International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

-

Interim Report IR-03-10

Age at Maturation Predicted from Routine Scale Measurements in Norwegian Spring-Spawning Herring (*Clupea harengus*) Using Discriminant and Neural Network Analysis

Georg H. Engelhard (Georg.Engelhard@imr.no)
Ulf Dieckmann (dieckman@iiasa.ac.at)
Olav Rune Godø (Olav.Rune.Godoe@imr.no)

Approved by

Leen Hordijk (hordijk@iiasa.ac.at)
Director

March 2003

Contents

Abstract

We evaluate two methods allowing the prediction of age at maturation from the widths of annual growth layers in scales (or otoliths) in a case study on Norwegian spring-spawning herring. For this stock, scale measurements have been collected routinely for many decades. We compare the performance in classifying age at maturation (at 3, 4, ..., 9 years) between conventional discriminant analysis (DA) and the new methodology of artificial neural networks trained by backpropagation (NN) against a 'control' comprising historical estimates of age at maturation obtained by visual examination of scales. Both methods show encouraging, and about equally high classification success. The marginal differences in performance are in favour of DA, if the proportion of correctly classified cases is used as criterion (DA 68.0%, NN 66.6%), but in favour of NN if other criteria are used including the prediction errors (error >1 year: DA 5.2%, NN 2.9%) and the average degree of under- or overestimation (DA: underestimation 1.1% of mean; NN: overestimation 0.2% of mean). We provide evidence that both methods approach the *a priori* limits to maximal classification success, caused by overlapping combinations of predictor variables between maturation groups. These methods will allow studies on age at maturation in this important fish stock over a very long time-span including periods well before, during, and after its collapse to commercial extinction. Similar techniques might well be applicable to any other fish stock with long-term data on scale or otolith growth layers available.

About the Authors

Georg H. Engelhard
Institute of Marine Research
PO Box 1870 Nordnes
N-5817 Bergen, Norway

Ulf Dieckmann
Adaptive Dynamics Network
International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

Olav Rune Godø
Institute of Marine Research
PO Box 1870 Nordnes
N-5817 Bergen, Norway

Acknowledgments

We thank Sébastien Barot, Mikko Heino, Jens Christian Holst, Ingolf Røttingen, Aril Slotte, and Reidar Toresen for valuable discussions during this study. We acknowledge Oddvar Dahl, Annlaug Haugsdal, and Jan Henrik Nilsen for providing GHE extensive information on various aspects of the analysis of herring scales. The study was supported by the European Research Training Network *ModLife* (Modern Life-History Theory and its Application to the Management of Natural Resources), funded through the Improving Human Potential Programme of the European Commission (Contract Number HPRN-CT-2000-00051).

Age at Maturation Predicted from Routine Scale Measurements in Norwegian Spring-Spawning Herring (*Clupea harengus*) Using Discriminant and Neural Network Analysis

Georg H. Engelhard (Georg.Engelhard@imr.no)
Ulf Dieckmann (dieckman@iiasa.ac.at)
Olav Rune Godø (Olav.Rune.Godoe@imr.no)

Introduction

Age at maturation, or the age at which animals attain the capacity to reproduce, is tightly linked to lifetime reproductive success of individuals (Stearns, 1992; Bernardo, 1993). At population level, age at maturation is thus a key factor influencing stock productivity. The large fluctuations in many commercially important fish stocks may be better understood if long-term trends in age at maturation can be established (e.g. Godø, 2000 on the dynamics of maturation in Northeast Arctic cod, *Gadus morhua* L.). Unfortunately, for many major fish stocks such knowledge is relatively limited or absent. By contrast, data on length and age composition are often routinely collected, normally based on studies of the annual growth layers in either scales or otoliths.

Can routine growth layer measurements be used to predict age at maturation, with a satisfactory degree of confidence? This question is addressed here for Norwegian spring-spawning herring, *Clupea harengus* L., a large herring stock of primary commercial significance and characterised by substantial fluctuations in biomass (Hjort, 1914; Toresen and Østvedt, 2000). From high abundance in the 1940s, this stock declined drastically during the 1950s and 1960s probably due to a combination of overfishing and unfavourable climatic conditions, until it collapsed to near-commercial extinction in the late 1960s. Since then it has recovered, slowly during the 1970s, but more rapidly since the late 1980s; it is now considered fully recovered (Toresen and Østvedt, 2000). It is known that the stock size fluctuations were often accompanied by pronounced changes in the patterns of growth and maturation (Runnström, 1936; Toresen, 1990). The long-term patterns of maturation in this stock, however, have not yet been properly described, and are particularly poorly known for the most recent three decades.

Age at maturation varies considerably in Norwegian spring-spawning herring, largely due to the wide range of environmental conditions experienced by the juveniles (Dragesund et al., 1980). The spawning areas are distributed over a wide latitudinal gradient along the west coast of Norway between 58°N and 69°N (Devold, 1963; Hamre, 1990). After hatching, the juveniles drift northeastward with the Norwegian coastal current and so either reach the fjords of western and northern Norway, or the Barents Sea (Dragesund, 1970). Those ending up in the temperate waters along the Norwegian west coast generally grow rapidly, and usually spend only 1-2 years in the nursery areas before migrating to the Norwegian Sea to mix with shoals of older fish. By contrast, those ending up in the Arctic water masses of the Barents Sea grow slowly usually remaining for about 3-5 years in the nursery areas (Barros and Holst, 1995). After, usually, 1-2 summers during which the herring forage and live more pelagically in the Norwegian Sea, the animals mature. This results in ages at maturation that can vary between 3 and 9 years, with the majority of individuals maturing at ages 4 to 8 (Runnström, 1936).

