

# Recent advances in statistical methodology applied to the Hjort liver index time series (1859–2012) and associated influential factors<sup>1</sup>

Gudmund H. Hermansen, Nils Lid Hjort, and Olav S. Kjesbu

**Abstract:** Certain recent advances in statistical methodology have promising potential for fruitful use in general biology and the fisheries sciences. This paper reviews and discusses some of the relevant themes, including accurate modelling via focused model selection techniques, dynamic goodness-of-fit testing of processes evolving over time, finding break points for phenomena experiencing regime shifts, prediction uncertainty, and optimal combination of information across diverse sources via confidence distributions. The methods are illustrated for the Hjort liver quality index time series. Its roots lie in the classic study by Hjort in 1914, where liver quality of the Northeast Arctic cod (*Gadus morhua*) for 1880–1912 was reported on and studied, along with related factors, making it one of the first teleost time series ever published. The series has been extended both backwards and forwards in time, to 1859–2012, due to comprehensive archival and calibration efforts of Kjesbu et al. in 2014, yielding one of the longest time series of marine science. Our study offers a detailed examination of this series and how it relates to and interacts with associated factors, including winter temperatures, length distribution parameters, cod mortality, and a certain index related to availability of food. We identify certain mild nonstationary aspects of the time series, show that there is a regime shift around 1990 in the ways the liver series interacts with winter temperatures, and demonstrate that mortality and food availability play important roles.

**Résumé :** Certaines avancées récentes en méthodologie statistique présentent un potentiel intéressant d'utilisation en biologie générale et dans les sciences halieutiques. L'article présente une synthèse et une discussion de certains des thèmes pertinents, dont la modélisation exacte grâce à des techniques de sélection de modèles ciblées, la vérification dynamique de la qualité de l'ajustement de processus qui varient dans le temps, la recherche de points de rupture pour des phénomènes caractérisés par des changements de régime, l'incertitude des prédictions et la combinaison optimale d'information provenant de sources variées par l'utilisation de distributions de confiance. Les méthodes sont illustrées pour une série chronologique de l'indice de qualité du foie de Hjort, une des premières séries chronologiques publiées concernant des téléostéens, dont les fondements se trouvent dans l'article classique de Hjort en 1914, qui rend compte de la qualité des foies de morues (*Gadus morhua*) de 1880 à 1912 et étudie cette dernière, en plus de facteurs connexes. La série a été prolongée vers le passé et vers le futur, de 1859 à 2012, grâce aux efforts d'archivage et d'étalonnage exhaustifs de Kjesbu et al. en 2014, ce qui a produit une des plus longues séries chronologiques en sciences de la mer. Notre étude présente un examen détaillé de cette série et de son lien et ses interactions avec des facteurs associés, dont les températures hivernales, des paramètres relatifs à la distribution des longueurs, la mortalité de la morue et un indice relié à la disponibilité de nourriture. Nous cernons certains aspects légèrement non stationnaires de la série chronologique et démontrons qu'un changement de régime a eu lieu vers 1990 dans l'interaction de la série sur les foies avec les températures hivernales et que la mortalité et la disponibilité de nourriture jouent des rôles importants. [Traduit par la Rédaction]

## Introduction

The first four chapters of the classic Hjort (1914) are essentially occupied with the **quantity** of fish (specifically, herring (*Clupea harengus*) and cod (*Gadus morhua*)) and the associated underlying causes driving their fluctuations. He was, however, also concerned with what he terms the **quality** of fish and devotes most of the book's chapter 5 to discuss how this can reasonably be defined and measured, also attempting to identify factors involved as it varies from year to year. He proposed using the liver quality index "No. of hectolitres of liver pr. 1000 skrei" (i.e., Northeast Arc-

tic (NEA) cod) for such a purpose and established a time series of such measurements for the skrei for the years 1880–1912. This is arguably one of the first comprehensive teleost time series ever published. A few points of the same series were used and studied in Helland-Hansen and Nansen (1909, chapter 7.3).

Work summarized in Kjesbu et al. (2014a) has made it possible to extend this liver quality time series both backwards and forwards in time, using data from both fisheries statistics (1859–1990) and Råfisklaget (the Norwegian Fishermen's Sales Organisation, 1991–2012) along with further archival and calibration efforts. In this process, it has been found fruitful to pass from the somewhat

Received 10 February 2015. Accepted 15 August 2015.

Paper handled by Associate Editor C. Tara Marshall.

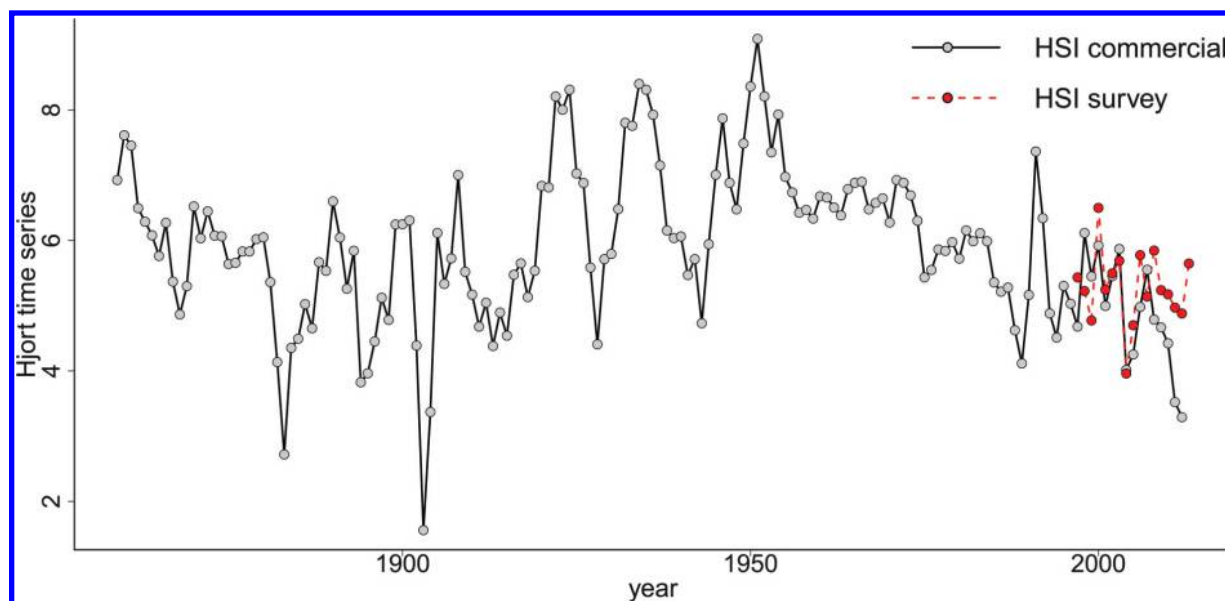
**G.H. Hermansen and N.L. Hjort.** Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo, Norway.

**O.S. Kjesbu.** Institute of Marine Research (IMR) and Hjort Centre for Marine Ecosystem Dynamics, P.O. Box 1870 Nordnes, 5817 Bergen, Norway; Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway.

