the sea-spawning Åland Island form was also included, its estimated proportion among all sea-spawning whitefish was only 0.4%. In the light of the 95% confidence interval, the proportion of mixed catches was $< 15\%$ and not significantly differing from zero. Hence, assuming no obvious pattern of selectivity in the sampling, we conclude that the stocked sea-spawning whitefish originating from the local hatchery contributed very little to the whitefish pool at the Åland Islands. Alternatively, the Åland form may differ in its behaviour (e.g., migration), being less likely captured in the mixed catches compared with the two other groups.

River-spawning whitefish dominated the mixed fishery catches by an overall estimated proportion of ca. 55%. Previous samplings with similar fishing gears resulted in similar estimated proportions (40–60%) of river-spawning whitefish at the Åland Islands (Lehtonen and Himberg 1979, Lehtonen 1981, Himberg *et al.* 2015) and in the Archipelago Sea (Lehtonen and Böhling

Table 5. Parameters and their estimates in the hierarchical mixture model, which allows for temporal variation in the proportion of river-spawning whitefish. The spatial coefficient (b_2) was fixed to zero, as it did not differ from zero. Parameter estimates are associated with their standard errors (SE) and 95% confidence intervals (CI_{low}, CI_{high}). These are based on a parametric bootstrapping procedure with 2000 repetitions.

Notation	Explanation	Estimate	SE	CI _{low}	CI _{high}
a	Intercept	0.827	0.432	0.105	1.850
b_1	Coefficient for temporal change	-0.652	0.277	-1.265	-0.201
b_2	Coefficient for spatial change	0	—	—	—
σ_{smpl}	Between-sample SD	0.573	0.314	0	1.079
μ_1	Mean of GRC, river	30.10	0.132	29.85	30.35
σ_1	SD of GRC, river	2.04	0.079	1.88	2.19
μ_2	Mean of GRC, sea (Archipelago)	27.78	0.119	27.55	28.01
σ_2	SD of GRC, sea (Archipelago)	2.24	0.074	2.10	2.37

1988). This makes sense, as the waters around Åland Islands and in the Archipelago Sea hold a high production of food (Lehtonen 1981, Himberg 1995, Himberg *et al.* 2015). A decrease in the proportion of river-spawners was observed in the catches from June to August. Presuming no changes in the selectivity of the two ecotypes in the catches, this decrease is probably due to spawning migration toward rivers of the recently matured river-spawning whitefish pool (Lehtonen 1981, Himberg *et al.* 2015).

Our study provides no evidence for a consistent difference in the proportion of river-spawners in the south–west compared with the north–east of the Åland Islands during the summer months. The south–west parts of Åland are beneficial feeding grounds and seem to be at least as important for river-spawning whitefish as the north–east parts of Åland, despite the longer distance from presumed spawning grounds. For example, the fairly high estimated proportion of river-spawning whitefish in catches from the southernmost sampling site of Lågskär (see Fig. 1), extends the known southern latitude of feeding migration habitats for river-spawning whitefish at the Åland Islands and shows that river-spawning whitefish can pass considerable deep open sea distances.

Notably, a considerable part of the river-spawning whitefish foraging at the Åland Islands is stocked, as indicated by a spot-check sample (Lill *et al.* 2015, Hägerstrand *et al.* 2015). In order to strengthen the whitefish stocks and increase fishery catches, several million one-summer-old whitefish of the river-spawning type are stocked at rivers along the Finnish west coast. River whitefish from the Kokemäenjoki river population are also stocked at the inner-part of the Archipelago Sea. The use of Kokemäenjoki whitefish is beneficial because it is the fastest growing form of river whitefish in the Bothnian Bay region (Lehtonen 1981).

Also, sea-spawning whitefish are stocked in the Archipelago Sea and in the waters around the Åland Islands to strengthen local sea-spawning stocks. The stocked sea-spawning whitefish in the Archipelago Sea mainly originates from Bengtsår (Hanko, Uusimaa). The stocked sea-spawning whitefish in Åland is largely raised in a hatchery in Guttorp (Åland Islands); with these

fish being originally caught from local populations (Geta and Saltvik, Åland). As shown in this study, the Guttorp whitefish, which has a characteristic distribution of gill rakers, make up an uncertain part in the mixed catches at the Åland Islands. However, the proportion is quite small, with total catches from 0–15% of the sea-spawning ecotype. The stocking of river whitefish into the sea should be avoided, due to the risk of hybridization between the two ecotypes.

While knowing the proportion of the two ecotypes is important for successful fishery management, reliable management actions simultaneously need to account also for the total numbers of fish present at different times and sites — or at least of relative abundance, e.g., measured as catch per unit effort. This component, combined with an estimated proportion of the ecotypes, contains similar information as time series of relative abundance of both ecotypes. In this study, where we consider only summer data from a three-month period and far outside spawning time, there are large numbers of fish; and temporal changes in the proportion alone may as such provide guidelines for the situation in early and late summer.