Since the early 1900s, extensive collections of Norwegian spring-spawning herring scales have been made by the Institute of Marine Research, Bergen, Norway (Hjort, 1914). In herring scales, the formation of gonads, characterising the process of maturation, is reflected by a subtle change in width and microstructure of the corresponding annual growth layer (Lea, 1928; Runnström, 1936). This, however, can only be observed in well-preserved scales from the lateral side of the body. Prior to circa 1970, the sampled herring were mainly caught using drift-nets and purse-seines, techniques that result in relatively minor scale loss, allowing the collection of scales in good condition. Experienced scale readers routinely counted the number of spawning rings based on visual evaluation of the scale structure, so that the age at maturation could directly be derived from observation. Since the 1970s, however, samples have mainly been taken from purse-seiners and to an increasing extent from pelagic trawlers. This has resulted in higher scale loss and hence, a relatively large proportion of scales of poorer quality. Direct observation of age at maturation from scales was, therefore, discontinued in 1974.

In contrast, we do have information on the widths of annual growth layers in Norwegian spring-spawning herring scales over a very long period, starting in the 1930s and extending to the present day. If it is possible to predict age at maturation from such data it would allow the establishment of a long-term time series on this important life-history characteristic. Such information will be particularly valuable in the light of changes in the stock, that may have occurred when it collapsed in the 1960s. The goal of the present paper is to describe and evaluate the efficiency of two methods, discriminant analysis and artificial neural network analysis, to classify age at maturation in adult Norwegian spring-spawning herring, based on the widths of annual growth layers in scales.

Materials and Methods

Data collection

The two classification methods were evaluated using mature individuals of Norwegian spring-spawning herring, collected by the Institute of Marine Research between 1935 and 1973. Samples of 100-200 herring were collected from drift-net, beach-seine, purse-seine, or trawl catches, caught by both commercial and research vessels. For each fish, standard measurements were taken, including body mass, total length, sex, and maturity stage.

When available, up to four scales were collected from the skin just behind the operculum, along the lateral body line. Scales were mounted on microscopic glass plates coated with gelatine and thus conserved for later analysis. By microscopic examination of the scales shortly after preparation, scale readers determined the age, based on the total number of growth layers (Lea, 1911) and moreover, the age at maturation, based on observations on each of the growth layers (Lea, 1928, 1929; Runnström, 1936). This implied a distinction between (1) 'coastal' rings corresponding to the juvenile stage (rather narrow to very wide summer zones divided by either diffuse or sharp winter rings), (2) 'oceanic' rings corresponding with the late immature stage when the animals live in the Norwegian Sea (wide summer zones divided by diffuse winter rings), and (3) 'spawning' rings corresponding to years during which the herring spawned (narrow to very narrow outer summer zones divided by sharp winter rings) (Fig. 1).

For a large number of these historical scale samples (collected before 1974), measurements on annual growth layers were carried out recently (during the 1990s). These measurements followed the new methodology of scale examination that has been in practice at the Institute of Marine Research since 1974 and has replaced the older method described above, which distinguished between 'coastal', 'oceanic', and 'spawning' rings. The new method was described by Barros and Holst (1995) and involves the measurement of the total radius of the scale and of the radius of each annual growth layer up to the 9th along a line running from the focus to the edge of the scale, by means of a stereomicroscope fitted with an ocular micrometer (Fig. 1).

Description of the data

The data used here included 45 386 individuals, that satisfied the following conditions: (1) they had been classified as mature based on evaluation of the gonads; (2) age and age at maturation had been interpreted by observation of the scales; and (3) the widths of annual scale increments had been quantified by direct measurement (unit: mm).

For classifying age at maturation, it is required that the data be stratified according to age for two simple reasons. First, the number of measured scale increments and as a result, that of explanatory variables is dependent on age. Second, while a mature individual caught at, say, age 4 can only have matured at an age of either 3 or 4, an individual caught at, say, age 11 may have matured at any age between 3 and 9. An overview of the age-stratified dataset is given in Table 1.

Discriminant analysis (DA)

Discriminant analysis is a well-known statistical procedure used to build predictive models of group membership based on observed characteristics of each individual. Two-group discriminant analysis generates one discriminant function based on linear combinations of the predictor variables that result in the best discrimination between two groups. Multiple discriminant analysis allows the discrimination between more than two groups by generating a set of discriminant functions. These functions are generated