**Corresponding author:** Gudmund H. Hermansen (e-mail: [gudmunhh@math.uio.no](mailto:gudmunhh@math.uio.no)).

<sup>1</sup>This article is part of the special issue based on the Johan Hjort Symposium on Recruitment Dynamics and Stock Variability, Bergen, Norway, 7–9 October 2014.

**Fig. 1.** The Hjort time series with the HSI, 1859–2012, from commercial fishing and IMR's Lofoten surveys (the latter from 1997 onwards); see Kjesbu et al. (2014a) and Table 1.



crude volumetric-based hectolitres of liver per 1000 fish used earlier to a more naturally standardized measurement called the hepatosomatic index (HSI). In bulk form this HSI is defined as total amount of liver (in kg) divided by total amount of fish (in kg). Passing from volumetric to mass scale has been achieved via an essentially linear relationship learned from regression analysis; see Kjesbu et al. (2014a, eq. 2). The resulting HSI data, perhaps the longest-running marine science time series there is, are displayed in Fig. 1. These authors also investigate the extent to which the Kola temperature (i.e., off the Kola Peninsula, Russia), where annual measurements are available from 1900 (with more detailed monthly mean temperatures from 1921 onwards, see Boitsov et al. 2012) can be seen to influence or interact with the HSI. Also, other covariates and their degree of relevance for the HSI are reviewed and examined in Kjesbu et al. (2014a), including fat content and body size.

That the HSI carries important biological information for the quality of a stock of fish has been demonstrated in several publications, from Helland-Hansen and Nansen (1909) and Hjort (1914) onwards. The background for this interest in amount of the liver from earlier days and up to today rests with the fact that this organ not only accumulates fat for subsequent metabolic costs in capital breeders (e.g. stored energy) like cod but also is a production site for yolk (vitellogenin) and eggshell (chorion) material, which are transported by the blood to the maturing ovary (Tyler and Sumpter 1996). Also, at the population level, HSI has shown clear positive links to the level of prey, in particular the stock size of capelin (*Mallotus villosus*), but also to individual fecundity and thereby total egg production (Marshall et al. 1999). Further elaborations of the importance of the liver in the present context are given in Sandeman et al. (2008) and Kjesbu et al. (2014a), where it is argued that the HSI is the universal expression of investment in liver size versus body size.

The aims of this article are twofold. The first goal is to provide more careful statistical analyses of the HSI parameter itself, from its definition, interpretation, and generalization from bulk index to individual index, to aspects of the full 1859–2012 time series. We also examine degrees of associations with related quantities and provide glimpses into the future, using data also to predict ahead as partly opposed to focusing on understanding the past. The covariates or influential factors we shall study later in our paper are as follows:

- (i) Kola temperatures (annual and winter), with data provided by Polar Research Institute of Marine Fisheries and Oceanography (PINRO; [www.pinro.ru/index-e.htm](http://www.pinro.ru/index-e.htm));
- (ii) mean length distribution, with data from the Institute of Marine Research (IMR) long-term commercial catch sampling program; see details in Kjesbu et al. (2014a);
- (iii) fish mortality rate,  $F$  (the usual parameter associated with continuous fishing and natural mortality; see, e.g., Hilborn and Walters 1992, chapter 10.3), with data from ICES (2014); and
- (iv) an index for the amount of food available (ICES 2014), the latter defined as the ratio of biomass for capelin with spawning stock biomass for the NEA cod (Kjesbu et al. 1998).

For convenient reference, the covariates are listed with brief annotations in Table 1.

Our second ambition is to use the opportunity to provide perspectives on and a brief overview of certain modern and relevant developments in statistical methodology. These relate to (i) choosing good models for complex phenomena, partly via the “focused viewpoint” used in the construction of certain focused information criteria for model selection (see Hjort and Claeskens 2003; Claeskens and Hjort 2003, 2008; Hermansen and Hjort 2015a, 2015b); (ii) assessing adequacy of fit using dynamical monitoring of log-likelihood maxima and other criteria, useful for examining processes evolving over time and for identifying regime shifts; (iii) the sometimes challenging statistical dividing line between “to explain” and “to predict” (cf. Breiman 2001; Shmueli 2010); and (iv) ways of optimally combining information across diverse data sources, via the theory of confidence distributions (see Xie and Singh 2013; Schweder and Hjort 2016). In subsequent sections, using these methods, we are able to identify certain mild aspects of nonstationarity of the HSI series and are also able to point to an apparent regime shift taking place around 1990 in the ways the HSI interacts with Kola winter temperatures. We shall also see that mortality and food availability appear to play more important roles for the HSI than Kola temperature and the length distribution.

## The HSI and targets for inference

In Hjort (1914, chapter 5), where the liver index is introduced as “No. of hectolitres of liver pr. 1000 skrei”, the underlying concern

**Table 1.** A summary including the main references for the most important data types used throughout the statistical analysis.

| Factor  | Data series       | Definition                                             | Time period          | Data source                                                                                |
|---------|-------------------|--------------------------------------------------------|----------------------|--------------------------------------------------------------------------------------------|
| HSI     | Commercial survey | Defined in eq. 1                                       | 1859–2012, 1997–2013 | Kjesbu et al. 2014a                                                                        |
| Kola    | Winter annual     | Mean October–March annual average                      | 1921–2013, 1900–2013 | Data from PINRO ( <a href="http://www.pinro.ru/index-e.htm">www.pinro.ru/index-e.htm</a> ) |
| Length  |                   | Mean length distribution                               | 1932–2012            | Data from the IMR long-term program; see Kjesbu et al. 2014a                               |
| F       |                   | Fish mortality rate; see e.g. Hilborn and Walters 1992 | 1913–2013            | Data from ICES 2014                                                                        |
| Capelin |                   | Proxy for available food; see Kjesbu et al. 1998       | 1981–2013            | Data from ICES 2014                                                                        |

and motivation is that of coming to grips with the **quality** of fish, supplementing the information of the **quantity**. It is clear from his discussion that also other and related versions of “quality” can be used. In this section we make some remarks pertaining to the statistical issues involved when defining and measuring appropriate indexes.

### What is the liver quality index (HSI)?

The definition mentioned above, used in, e.g., Hjort (1914, his figure 107), is a practical one, working in bulk modus, so to speak, without necessitating detailed examination of each individual fish. It may be represented as

$$(1) \quad \widehat{\text{HSI}}_{\text{bulk}} = 100 \times \frac{\text{total amount of liver}}{\text{total amount of fish}} = 100 \times \frac{\bar{x}}{\bar{y}}$$

where  $(x_i, y_i)$  represent the mass of liver and the total mass for fish  $i$ , and  $\bar{x}$  and  $\bar{y}$  are the respective means, over a sample of say  $n = 1000$  fish (in the Lofoten fishery, millions of fish are actually landed). This is the bulk liver index used in Kjesbu et al. (2014a), marked “commercial” in Fig. 1. An alternative definition, relating more directly to the individual fish, is

$$(2) \quad \widehat{\text{HSI}}_{\text{ind}} = 100 \times \left( \frac{1}{n} \sum_{i=1}^n \frac{\text{mass of liver in fish } i}{\text{mass of fish } i} \right) \\ = 100 \times \left( \frac{1}{n} \sum_{i=1}^n \frac{x_i}{y_i} \right)$$

This per-fish index has been measured for the years 1997 onwards as part of IMR’s Lofoten research survey, marked “survey” in Fig. 1.