Using Gaussian mixture models on GRC data is a practical approach for approximating the proportion of river- and sea-spawning whitefish in catches at the Åland Island. It is a simple and robust tool for local implementation. We recommend that statistically large samples (preferably > 200) are analysed from each sampling area and season and that each sample contains fish from several locations (e.g., at least 5 clusters). A more laborious and expensive method, combining GRC with genetic data (Ozerov *et al.* 2015, 2016), may be less practical for immediate use by local authorities. Even for these types of data sets, hierarchical mixture models can be useful for analysing spatio-temporal variation in the proportion of ecotypes.

Conclusion

Our study sheds knowledge on the occurrence of whitefish forms in summer catches at the Åland Islands. From the conservation perspective, the clearly decreasing proportion of the

river-spawning ecotype throughout the summer, suggests that any fishing restrictions targeting river-spawning whitefish should focus on the early summer months (and perhaps the spring months). However, considering the fairly low fishing efforts and small numbers of whitefish caught at the Åland Islands during the spring and summer; and that only a part of the river-spawning whitefish is naturally born, the effectiveness of such measures is questionable. This result is conditional based on the fact that the proportions in the catches reflect the actual whitefish pool without selectivity patterns. Further data on total numbers are needed to consolidate the patterns. Naturally, the successful conservation of the river-spawning ecotype also requires reconstruction of deteriorated spawning rivers. Conservation actions targeting the local sea-spawning whitefish population, in turn, should focus on fishing restrictions at spawning time (15 October–15 November) in spawning areas. Furthermore, selling whitefish roe could be prohibited. The models presented here provide tools for estimating the proportions of river- and sea-spawning whitefish in fishery catches at the Åland Islands and facilitates authorities to follow spatio-temporal frequency distribution alterations. Together with information on the total number of whitefish around, this type of information is relevant for taking evidence-based actions for regulating the fishing pressure on either whitefish ecotype. Better knowledge on the population dynamics, larger sample sizes with more recent baseline data, as well as accounting for possible temporal trends in baseline distribution parameters would likely improve the insights on the whitefish populations composing the whitefish catch at the Åland Islands. The feeding behavior and migration of the two sympatric ecotypes in the sea should be further studied. Other important questions from a conservation point of view are to investigate whether stocked river whitefish spawn in the sea at Åland and whether they hybridize with local sea spawners.

Acknowledgements: The work was funded jointly by the Åland Provincial Government and the European Union (European fisheries fund – EFF; European Maritime and Fisheries Fund – EMFF), through two project grants: ÅLR 2012/#7681 at Åbo Akademi University and ÅLR 2018/#2955 at Novia University of Applied Sciences. We

thank the local fishermen at the Åland Islands for their cooperation in collecting whitefish samples and Husö Biological Station for offering excellent laboratory facilities.

References

- Amundsen P.A., Bøhn T. & Våga G.H. 2004. Gill raker morphology and feeding ecology of two sympatric morphs of European whitefish (*Coregonus lavaretus*). *Ann. Zool. Fennici* 41: 291–300.
- Arnold T.W. 2010. Uninformative Parameters and Model Selection Using Akaike's Information Criterion. *J. Wildl. Manag.* 74: 1175–1178.
- Blocker A.W. 2014. *fastGHQuad: Fast Rcpp implementation of Gauss-Hermite quadrature*. R package ver. 0.2. available at <https://CRAN.R-project.org/package=fastGHQuad>
- Burnham K.P. & Anderson D.R. 2002. *Model selection and multimodel inference: a practical approach*. Springer, New York.
- Himberg M. 1970. A systematic and zoogeographic study of some North European Coregonids. In: Lindsey C. & Woods C. (eds), *Biology of Coregonid Fishes*, Manitoba University Press, Winnipeg, pp. 219–250.
- Himberg M. 1995. Sikens biologi och lekplatser i Skärgårds och Bottenhavet [Whitefish biology and spawning locations in the Archipelago Sea and Gulf of Bothnia]. Maa- ja Metsätalousministeriö. *Finnish Game and Fisheries Research Report* 16: 2–21. [In Swedish].
- Himberg M., Numers M., Vasemägi A., Wiklund T., Lill J.O. & Hägerstrand H. 2015. Gill raker counting for approximating the ratio of river- and sea-spawning whitefish (*Coregonus lavaretus* L.) in the Gulf of Bothnia, Baltic Sea. *Acta Ichthyol. Pisc.* 45: 125–131.
- Hudd R., Veneranta L. & Vanhatalo J. 2013. Havslekande sikens och sikløjans yngelproduktionsområden [The reproduction areas of sea-spawning whitefish and vendace]. *Finnish Game and Fisheries Research Report* 7. [In Swedish].
- Hägerstrand H., Heimbrand Y., Jokikokko E., Florin A.B., Himberg M., Wiklund T., Slotte J.M.K. & Lill J.O. 2015. Life history inhomogeneity in Baltic Sea whitefish populations revealed by otolith strontium signatures – identification of stocked fish. *Front. Mar. Sci.*, doi: 10.3389/conf.FMARS.2015.03.00113.
- Jokikokko E. & Huhmarniemi A. 1998. Stocking practices of anadromous whitefish, *Coregonus lavaretus* lavaretus, in Bothnian Bay, Finland; evidence from gillraker numbers. *Adv. Limnol.* 50: 507–515.
- Jokikokko E. & Huhmarniemi A. 2014. The large-scale stocking of young anadromous whitefish (*Coregonus lavaretus*) and corresponding catches of returning spawners in the River Tornionjoki, northern Baltic Sea. *Fisheries Manag. Ecol.* 21: 250–258.
- Larsson S., Byström P., Berglund J., Carlsson U., Veneranta L., Larsson S.H. & Hudd R. 2013. Characteristics of anadromous whitefish (*Coregonus lavaretus* (L.)) rivers in the Gulf of Bothnia. *Adv. Limnol.* 64: 189–201.