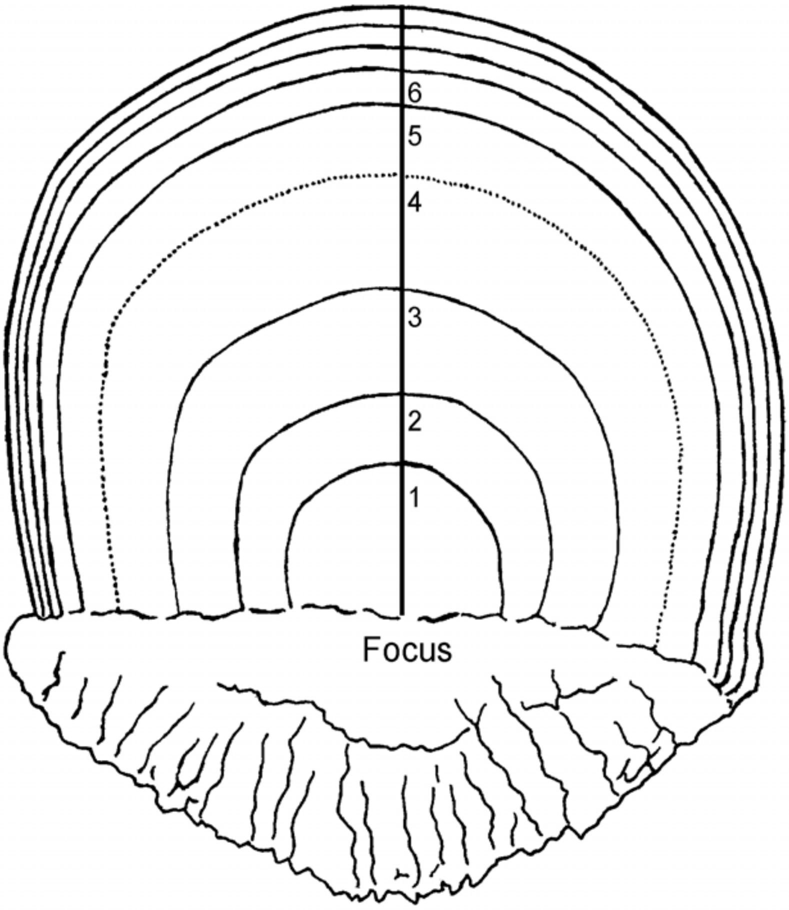

Figure 1. Lateral scale from a 9-year-old herring caught in February. The widths of annual growth layers are measured along the imaginary line indicated, which runs from the focus of the scale of its periphery. Note that this scale shows 3 wide ‘coastal’ rings separated by sharp winter rings; 2 wide ‘oceanic’ rings separated by a diffuse winter ring; and 4 narrow ‘spawning’ rings separated by sharp winter rings. Thus, the age at maturation is 5.

from individuals with known group membership; the functions can then be applied to other individuals of unknown group membership; given that measurements for the predictor variables are available.

For our application to predict age at maturation from scale measurements, it was practical to stratify the data according to age (Table 1). This implied that for each age group (age 3, 4, ..., 8, 9+) distinct discriminant analyses were carried out. Predictor variables were the widths of annual scale increments up to the 9th, after log-transformation of the data to obtain normality and increase the homogeneity of variance. Dependent variable was age at maturation (at 3, 4, ..., 9 years), defined *a priori* based on the direct observations by scale readers. The discriminant analyses were carried out using the SPSS 10.0.7 Windows package (SPSS Inc., 1989-1999).

Table 1. Sample sizes for herring individuals where both scale increments were measured and age at maturation was determined based on the observation of ‘spawning rings’, used to obtain the proposed procedures for classifying age at maturation.

	Age at maturation							
Age	3	4	5	6	7	8	9	All
3	129							129
4	52	2 865						2 917
5	68	1 164	3 658					4 890
6	47	1 163	1 330	3 602				6 142
7	37	950	2 429	1 233	1 946			6 595
8	26	1 033	2 597	2 071	650	972		7 349
9+	65	1 968	4 948	5 879	2 768	1 358	378	17 364
All	424	9 143	14 962	12 785	5 364	2 330	378	45 386

Artificial neural network (NN)

Artificial neural networks were also tried as an alternative method to predict age at maturation from scale measurements in Norwegian spring-spawning herring. Such neural networks imitate human neuron functioning, transforming an activating variable into a non-linear response. They can be applied as an alternative to various statistical procedures, and are particularly useful in cases of non-linear relationships between predictor and dependent variables (Fausett, 1994; Basheer and Hajmeer, 2000). Several recent studies showed that, for the purpose of classification, artificial neural networks often have superior predictive performance when compared to conventional statistical procedures, i.e. discriminant analysis and logistic regression (e.g. Edwards and Morse, 1995; Simmonds et al., 1996; Lek et al., 1996; Özesmi and Özesmi, 1999; Manel et al., 1999a, 1999b).

We constructed artificial neural networks written in the programming language C. For each of the age groups (age 3, 4, ..., 8, 9+), separate neural network analyses were carried out (as was the case in discriminant analyses). These networks could be characterised as three-layer feed-forward neural networks, and the architecture is described in the Appendix to this paper. The networks were trained by means of the back-propagation learning algorithm (Rumelhart et al., 1986). The prediction of age at maturation from scale measurements by use of neural networks occurred in two major phases. First, during the *training phase* internal parameters within the network (weights) were adjusted iteratively such that the performance of the network, equivalent to accurately predicting age at maturation, was maximised; this stage continued until there was no further increase in network performance, or classification success (see Appendix for more detail on the training procedure). Second, during the *prediction phase* the final, optimal network obtained during the training phase was used to predict age at maturation for all individuals in the database.

Performance of classification methods

We used three indicators to judge the quality of the results obtained using discriminant analysis and neural networks:

(1) *Classification success*, defined as the proportion of correctly classified individuals, assessed per age group, as a general indicator of classification success.

(2) *Prediction errors*, defined as the absolute differences between observed and predicted values for age at maturation, and expressed either as the mean prediction error averaged over all cases, or as the proportion of cases where age at maturation was misclassified by more than 1 year.

(3) *Degree of under- or overestimation*. Using the Wilcoxon signed-rank test, we examined to what extent there was a tendency for estimated values for age at maturation, predicted using either DA or NN, to be either higher or lower than the observed values.

Quantification of overlap between maturation groups

If the different maturation groups show considerable overlap in (combinations of) the explanatory variables (i.e. scale measurements), then it can be expected that classification success will be *a priori* limited to a certain extent, regardless of the classification method used. We examined the *a priori* limitations to classification success for both discriminant and neural network analyses, by quantifying the degree of overlap in the multidimensional character space between the different maturation groups.

Overlap was quantified as follows. For all individuals belonging to a certain age group (age 3, 4, ..., 9+), the coordinates describing their locations in the multidimensional character space were determined by their values for the scale measurements. Next, for each individual (1) the number of ‘neighbouring’ individuals situated closer to it than a certain maximal distance in the multidimensional character space was calculated; and (2) out of these, the fraction of *similar neighbours* was computed (i.e., other individuals characterised by the same age at maturation as the focal individual). The fraction of similar neighbours within a certain multivariate distance was then averaged over all individuals in a given age group. This procedure, simplified to a two-dimensional character space, is visualised in Figure 2.

The mean fraction of similar neighbours situated within a very small multivariate distance will provide an indication of the hypothetical limitations to maximal classification success. Unfortunately, when the distance defining neighbourhood decreases, also the total number of neighbours will decrease; as a result, the estimation of the fraction of similar neighbours will become less accurate when distance decreases. Therefore, the procedure was repeated for a range of different neighbourhood-defining distances (0.05, 0.10, 0.15, ..., 2.50 mm). The results were plotted and examined graphically.