These indexes are both biologically meaningful and are of course related but not equivalent. The underlying statistical parameters are respectively  $\widehat{\text{HSI}}_{\text{bulk}} = E(X/Y)$  for the bulk and  $\widehat{\text{HSI}}_{\text{ind}} = E(X/Y)$  for the per-fish index, with  $(X, Y)$  denoting liver mass and full mass for a randomly selected fish in the population in question, and “E” as usual denoting mathematical expectation of a random variable. The degree to which the two parameters differ is determined by aspects of the joint distribution of the two quantities  $X$  and  $Y$ , including their internal correlation, the spread of the distribution of  $Y$ , and the latter’s distance from zero.

The  $\widehat{\text{HSI}}_{\text{bulk}}$  time series cannot alone provide information on values of the individual-based parameter  $\widehat{\text{HSI}}_{\text{ind}}$ . For a given population, we may use research data on individual body metrics, however, collated from fishing ports over many years consulting “skrei” catches as in Hjort (1914) (Kjesbu et al. 1998, 2010) to analyse the relevant joint distribution of  $(X, Y)$ , leading also to a mechanism for predicting one HSI parameter from the other. Examining these data, with  $n = 439$  pairs of  $(x_i, y_i)$ , one learns first that  $X$  and  $Y$  are separately well modelled using gamma distributions and also that a five-parameter bivariate gamma model pro-

vides a fully adequate fit to the joint distribution. The model in question takes

$$(3) \quad X = G_1^{-1}(\Phi(U), a_1, b_1) \quad \text{and} \quad Y = G_2^{-1}(\Phi(V), a_2, b_2)$$

involving the inverses  $G_1^{-1}$  and  $G_2^{-1}$  of the gamma cumulative distribution functions with parameters  $(a_1, b_1)$  and  $(a_2, b_2)$ , respectively, the cumulative standard normal distribution function  $\Phi$ , and a standardized binormal pair  $(U, V)$  with correlation parameter  $\rho$ . Parameter estimates obtained by maximum likelihood were 2.51, 6.52, 3.99, 0.63, and 0.83, corresponding in particular to means and standard deviations 6.23 and 3.28 for  $x$  and 0.38 and 0.25 for  $y$ . The data fit the estimated gamma densities well (Fig. 2), passing Pearson-type goodness-of-fit tests.

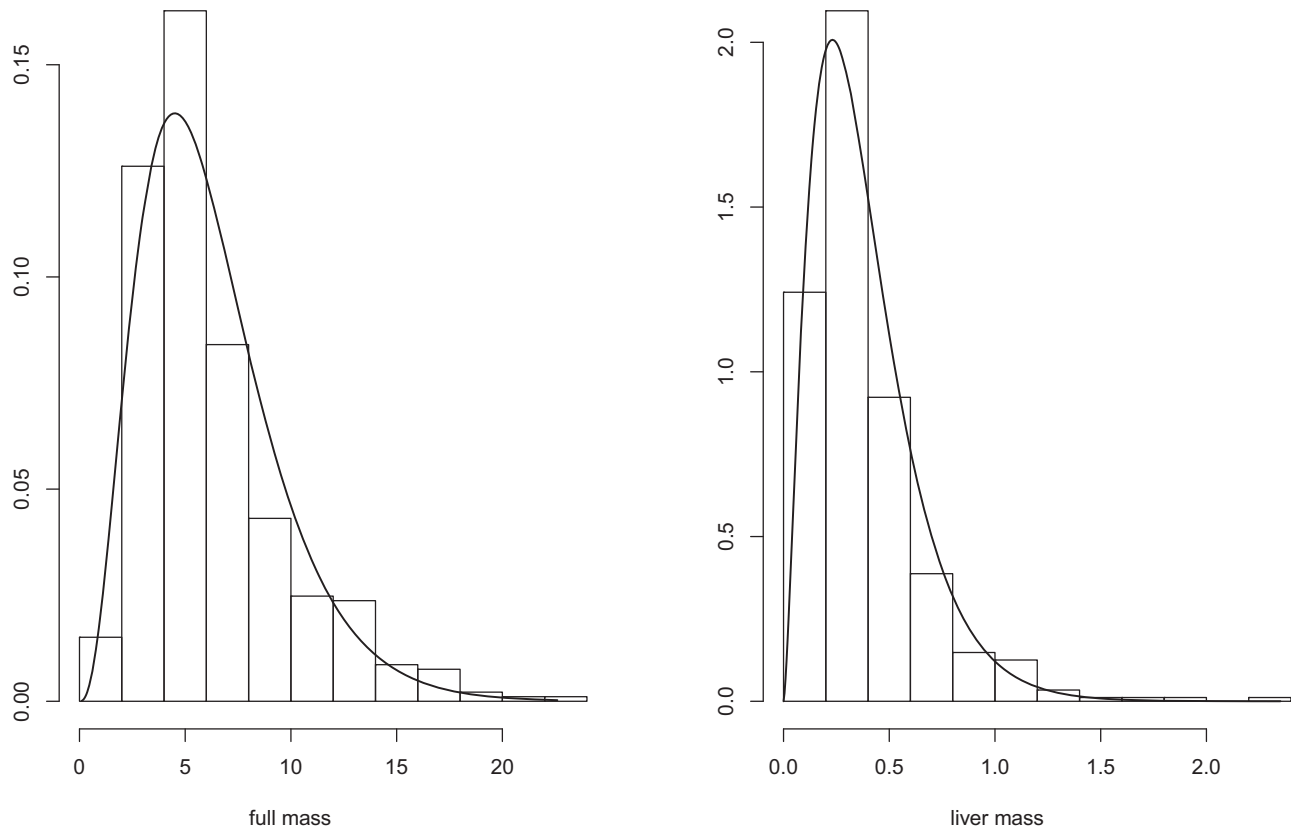
We may use the bivariate gamma model (eq. 3) to infer aspects of the connection between the bulk index  $\widehat{\text{HSI}}_{\text{bulk}}$  and the individual-fish index  $\widehat{\text{HSI}}_{\text{ind}}$  within the same stratum of a population. Figure 3 displays simulated pairs from the relevant distribution, taking for this illustration the sample size  $n$  above to be 1000. It leads to a correlation of 0.83 between the bulk and the per-fish HSI. This investigation also leads to the tentative formula

$$(4) \quad \widehat{\text{HSI}}_{\text{bulk}} = 1.581 + 0.786 \widehat{\text{HSI}}_{\text{ind}}$$