- Lehtonen H. 1981. Biology and stock assessments of Coregonids at the Baltic coast of Finland. *Finn. Fish. Res.* 3: 31–83.
- Lehtonen H. & Böhling P. 1988. Management of the whitefish (*Coregonus lavaretus* L. s.l.) fishery in the Gulf of Bothnia. *Finn. Fish. Res.* 9: 373–387.
- Lehtonen H. & Himberg M. 1979. Sikbestånden och sikfångster vid våra kuster [Whitefish stocks and catches at our coasts]. *Fiskeritidskrift för Finland* 4: 68–72. [In Swedish].
- Lehtonen H. & Himberg M. 1992. Baltic Sea migration pattern of anadromous, *Coregonus lavaretus* L, S. Str and sea-spawning European whitefish, *C.L. Widegreni* Malmgren. *Pol. Arch. Hydrobiol.* 39: 463–472.
- Lill J.O., Heimbrand Y., Slotte J., Himberg M., Florin, A.B. & Hägerstrand H. 2015. PIXE analyses of polished otoliths for identification of anadromous whitefish in the Baltic Sea. *Nucl. Instr. Meth. Phys. Res. B* 363: 66–69.
- McLachlan G. & Peel D. 2000. *Finite mixture models*. John Wiley & Sons, Inc., New York.
- Ozerov M.Y., Himberg M., Aykanat T., Sendek D.S., Hägerstrand H., Verliin A., Krause T., Olsson J., Primmer C.R. & Vasemägi A. 2015. Generation of neutral FST baseline for testing local adaptation on gill raker number within and between European whitefish ecotypes in the Baltic Sea basin. *J. Evol. Biol.* 28: 1170–1183.
- Ozerov M., Himberg M., Debes P., Hägerstrand H. & Vasemägi A. 2016. Combining genetic markers with an adaptive meristic trait improves performance of mixed-stock analysis in Baltic whitefish. *ICES J. Mar. Sci.* [In Press], doi: 10.1093/icesjms/fsw122
- R Core Team 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org/>
- Roy D. 2003. The discrete normal distribution. *Commun. Stat. Theory Methods* 32: 1871–1883.
- Säisä M., Rönn J., Aho T., Björklund M., Pasanen P. & Koljonen M.L. 2008. Genetic differentiation among European whitefish ecotypes based on microsatellite data. *Hereditas* 145: 69–83.
- Tuikkala A. & Pirttijärvi J. 1976. Storsik och fiskevård i Södra Botten viken [Great whitefish and fishery management in Southern Bothnian Bay]. *Finlands fiskeritidskrift*. 20 (ny serie) 6: 125–133. [In Swedish].
- Vanhatalo J., Veneranta L. & Hudd R. 2012. Species distribution modeling with Gaussian processes: A case study with the youngest stages of sea spawning whitefish (*Coregonus lavaretus* L. s.l.) larvae. *Ecol. Model.* 228: 49–58.
- Veneranta L., Hudd R. & Vanhatalo J. 2013. Reproduction areas of sea-spawning coregonids reflect the environment in shallow coastal waters. *Mar. Ecol. Prog. Ser.* 477: 231–250.
- Wikgren B.J. 1962. Resultaten av sikmärkningarna inom Åland och vid Luvia [Results from whitefish tagging at the Åland Islands and Luvia]. Åbo Akademi-Ålands landskapsstyrelse. *Husö Biologiska Station Meddelanden* 3: 1–26. [In Swedish].

Appendix

The authors are happy to share the gill raker count data (CSV-file) and R-code (functions and script) applied in this study. If you wish to view them, please contact the corresponding author. The R-code used in our analyses include three readily applicable functions: first, for testing the null hypothesis of discrete normality; second, for estimating the proportions of, at most, three ecotypes in the case of one mixed-catch; and third, for the hierarchical model, which can be used for analysing spatial and temporal patterns in the proportion of ecotypes.