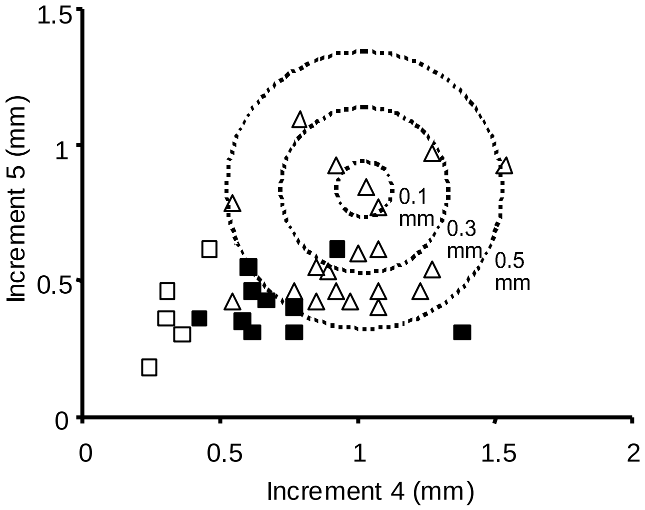

Figure 2. Two-dimensional visualisation of the procedure used to quantify the degree of overlap in scale measurement data between different herring maturation groups. For 30 individuals aged 5, the multidimensional character space (widths of all scale increments) is here reduced to a two-dimensional character space (widths of increments 4 and 5 only). Open squares, filled squares, and triangles represent the coordinates of individuals that matured at ages 3, 4, and 5, respectively. For each individual, (1) the number of 'neighbours' situated within a certain multivariate, neighbourhood-defining distance is computed (e.g., within either 0.1 mm, 0.3 mm, or 0.5 mm, as illustrated here for one focal individual); and (2) out of these, the fraction of neighbours similar in age at maturation is computed. Fractions of similar neighbours are then averaged over all individuals, for different neighbourhood-defining distances.

Results

Classification success

The overall proportion of correct classifications of age at maturation in herring based on scale measurements was 68.0% if discriminant analysis (DA) was used, and 66.6% if neural networks (NN) were used. The overall success rate was only marginally, but nevertheless significantly ($P < 0.0001$) higher for DA than for NN (Table 2).

An assessment of classification success per age group (Table 2) shows that the proportion of successful classifications, as expected, decreased with age, from 100% in mature fish caught at age 3 (only maturation at age 3 possible) to 56.4% (DA) or 54.9% (NN) in mature fish caught at age 9 or higher (maturation at all ages from 3 to 9 possible). For all age classes, there were only minor differences in classification success based on the two proposed methods. In particular, methodological differences were negligible for fish caught at ages 3-6; for fish caught at higher ages, classification success was marginally higher using DA than using NN.

Prediction errors

Performance as indicated by prediction errors was marginally better for NN than for DA (Table 3). The overall mean prediction error, defined as the absolute difference between

Table 2. Classification success (proportion of correct classifications) of age at maturation in herring based on scale measurements, using either discriminant (DA) or neural network (NN) analysis. The Wilcoxon signed-rank test is used to examine for differences in classification success between the two methods; positive values of *Z* indicate higher success for DA. Values of $P < 0.05$ are emboldened.

		DA		NN		Wilcoxon	
Age	*n*	Correct *n*	Correct %	Correct *n*	Correct %	*Z*	*P*
3	129	129	100.0 %	129	100.0 %	0	1
4	2 917	2 867	98.3 %	2 868	98.3 %	−0.180	0.857
5	4 890	4 074	83.3 %	4 055	82.9 %	1.420	0.156
6	6 142	4 880	79.5 %	4 912	80.0 %	−1.482	0.138
7	6 595	4 536	68.8 %	4 193	63.6 %	8.760	< **0.0001**
8	7 349	4 596	62.5 %	4 514	61.4 %	1.988	**0.047**
9+	17 364	9 788	56.4 %	9 540	54.9 %	3.652	**0.0003**
All ages	45 386	30 870	68.0 %	30 211	66.6 %	7.138	< **0.0001**

observed and predicted values for age at maturation, was 0.382 year when using DA, and 0.367 year when using NN. The proportion of cases where the prediction error was over 1 year was higher when using DA (5.2%) than when using NN (2.9%). The difference in prediction errors between the two methods was small but significant ($P < 0.0001$; Table 3).

Degree of misclassification, as expected, increased with age at catch (Table 3). In fish caught at age 4, prediction errors averaged 0.017 year with both methods (range of errors 0-1 year). In fish caught at age 9 or higher, prediction errors averaged 0.521 year if DA was used, and 0.503 year if the NN was used (range of errors 0-5 year). While differences in the degree of misclassification between the two proposed methods were negligible for fish caught at low ages (3-5, 7), the differences were more pronounced for fish caught at higher ages (6, 8 and higher).

Prediction errors resulting from DA were significantly correlated with those resulting from NN (Table 3). This implied that if age at maturation for a given individual was misclassified using DA, it was also likely to be misclassified to a similar extent using NN.

Degree of under- or overestimation

Over the whole dataset, age at maturation was slightly underestimated using DA, but on average only by 0.06 year which represented less than 1.1% of the mean age at maturation (Table 4). When the NN was used for prediction, age at maturation was on

Table 3. Prediction errors in estimating age at maturation in herring from scale measurements, using either discriminant (DA) or neural network (NN) analysis. Prediction errors represent the absolute differences between observed and predicted values for age at maturation. The Wilcoxon signed-rank test is used to examine if prediction errors were significantly higher (positive values of *Z*) or lower (negative values of *Z*) using DA than using NN. The Spearman rank correlation is used to examine for associations between prediction errors using either DA or NN. Values of *P* < 0.05 are emboldened.