for translating the per-fish index to the bulk liver index. Equation 4 suggests possibilities for transforming back and forth between the two indexes, as with survey and bulk data (cf. Fig. 1). Such a formula would need to be used with care, however, as its precise coefficients relate to the population being sampled. To illustrate such ideas, also in connection with methods of combining information across sources, let us go back to the Hjort time series displayed in Fig. 1, now zooming in on 1997–2012. We wish to supplement the information there in two ways. First, assuming for this illustration that the formula above is reasonable for the grander population, we find bulk index estimates from the commercial series. Second, the methodology of optimal combination of information across data sources, involving confidence distributions and reviewed in a later section, may be used to provide the best estimates of the  $\widehat{\text{HSI}}_{\text{bulk}}$  for these years, utilizing both the commercial and the survey data (Fig. 4). Note for this application that the resulting estimated bulk HSI is higher than both the commercial  $\widehat{\text{HSI}}_{\text{bulk}}$  numbers and the survey  $\widehat{\text{HSI}}_{\text{ind}}$  numbers, due in part to the fact that eq. 4 yields bulk scores higher than individual scores for the range of the latter met here (cf. the scales involved in Fig. 3).

The relatively high correlation between mass of liver and that of the fish (it is about 0.83), in addition to the moderate variances, helps secure that the difference between  $\widehat{\text{HSI}}_{\text{ind}}$  and  $\widehat{\text{HSI}}_{\text{bulk}}$  is fairly small. Simulation studies indicate that there is a nearly linear relationship, with perfect correspondence in HSI, if the correlation is high, but also that the  $\widehat{\text{HSI}}_{\text{ind}}$  has rather higher

**Fig. 2.** Gamma distribution densities fitted to full-mass and liver mass data (both in kg), and with correlation 0.82; all applied data refer to Northeast Arctic cod (skrei), classified from otolith readings (Mjanger et al. 2006).



values compared with  $HSI_{bulk}$  when the correlation is small. This points again to the importance of understanding the actual target for estimation and also a need for additional fundamental analyses for different strata of populations. This might be used for establishing more general translation formulae than that of eq. 4.

### Statistical modelling of the HSI time series

In the following, we will introduce different statistical models to study and analyse the bulk liver quality index  $HSI = HSI_{bulk}$  of Fig. 1. A good statistical model allows us to answer various pertinent questions, such as predicting the future behaviour of the HSI, checking for anomalies, discovering potential outliers, and investigating structural changes or regime shifts in observed series. Moreover, a full stochastic model makes it possible to study different joint relationships and interactions among our main target for inference, the HSI, and several explanatory series, like the Kola temperatures, the length distribution of the population, the mortality rate, and the food supply, as briefly pointed to in the Introduction above and further specified below.

Before we discuss and examine these covariate series, we carry out a separate investigation of the bulk HSI series from Fig. 1 in itself. To do so, let  $z_i$  represent liver quality index HSI for year  $i$  and consider the model where

$$(5) \quad z_i = \beta_0 + \beta_{year} t_i + \varepsilon_i$$

where  $t_i$  relates to calendar year  $i$ , from 1859 to 2012, for convenience defined here as year  $i - 1880$ , and  $\varepsilon_i$  taken to be a stationary zero-mean Gaussian time series. Typically, this stationary part of eq. 5 will be modelled as a low-order autoregressive process. For the  $k$ th-order autoregressive model, there is a representation

$\varepsilon_i = \rho_1 \varepsilon_{i-1} + \dots + \rho_k \varepsilon_{i-k} + \sigma \delta_i$ , where the noise terms  $\delta_i$  are independent and standard normally distributed (some alternative and more general types of models will be discussed below). Without going into the details, we point out that some technical conditions on the  $\rho_j$  parameters are needed to ensure that the resulting model is indeed stationary; see Brockwell and Davis (1991, chapter 3) or Brillinger (1975) for introductions to autoregressive models as well as to general time series modelling.

The stationary part of the model (eq. 5) entails lagged dependencies between consecutive years of the HSI and is by the simple structure of the autoregressive models not difficult to interpret. The residuals  $z_i - \beta_0 - \beta_{year} t_i$  depend linearly on the previous values, with the degree to which this happens dictated by the size of the autoregression coefficients. We also take the opportunity to point out that the class of autoregressive models is able to approximate any type of stationary dependency structure (Brockwell and Davis 1991, chapter 4.4).

In the model given in eq. 5, we have also included a linear drift or trend, and as mentioned more complex (and perhaps more realistic) relationships including covariates will be discussed later. Below we will question whether  $\beta_{year}$  is zero or not. If  $\beta_{year}$  is significantly smaller than zero, for example, it might have important implications for the future behaviour and the general understanding of the liver quality index, and indeed for the fish population itself.

Figure 5 sheds light on the predictive ability of the model of eq. 5, for the case of 1-year-ahead predictions for the HSI series, using an autoregressive model of order two for the  $\varepsilon_i$ . It demonstrates that the mean prediction error, say  $\hat{z}_i - z_i$  when using  $\hat{z}_i$  to predict  $z_i$ , is smaller in size when deriving predictions via the autoregressive model than when using the mean of the three previous values. This is a pertinent comparison, as the latter

**Fig. 3.** Simulated per-fish and bulk HSI values using the model of eq. 3, as per eqs. 1 and 2. The respective mean and standard deviations are 5.84 and 0.03 for the HSI per-fish distribution and 6.17 and 0.03 for the HSI bulk distribution. The correlation is 0.83.

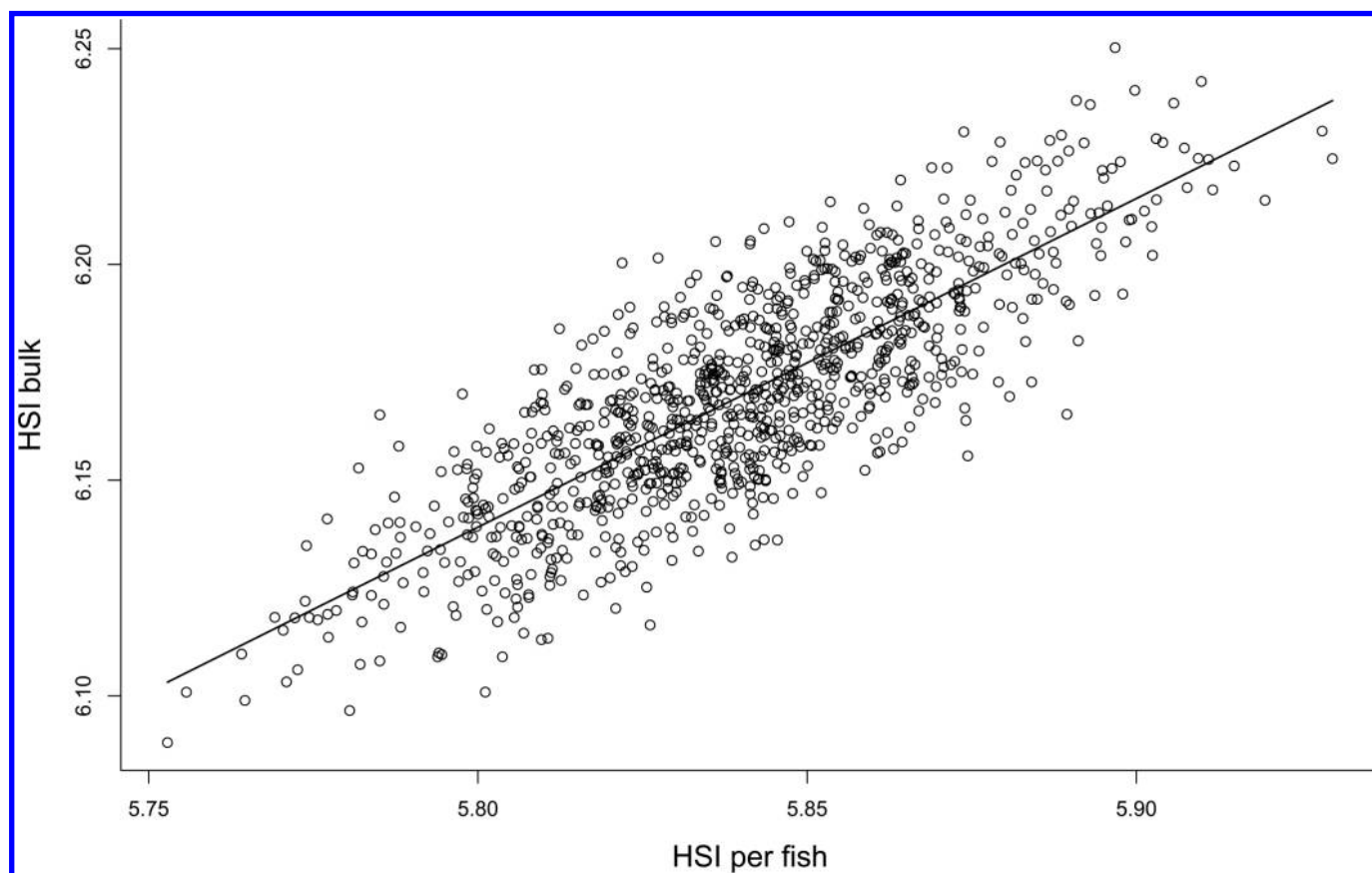
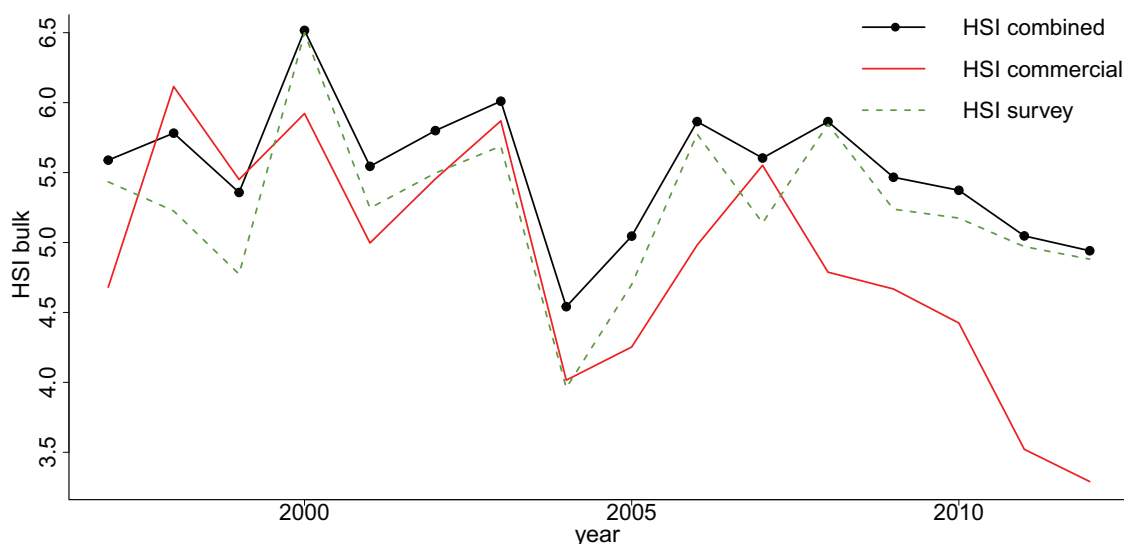


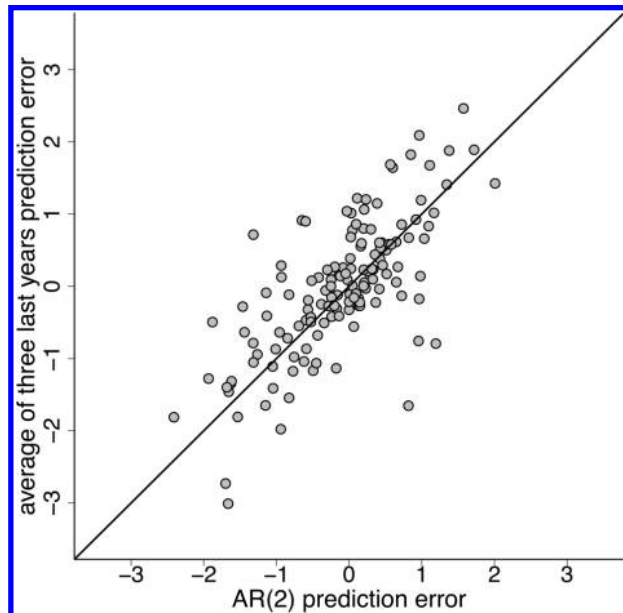
**Fig. 4.** Expanding on data from 1997 to 2012 of Fig. 1, this figure displays both the HSI bulk index from the commercial data (solid line, no circles) and the HSI individual index from the survey data (dashed line), along with the optimal prediction of the HSI bulk index based on these series (solid line with circles), using confidence distribution theory.



method is used with ICES forecasting routines for catch and stock mass of NEA cod (ICES 2014). Also, other plots and tests (not reported in detail here) show that the autoregressive model of this order does well and typically better than those for other competing models. One particularly useful plot is that of  $m_t = \Phi(d_t)$ , where  $d_t = (z_t - \hat{z}_t)/\hat{\sigma}_t$ , involving the standardized prediction error associ-

ated with  $\hat{z}_t$  and with  $\Phi$  the standard normal distribution function. The idea is that if the model used to derive the prediction formula is good, the  $d_t$  numbers will be close to having a standard normal distribution, so that the  $m_t$  values will be close to uniformly distributed on the unit interval. If the model does not fit well, the  $m_t$  values will tend to be closer to either zero or one. Our plots of this

**Fig. 5.** Sequential 1-year-ahead predictions (1870–2012) from an autoregressive model of order two compared with the predictions obtained by using the mean of the HSI of the 3 preceding years as a predictor. On average, over the period of 143 years, using the autoregressive model results in a mean absolute error of size 0.66 and root-mean-squared error (RMSE) of 0.86, while the simpler mean of the preceding years gives an absolute error of 0.73 and RMSE of 0.99. This illustrates the potential gain by building more complex statistical models.



kind for different competing models support using the autoregressive model given in eq. 5 of order two.