		Mean prediction error		Prediction error >1		Test		Correlation	
Age	*n*	DA	NN	DA	NN	*Z*	*P*	r_S	*P*
3	129	0.000	0.000	0.0 %	0.0 %	0	1		
4	2 917	0.017	0.017	0.0 %	0.0 %	0.180	0.857	0.682	< **0.0001**
5	4 890	0.172	0.173	0.5 %	0.2 %	–0.205	0.838	0.869	< **0.0001**
6	6 142	0.224	0.209	1.8 %	0.9 %	3.70	**0.0002**	0.769	< **0.0001**
7	6 595	0.393	0.391	7.4 %	2.5 %	0.413	0.680	0.520	< **0.0001**
8	7 349	0.466	0.428	7.5 %	3.6 %	5.43	< **0.0001**	0.534	< **0.0001**
9+	17 364	0.521	0.503	6.7 %	4.7 %	3.75	**0.0001**	0.518	< **0.0001**
All ages	45 386	0.382	0.367	5.2 %	2.9 %	6.24	< **0.0001**	0.598	< **0.0001**

average overestimated, however to the even lesser extent of 0.01 year (0.2% of mean age at maturation; Table 4). For the different age groups, age at maturation predicted using DA or NN was on average either similar to, or slightly lower, or slightly higher than the averages for the observed values. The degree of under- or overestimation was always very small, never accounting for more than 2.3% of the average observed values.

Overlap in scale measurement data between maturation groups

There was considerable overlap in data on scale measurements between the different maturation groups. This was shown by the analysis, per age group, of the average fraction of neighbours similar in age at maturation as a function of multivariate distance defining neighbourhood (Fig. 3). If a relatively low neighbourhood-defining distance was chosen (range 0.05-0.20 mm), there was close agreement between the percentage of neighbours similar in age at maturation per age groups (Fig. 3), and classification successes using both DA and NN (Table 2). As an example, for fish aged 5, within a multivariate distance of 0.1 mm the mean fraction of neighbours similar in age at maturation was 81.0%, in accordance with classification successes of 83.3% using DA and 82.9% using NN. For fish aged 9 or more, the mean fraction of similar neighbours within the same distance was 53.3%, in accordance with classification successes of 56.4% using DA and 54.9% using NN.

Table 4. Degree of under- or overestimation of age at maturation, using either discriminant (DA) or neural network (NN) analysis. Means with standard deviations describe age at maturation as observed or predicted using DA or NN. The Wilcoxon signed-rank test is applied to examine if age at maturation is significantly under- or overestimated. Negative and positive values of *Z* indicate under- and over-estimation of age at maturation, respectively. Values of $P < 0.05$ are emboldened.

		Observed	Predicted using DA			Predicted using NN		
Age	*n*	Mean ±SD	Mean ±SD	*Z*	*P*	Mean ±SD	*Z*	*P*
3	129	3.00 ±0.00	3.00 ±0.00	0	1	3.00 ±0.00	0	1
4	2 917	3.98 ±0.13	3.99 ±0.11	2.26	**0.024**	4.00 ±0.04	6.71	<**0.0001**
5	4 890	4.73 ±0.47	4.77 ±0.45	5.66	<**0.0001**	4.79 ±0.41	9.42	<**0.0001**
6	6 142	5.38 ±0.81	5.38 ±0.86	0.31	0.758	5.39 ±0.80	1.77	0.077
7	6 595	5.62 ±1.07	5.55 ±1.09	−7.89	<**0.0001**	5.63 ±0.95	1.05	0.294
8	7 349	5.71 ±1.21	5.65 ±1.22	−6.31	<**0.0001**	5.58 ±1.11	−14.7	<**0.0001**
9+	17 364	5.85 ±1.19	5.76 ±1.12	−16.2	<**0.0001**	5.93 ±1.01	12.0	<**0.0001**
All ages	45 386	5.49 ±1.16	5.43 ±1.12	−16.5	<**0.0001**	5.50 ±1.06	5.64	<**0.0001**

Discussion

The study shows that age at maturation in Norwegian spring-spawning herring can be estimated from scale measurements with reasonably high prediction success. Two entirely different classification methods, discriminant (DA) and neural networks (NN) analysis, both predicted age at maturation correctly in about 67-68% of all cases (Table 2). These success rates are especially encouraging, if one considers that maturation in Norwegian spring-spawning herring may occur at no less than 7 different ages, ranging from 3 to 9 (Runnström, 1936). For adult herring caught at the age of 4, where there are only two possible ages at maturation (3 and 4), classification success was as high as 98.3% regardless of the method used. There was only a very small fraction of individuals where the difference between observed and predicted age at maturation was more than 1 year; this fraction was 5.2% using DA and 2.9% using NN (Table 3). Over the whole dataset there was a significant, but only minor degree of underestimation of age at maturation if DA was used (by 1.1% of mean age at maturation), and a significant but even lower degree of overestimation if NN were used (by 0.2% of mean age at maturation; Table 4). The significant levels of under- and overestimation using DA