### Is the HSI time series stationary?

Analysing the complete  $HSI_{bulk}$  time series of Fig. 1 using standard tests for stationarity, there is little evidence suggesting that the HSI does not satisfy the conditions for being stationary. This depends, however, to some extent on the time window considered. A Dickey–Fuller-type test (Dickey and Fuller 1979), for example, used on the full HSI series 1859–2012, rejects the null hypothesis (with a  $p$  value less than 0.01) that the series has a unit root. Along with graphical diagnostics, this provides evidence that the Hjort series is well modelled via low-order autoregressive models. By making sequential AIC analyses, in the same spirit as with the monitoring bridge plots reported in connection with Fig. 6, however, we have a less conclusive picture and observe that the effect of including the linear  $\beta_{year} \times year_i$  in the model has major importance for several long periods of time. We will therefore keep the linear component in our baseline model for now; this also serves as a minimal model to attempt to improve upon when we introduce more complex models and covariates in later sections.

While the brief considerations above concerned the stationarity or not of the mean function, a different aspect of the HSI series is its variability level around the mean function. Visual inspection of the series might indicate a nonconstant level of variability. The standard deviation parameter examined now, say  $\sigma_t$  at year  $t$ , is that associated with the distribution of  $HSI_{bulk}$  over time, not that of the precision of an individual data point. That the  $\sigma_t$  is not constant over time is indeed borne out of careful estimation and testing procedures, indicating a certain decrease in variability level during the years 1955–1990. The drop in standard deviation appears to stem from after 1950 (Fig. 7), at a time when the use of trawling in the Barents Sea flattens out at a high level (Høyen et al.

2008). This drop in  $\sigma_t$  is also associated with generally higher exploitation rates (Jørgensen 1990; ICES 2014; Kjesbu et al. 2014b), causing an abrupt fall in overall body size (Jørgensen 1990; Kjesbu et al. 2014a) and thereby in HSI, which is positively size-dependent (Kjesbu et al. 2014a). The same period shows up in other graphical checks as moderate irregularities and is also reflected in models we have investigated of the time-varying coefficients type (see our concluding remark D below). There appears to be an increase in variability of the recorded HSI again, around the end of the 1990s. This might be due to the gradual reconstruction of the stock demography up to recent time (Kjesbu et al. 2014b) for the time point 1990 preceded by a temporally sharp drop in fishing mortality.

For the illustrative purpose of the present section, the deviation is not too severe, and approximating the standard deviation function with a constant will not become statistically troublesome, for most purposes (see the appropriate FIC discussion below).

### Regime shifts and structural changes

There are occasions where some statistical model is in force for a number of years, after which the underlying parameters change dramatically and perhaps rapidly, pushing certain aspects of interest into a new state of equilibrium. Such regime shifts have arguably taken place regarding the ways in which the Kola winter temperature and the HSI series and their interplay have developed. For a general biological account of regime shifts in relation to climate changes, also of relevance for the NEA cod, see Brander (2010) and Vasilakopoulos and Marshall (2015).

There are various statistical techniques devoted to studying such phenomena, from testing the null hypothesis that no significant change has taken place during a certain time period to estimating the position of a break point in case such a discontinuity has taken place; see Frigessi and Hjort (2002) for a general discussion and overview and Hjort and Koning (2002) for a class of such methods. Here, we briefly outline one particular method, associated with a certain graphical plot for checking constancy and looking for break points. Assume a certain model is put to work, involving a parameter vector  $\theta = (\theta_1, \dots, \theta_p)$  of length  $p$ , leading in particular to the model-based log-likelihood function  $\ell_j(\theta)$  associated with years  $1, \dots, j$ . Thus, we may for each time period 1 to  $j$  compute the maximum likelihood estimate  $\hat{\theta}_j$  and associated log-likelihood maximum value  $\ell_{max,j}$ , say, as long as  $j \geq p$ . From the full sequence of observations, over years  $1, \dots, n$ , we may thus monitor both how the parameter estimates and the log-likelihood maxima develop over time. One particular monitoring function is then

$$(6) \quad B_{n,j} = \sqrt{n} [n^{-1} \ell_{max,j} - (j/n) \hat{a}] / \hat{\kappa} \quad \text{for } j = p, \dots, n$$

where  $\hat{a} = n^{-1} \ell_{max,n}$  and  $\hat{\kappa}$ , the latter an estimate of  $1/\sqrt{n}$  times the standard deviation of  $\ell_n(\theta)$ , are both based on the full dataset. Note that the process ends in  $B_{n,n} = 0$ . The point now is that (i) if the underlying model in fact has not changed after all, then the monitoring bridge process  $B_{n,j}$  of eq. 6 behaves for large  $n$  as a so-called Brownian bridge, with controlled fluctuations; and (ii) if there is a break point, the plot will help in identifying its position.

Figure 6 provides an application of this method, associated with the autoregressive model

$$HSI_i = \beta_0 + \beta_1 x_{i-1} + \sigma \varepsilon_i \quad \text{for 1921 to 2012}$$

where  $x_{i-1}$  is the mean Kola winter temperature from the year preceding  $HSI_i$  (cf. further discussion below in the section on Covarying factors and further models), and the  $\varepsilon_i$  for this illustration follows an autoregressive model of order one. The two series are shown in Fig. 8. A Brownian bridge stays within  $\pm 1.358$  with probability 0.95, so any  $B_{n,j}$  values observed outside this band indicate that the underlying model has not stayed constant over the time

**Fig. 6.** Monitoring bridge plot for the model  $HSI_t = \beta_0 + \beta_1 x_{t-1} + \sigma \varepsilon_t$  for the  $HSI_{bulk}$  series, with  $x_{t-1}$  the previous year's mean winter Kola temperature and the  $\varepsilon_t$  a standardized AR(1) process. The plot suggests there is a regime shift around year 1990, with model parameters taking on new values.