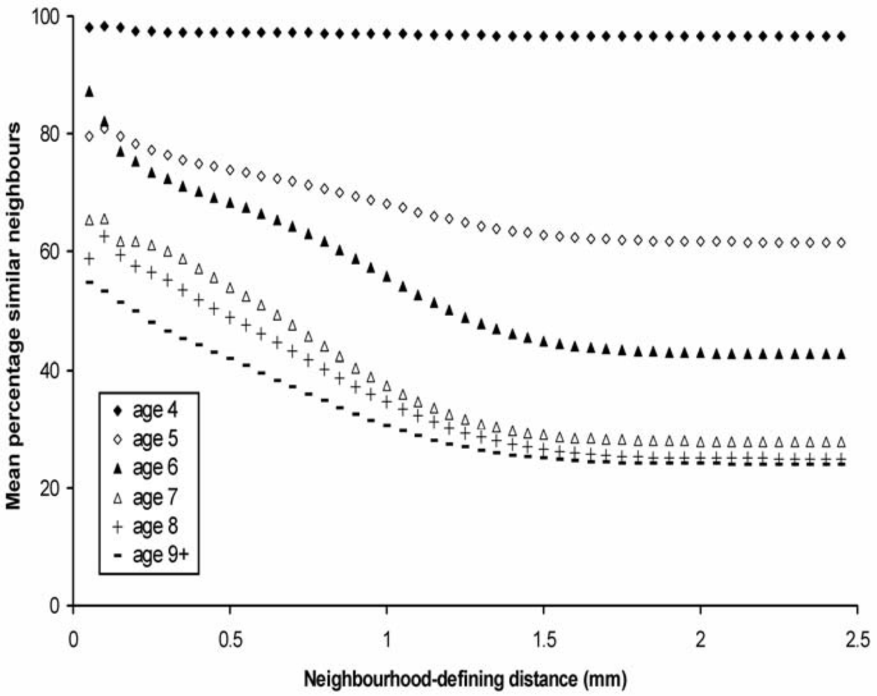

Figure 3. Mean percentage of neighbors similar in age at maturation as a function of multivariate distance defining neighbourhood (compare with Fig. 2), shown separately for age groups 4,5, …, 9+. For neighbours situated within small distances, note high correspondence of percentage similar neighbour with classification successes of both DA and NN (Table 2).

and NN, respectively, may be explained by our very large sample size (*n* = 45 386) rendering it extremely likely that levels of significance are reached even with very small differences, in particular in the case of paired comparisons. These combined indicators of classification performance underline that DA and NN are promising tools to predict age at maturation in herring from routine scale measurements.

The methods described here will therefore allow the analyses of long-term trends in age at maturation in one of the world's economically most important fish stocks (ICES, 2001), and in addition will allow studies on the effect of age at maturation on a range of important parameters, including size and condition at age, recruitment, and the total number of reproductive events throughout the lifespan of the herring under different fishing pressures (Beverton, 1992; Beverton et al., 1994; Godø, 2000). Moreover, such data will allow the analysis of reaction norms for age and size at maturation, with the possibility for disentangling genetic from phenotypic aspects of maturation (Heino et al., 2002). It is also important to note that similar methods might well be applicable to other commercially significant fish stocks, where either scales or otoliths are being collected routinely and the annuli examined for age reading and growth studies.

In this case study on Norwegian spring-spawning herring, the quality of the results obtained with either DA or NN was about equally high, and the question as to which method performed better in classifying age at maturation depended on the performance criterion used. If classification success, or the total number of cases classified at the exact, correct age at maturation, was used as performance criterion, then DA gave marginally but significantly better results than NN (Table 2). The two other performance criteria, however, indicated better results with NN. First, prediction errors were slightly, though significantly lower with NN (Table 3), implying that *if* a case was misclassified, the error was likely to be smaller with this method than using DA. Second, using NN age at maturation was on average overestimated to an (even) lower degree, than it was underestimated using DA (Table 4). Combined, these differences in classification performance between the two methods are of such a small magnitude, that DA and NN may be considered equally successful procedures to derive age at maturation from scale measurements in Norwegian spring-spawning herring.

In fact, it appears that both methods approach the maximal, hypothetical limits to classification as imposed *a priori* by the overlap in the predictor variables between the different groups to be classified. A clear indication of this was given by the analysis quantifying the degree of overlap in scale measurement data in the multidimensional character space between maturation groups (Fig. 3). At small neighbourhood-defining distances, the mean fraction of neighbours similar in age at maturation did not increase to approximately 100% as it would be expected were there no overlap in explanatory variables between the different groups. By contrast, for each of the age groups 4-9 this fraction, with decreasing neighbourhood-defining distances, approached percentages very similar to those characterising classification success using either DA or NN (Table 2). This implies that the DA and NN approach the limit to classification success imposed by the nature of the predictor data, and that no considerable improvement may be expected from any classification method.

To some extent, the overlap in scale data between maturation groups could be due to measurement inaccuracy when the radii of the scale annuli were measured, since the smallest unit of measurement was approximately 0.05 mm. Moreover, even though several skilled readers in a team with many years of experience were involved in the original process of *observing* age at maturation, some ‘judgement’ was often necessary in the distinction between typical ‘oceanic rings’ in the scale corresponding with the late immature stage (relatively wide summer zones separated by diffuse winter rings) and ‘spawning rings’ corresponding with the mature stage (narrow summer zones separated by sharp, fine winter rings: Lea, 1928; Runnström, 1936). The observations may therefore not always have provided the correct classifications for age at maturation. In particular, the annulus corresponding to the process of maturation is often intermediate in width to ‘spawning’ and previous ‘oceanic’ rings. Although generally the processes of maturation such as gonad formation and the first spawning migration lead to a considerable reduction in body growth rate, there are also many individuals that in the year of maturation form only (very) small gonads (Slotte, 2000) or may migrate to less distant spawning grounds (Slotte, 1999). This will result in a lower reduction in the rate of body growth and a less clear ‘mark’ of maturation in the scale structure (Runnström, 1936). Environmental influences in the year of maturation and natural variation in maturation between individuals are other possible causes of overlapping scale data among groups.

The very similar performances of NN and DA in predicting age at maturation did not agree with several recent studies, that report better, sometimes far better, classification results obtained with the relatively newly developed NN techniques when compared to the conventional, statistical methods of classification, discriminant analysis and logistic regression (e.g., Lek et al., 1996; Özesmi and Özesmi, 1999; Manel et al., 1999a, 1999b). In solving other biological problems, NNs also appear to be a superior alternative when compared to parametric modelling techniques such as multiple linear regression (e.g., Lek et al., 1995; Mastrorillo et al., 1997; Baran et al., 1999; Chen and Ware, 1999). In general, the high predictive power of NNs can be attributed to their ability to handle non-linear relationships between predictor and dependent variables particularly well, through the presence of many intervening information-processing units that each use the binary logistic activation function (Fausett, 1994; Basheer and Hajmeer, 2000). This advantage of NNs over parametric statistics, however, probably does not apply to our data, since the relationship between maturation and growth (and hence, scale structure) appears to be rather linear, or log-linear (Holst, 1996; Slotte, 1999). A further advantage of NNs is the small impact of extreme values on prediction success, and the absence of any specific assumptions on the distribution of the data (although data transformation may improve computational speed), while Gaussian data is an important assumption to be met in DA. Indeed, the data on scale increments, if log-transformed, were normally distributed, with virtually no extreme values.