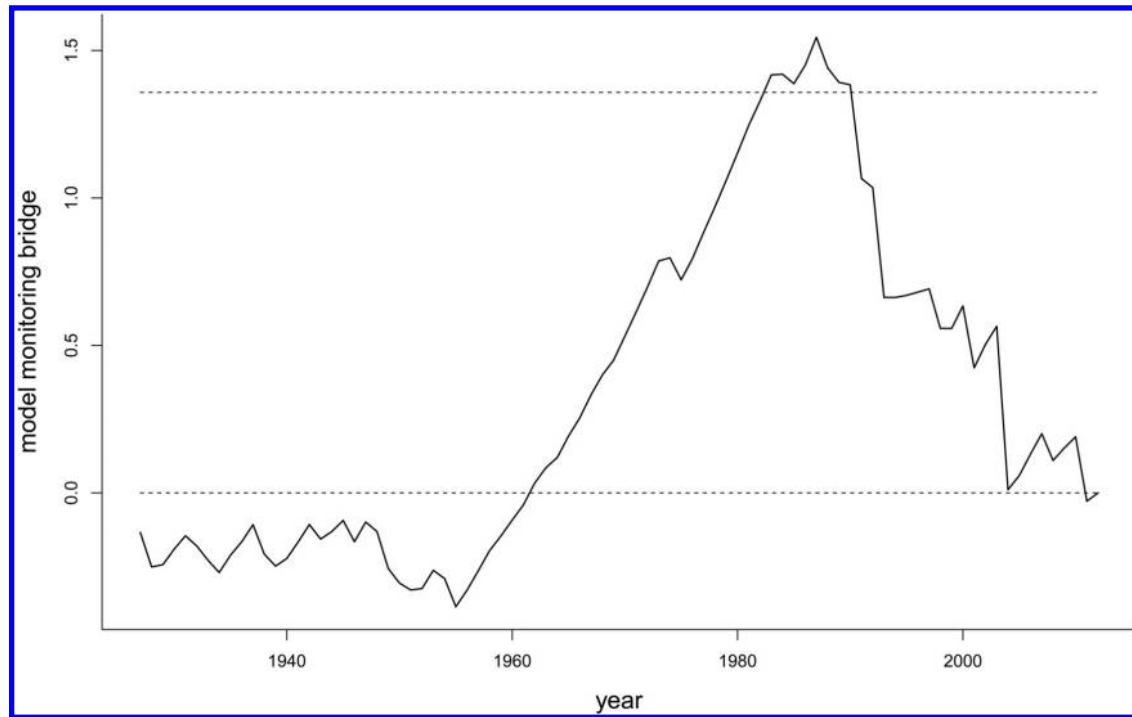
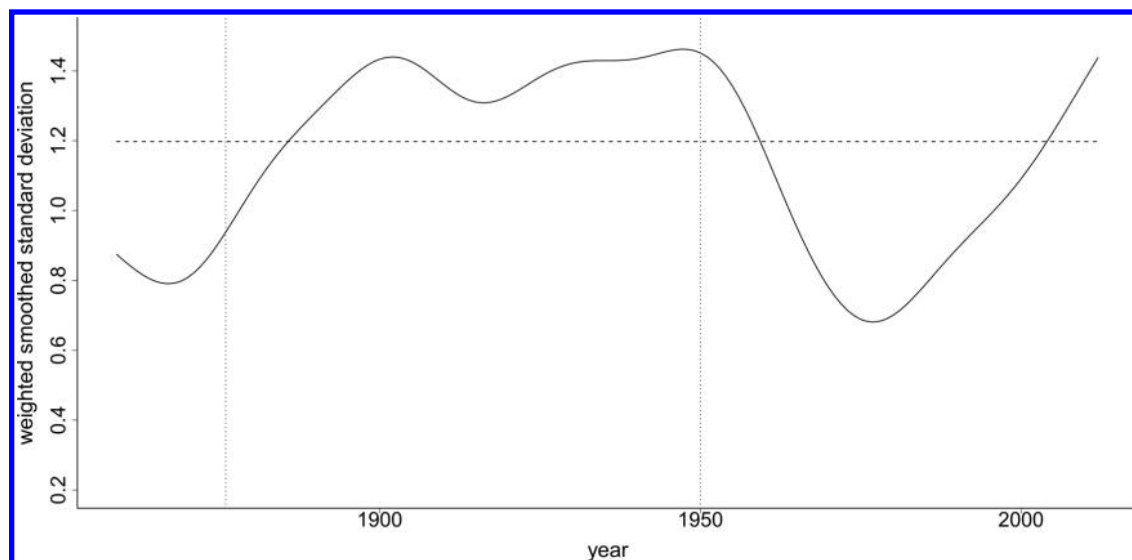


**Fig. 7.** Estimated standard deviation for the HSI bulk series, demonstrating a varying level of variability over time. The main drop in variability starts after 1950 (see the text). The other vertical line shows the start of the observed HSI at year 1876. The horizontal line indicates the overall standard deviation estimate using the complete series.

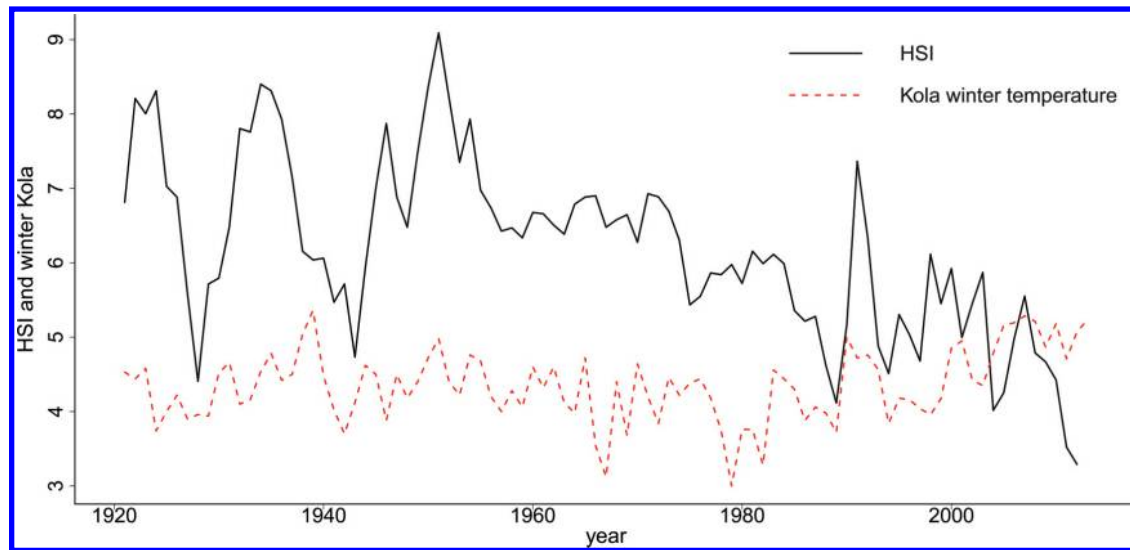


window considered (corresponding to testing the hypothesis of a constant model with significance level 0.05). Here, the maximum value is indeed higher than 1.358, as indicated in the figure. The data hence suggest there is a regime shift for the model around year 1990, with the model switching parameter values around that time. More information on the statistical methodology used here, along with further applications, can be found elsewhere (C. Cunen, G.H. Hermansen, and N.L. Hjort, unpublished data).

This suggestion of a “regime shift” (or alternatively “state shift” or “critical transition”) has been discussed also in marine ecology

(see Brander 2010; Vasilakopoulos and Marshall 2015 and references therein). The term is characterized by “relatively rapid change (occurring within a year or two) from one decadal-scale period of a persistent state (regime) to another decadal-scale period of a persistent state (regime)” (King 2005). Apparently,  $HSI_{bulk}$  and Kola temperature patterns started to depart in the 1980s (Kjesbu et al. 2014a, their figure 6: 9-year smoothed mean). Vasilakopoulos and Marshall (2015) identified a regime shift having taken place in 1981 using principal component analyses on 13 NEA cod population descriptors (including HSI) and five “stressors”

**Fig. 8.** The HSI series for 1921–2012 (solid line), along with mean Kola winter temperature (dashed line, in degrees Celsius).