Hence, in this study on Norwegian spring-spawning herring the data fulfilled very well the conditions required for the conventional statistical procedure of DA, and the extra advantages provided by NNs were relatively small. Such conditions might, however, be different in the case of other fish stocks where investigations on age at maturation based on growth layer data are being considered. NNs could still be the preferred option, in particular if data assumptions for DA are not fulfilled, and/or maturation is related with scale growth, or otolith growth, in an irregular, non-linear fashion.

References

Baran, P., Lek, S., Delacoste, M., and Belaud, A. 1996. Stochastic models that predict trout population density or biomass mesohabitat scale. Hydrobiologia 337: 1-9.

Barros, P., and Holst, J. C. 1995. Identification of geographic origin of Norwegian spring-spawning herring (*Clupea harengus* L.) based on measurements of scale annuli. ICES Journal of Marine Science 52: 863-872.

Basheer, I. A., and Hajmeer, M. 2000. Artificial neural networks: fundamentals, computing, design, and application. Journal of Microbiological Methods 43: 3-31.

Bernardo, J. 1993. Determinants of maturation in animals. Trends in Ecology and Evolution 8: 166-173.

Beverton, R. J. H. 1992. Patterns of reproductive strategy parameters in some marine teleost fishes. Journal of Fish Biology 41 (Suppl. B**)**: 137-160.

Beverton, R. J. H., Hylen, A., and Østvedt, O. J. 1994. Growth, maturation, and longevity of maturation cohorts of Northeast Arctic cod. ICES Marine Science Symposia 198: 482-501.

Chen, D. G., and Ware, D. M. 1999. A neural network model for forecasting fish stock recruitment. Canadian Journal of Fisheries and Aquatic Sciences 56: 2385-2396.

Devold, F. 1963. The life history of the Atlanto-Scandian herring. Rapports et Procès-Verbaux des Réunions du Conseil pour l’Exploration de la Mer 154: 98-108.

Dragesund, O. 1970. Distribution, abundance and mortality of young and adolescent Norwegian spring spawning herring (*Clupea harengus* Linné) in relation to subsequent year-class strength. Fiskeridirektoratets Skrifter, Serie Havundersøkelser 15: 451-556.

Dragesund, O., Hamre, J., and Ulltang, Ø. 1980. Biology and population dynamics of the Norwegian spring-spawning herring. Rapports et Procès-Verbaux des Réunions du Conseil pour l’Exploration de la Mer 177: 43-71.

Edwards, M., and Morse, D.R. 1995. The potential for computer-aided identification in biodiversity research. Trends in Ecology and Evolution 10: 153-158.

Fausett, L. 1994. Fundamentals of neural networks: architectures, algorithms, and applications. Prentice-Hall, Englewood Cliffs, New Jersey.

Godø, O. R. 2000. Maturation dynamics of Arcto-Norwegian cod. IIASA Interim Report IR-00-024, IIASA, Laxenburg, Austria, 16 pp.

Hamre, J. 1990. Life history and exploitation of the Norwegian spring spawning herring. *In* Biology and fisheries of the Norwegian spring spawning herring and blue whiting in the Northeast Atlantic, pp. 5-39. Ed. by T. Monstad. Institute of Marine Research, Bergen, Norway. 358 pp.

Heino, M., Dieckmann, U., and Godø, O. R. 2002. Measuring probabilistic reaction norms for age and size at maturation. Evolution 56: 669-678.

Hjort, J. 1914. Fluctuations in the great fisheries of Northern Europe viewed in the light of biological research. Rapports et Procès-Verbaux des Réunions du Conseil pour l'Exploration de la Mer 20: 1-228.

Holst, J. C. 1996. Long term trends in the growth and recruitment pattern of the Norwegian spring-spawning herring (*Clupea harengus* Linnaeus 1758). Ph.D. thesis, University of Bergen, Norway. 131 pp.

ICES 2001. Report of the Northern Pelagic and Blue Whiting Fisheries Working Group, Reykjavik, Iceland, 18-27 April 2001. ICES CM 2001/ACFM: 17.

Lea, E. 1911. A study on the growth of herrings. Publications de Circonstance 61: 35-57.

Lea, E. 1928. Undersøkelser over den norske sild: fra umoden til moden. Årsberetning vedkommende Norges Fiskerier 1927 (IV): 1-36.

Lea, E. 1929. The oceanic stage in the life history of the Norwegian herring. Journal du Conseil 4: 1-42.

Lek, S., Belaud, A., Dimopoulos, I., Lauga, J., and Moreau, J. 1995. Improved estimation, using neural networks, of the food consumption of fish populations. Marine and Freshwater Research 46: 1229-1236.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., and Aulagnier, S. 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecological Modelling 90: 39-52.

Manel, S., Dias, J. M., Buckton, S. T., and Ormerod, S.J. 1999a. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. Journal of Applied Ecology 36: 734-747.

Manel, S., Dias, J. M., and Ormerod, S. J. 1999b. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecological Modelling 120: 337-347.

Mastrorillo, S., Lek, S., and Dauba, F. 1997. Predicting the abundance of minnow *Phoxinus phoxinus* (Cyprinidae) in the River Ariège (France) using artificial neural networks. Aquatic Living Resources 10: 169-176.

McClelland, J.L., and Rumelhart, D.E. 1988. Explorations in parallel distributed processing: a handbook of models, programs, and exercises. Massachusetts Institute of Technology, Cambridge, MA.

Özesmi, S. L., and Özesmi, U. 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. Ecological Modelling 116: 15-31.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning representations by back-propagating errors. Nature 323: 533-536.

Runnström, S. 1936. A study on the life history and migrations of the Norwegian spring-spawning herring based on the analysis of the winter rings and summer zones of the scale. Fiskeridirektoratets Skrifter, Serie Havundersøkelser 5 (2): 1-103.

Simmonds, E. J., Armstrong, F., and Copland, P.J. 1996. Species identification using wideband backscatter with neural network and discriminant analysis. ICES Journal of Marine Science 53: 189-195.

Slotte, A. 1999. Effects of fish length and condition on spawning migration in Norwegian spring-spawning herring (*Clupea harengus* L.). Sarsia 84: 111-127.

Slotte, A., and Fiksen, Ø. 2000. State-dependent migration in Norwegian spring-spawning herring. Journal of Fish Biology 56: 138-162.

Stearns, S. C. 1992. The evolution of life histories. Oxford University Press, New York. 249 pp.

Toresen, R. 1990. Long-term changes in growth of Norwegian spring-spawning herring. Journal du Conseil International pour l’Exploration de la Mer 47: 48-56.

Toresen, R., and Østvedt, O. J. 2000. Variation in abundance of Norwegian spring-spawning herring (*Clupea harengus*, Clupeidae) throughout the 20th century and the influence of climatic conditions. Fish and Fisheries 1: 231-256.

Appendix

Architecture of neural network to derive age at maturation from scale measurements

The architecture of the artificial neural networks applied here can be described as three-layer feedforward neural networks trained by backpropagation. The first layer (input layer) comprises 3-9 *input units* containing the information on the predictor variables (i.e. data on annual scale increments, after log-transformation); the number of input units is thus equivalent to the number of measured increments. The second layer (hidden layer) comprises 14-19 *hidden units*; the total number of hidden units is chosen based on a trade-off between limiting computation time and obtaining sufficiently satisfactory classification results. The third layer (output layer) comprises a single *output unit* determining the output of the network, representing the variable to be predicted (i.e. age at maturation). All units in the input layer are connected with all units in the hidden layer, and these in turn are all connected with the output unit. Specific, modifiable weights are attributed to each of the connection links between units of successive layers. These weights are the link between the problem and the solution and are therefore said to contain the 'knowledge' of the neural network about the overall problem (Baran et al., 1996). At the start of the training phase, all weights in the network are initialised to small random values (–0.1, ..., 0.1); over the course of the training phase, the weights are gradually modified using the back-propagation learning algorithm until network performance is optimised (Rumelhart et al., 1986).

Each unit has an *activation* which determines its output signal. The activations of the input units are equal to the values for the predictor variables of a given case. The activations of the hidden units are computed in two steps. First, each of the input units emits a weighted output signal to all hidden units, equal to its activation multiplied by the weight associated to the specific connection link. Each of the hidden units summarizes the weighted input signals to compute its net input, as follows (Fausett, 1994):

$$net_h = bias_h + \sum_{i=1}^{I} a_i w_{ih} \tag{1}$$

where net_h is the net input received by the hth hidden unit, a_i is the activation of the ith input unit, w_{ih} is the weight associated with the connection link between input unit i and hidden unit h, I is the total number of input units, and $bias_h$ is a bias on hidden unit h. The bias may be compared to the constant in parametric statistical analyses. Next, each hidden unit computes its activation a_h by applying the activation function f to its net input; here, the binary sigmoid function $f(x) = (1 + e^{-x})^{-1}$ is used, which is one of the most typical activation functions:

$$a_h = f(net_h) \tag{2}$$

Next, each of the hidden units sends its weighted output signal to the output unit o, which summarizes its incoming signals to compute its own net input, as follows:

$$net_o = bias_o + \sum_{h=1}^{H} a_h w_h \tag{3}$$

where net_o is the net input received by the output unit, a_h is the activation of the hth hidden unit, w_h is the weight associated with the connection link between hidden unit h and the output unit, H is the total number of hidden units, and $bias_o$ is a bias on the output unit. The output unit then applies the binary sigmoid activation function to compute its activation a_o, which is the actual output of the network:

$$a_o = f(net_o) \tag{4}$$

The back-propagation learning rule (Rumelhart et al., 1986) to train the network implies that weights are modified in a backward sweep, according to the generalised delta rule (Rumelhart et al., 1986; McClelland and Rumelhart, 1988). First, an error information term δ_o is computed for the output unit by comparison of the actual and desired output of the net:

$$\delta_o = (d_o - a_o) f'(net_o) \tag{5}$$

where d_o and a_o are the desired and actual activation of the output unit, and f' is the derivative of the binary sigmoid activation function; hence, $f'(x) = f(x)\,[1 - f(x)]$. The weights on links from the hidden to the output layer are then corrected, according to the formula:

$$\Delta w_h = \varepsilon \delta_o a_h \tag{6}$$

where Δw_h is the weight correction term for the link from hidden unit h to the output unit, and ε is the learning rate parameter (here, a value of 0.1 was chosen). Next, for each of the hidden units an error information term δ_h is computed:

$$\delta_h = \delta_o w_h f'(net_h) \tag{7}$$

where w_h is the weight on the link from hidden unit h to the output unit. The weights on links from input units to hidden units are then corrected, according to the formula:

$$\Delta w_{ih} = \varepsilon \delta_h a_i \tag{8}$$

The training phase was terminated based on the mean-squared-error convergence criterion (McClelland and Rumelhart, 1988). The dataset was divided into a training subset comprising 67% of the cases, and a testing subset comprising 33% of the cases. During training epochs, the network is first adapted based on all cases in the training subset; its performance is then monitored based on the independent testing subset. A measure of performance is the mean-squared-error (E) calculated over all cases in the testing subset, as follows:

$$E = \frac{1}{P} \sum_{p=1}^{P} (d_p - a_p)^2 \tag{9}$$

where d_p and a_p are the desired and actual output of the network of a testing example p, and P is the total number of testing examples. Training epochs continue as long as there is an increase in the performance of the network for the testing subset, i.e. as long as E decreases.