(including Kola temperature). Hence, the outcome very much depends on the level of collated information and the statistical analyses undertaken. Nevertheless, according to these authors, the general regime shift in 1981 was largely driven by the combined effect of low temperature (Fig. 8), high  $F$  (see further below), and low stock size (cf. for instance figure 8 in Kjesbu et al. 2014a). It is noteworthy that the HSI for that time was rather typical (Fig. 8). Clearly, in the context of anthropogenic climate change, these types of studies will be paramount to follow up, especially in current high-latitude waters, expected to warm the most in the future, although to which degree is highly dependent upon future greenhouse gas emission pathways (Stocker et al. 2013).

## Model selection

Selecting an appropriate model is an important and integrated part of the statistical inference process. In many or most situations, the statistician will have more than one reasonable candidate for modelling the phenomena under study. Selecting an appropriate model for the final analyses and report is therefore of utmost importance, hence necessitating the notion of and research field of statistical model selection.

Although not explicitly pointed to, parts of the above discussion regarding the validity of the model and reliability of the underlying assumptions (e.g., checking for nonstationarity in the HSI series and examining potential trends) are actually questions related to model selection. Selecting the “best” model among a set of potential and reasonable candidates has an ongoing and long history, with techniques ranging from visual inspection, goodness-of-fit testing, and so-called model information criteria; for a general introduction to these themes see Claeskens and Hjort (2008).

Models with many parameters may become ineffective because of estimation variability, whereas slimmer models with fewer parameters might suffer from modelling bias. A “good” model selection strategy should balance complexity against simplicity and precision in a reasonable way. Models should be as simple as possible, but not simpler, as Einstein implied. (The exact quote from Einstein (1934) is actually “It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.”) The preferred model should be rich enough to capture the essential features and with high enough precision to be useful and at the same time still be simple enough to be comprehensible and possible to handle regarding fitting and inference techniques.

Different model selectors balance the desiderata of “low variance” and “low bias” in different ways.

## The AIC and the BIC

Among the more popular model selection strategies are Akaike’s information criterion (AIC; Akaike 1998) and the Bayesian information criterion (BIC; Schwarz 1978); also see the general survey by Claeskens and Hjort (2008), which also covers various fine-tuned versions and other variations, including links to cross-validation. The AIC and BIC have considerable appeal, since they are typically simple in both structure and use, resulting in model scores that can be used to rank candidate models from best to worst in accordance with a well-defined measure of discrepancy. The practical simplicity of AIC and BIC has perhaps led to uncritical use. This is especially true for the AIC, which is often used without any concern for the underlying motivation (see Hermans and Hjort 2014).

The AIC is defined as  $AIC = 2 \log\text{-lik}_{\max} - 2p$ , with  $\log\text{-lik}_{\max}$  the maximal value of the log-likelihood function and  $p$  the number of parameters used in the model. The BIC has a similar structure, viz.  $BIC = 2 \log\text{-lik}_{\max} - p \log n$ , with  $n$  the number of observations, but stems from a quite different motivation, to be commented on in a moment. Both lead to one “best” model; the AIC aims at the one minimizing a certain Kullback–Leibler divergence from the underlying true data-generating mechanism to the model in question, whereas the BIC aims at the model maximizing the posterior model probability in a Bayesian framework. These are global perspectives, preferring models that aim at capturing the main characteristics of the underlying data-generating process. The BIC is more parsimonious, tending to select models with fewer parameters than those favoured by the AIC. Generally speaking, and for large  $n$ , the BIC does better when it comes to finding the true model (if the data-generating process is captured by one of the candidate models), whereas the AIC does better regarding prediction and accurate estimation (Claeskens and Hjort 2008, chapter 4).

To illustrate the use of these criteria, consider modelling the HSI series as  $z_t = \beta + \sigma \varepsilon_t$ , with the  $\varepsilon_t$  forming a zero-mean standardized normal autoregressive process AR( $k$ ) of orders 0 (corresponding to independence), 1, 2, 3, 4, and 5. The parameter dimensions of these models are respectively 2, 3, 4, 5, 6, and 7 (since  $\beta$  and  $\sigma$  are parameters to be estimated for each of them). Table 2 gives the AIC and BIC scores, yielding in particular two different suggestions.

**Table 2.** AIC and BIC scores when fitting stationary autoregressive models of orders 0, 1, 2, 3, 4, and 5 to the HSI bulk index time series of Fig. 1.

|           | <i>k</i> |         |         |        |        |        |
|-----------|----------|---------|---------|--------|--------|--------|
|           | 0        | 1       | 2       | 3      | 4      | 5      |
| Dimension | 2        | 3       | 4       | 5      | 6      | 7      |
| AIC       | -495.5   | -363.7  | -363.5* | -365.0 | -367.0 | -368.9 |
| BIC       | -501.6   | -372.9* | -375.6  | -380.2 | -385.2 | -390.2 |

**Note:** The best models are indicated with an asterisk. The order zero model, corresponding to independence, is judged too simple by both criteria. The BIC has the AR(1) as the best model, whereas the AIC scores suggest that AR(1) and AR(2) are about equally good, with a slight preference for the latter.

We have also worked with certain dynamic plots and associated statistical model determination techniques (though not included in our paper), displaying relative AIC scores for models AR(1) to AR(5) as these progress with more data accumulated over time. The relative AIC score in question is  $AIC_{AR(k)} - AIC_{AR(2)}$ . These and similar plots and tests indicate firstly that higher-order models do not contribute significantly and secondly that something noteworthy takes place around 1903 and around 1990, with a drop in exploratory power. This is associated with the all-time record low value for the HSI in 1903 and the above-mentioned regime shift, respectively.

### The focused information criterion (FIC)

Instead of aiming at a model that is “reasonably good at everything”, the motivation underlying the FIC is that the intended use of the model and the focus of the investigation should play a central part of the selection procedure. Only rarely are both aspects good for all purposes. This is, for example, evident for regression models, where some covariates may be important for some types of questions but of lesser importance for other aspects of what is being studied. Compared with other classical information criteria, the FIC allows the precise intention of the analysis to be taken into account when selecting the model. The FIC sidesteps the often unachievable goal of finding one “correct” model for all uses and aims instead at finding the model that is best suited for answering focused questions, one focus at a time.

The FIC was introduced in Hjort and Claeskens (2003) and Claeskens and Hjort (2003) and is based on estimating and comparing the accuracy of individual model-based estimators for a chosen focus parameter, which we here denote  $\mu$ . The focus  $\mu$  ought to have a clear statistical interpretation across candidate models. For a given candidate model,  $\mu$  is then expressed as a function of this model’s parameters. For the model of eq. 5, the parameters are  $\beta_0$ ,  $\beta_1$ , and  $\sigma$ , along with the autoregressive parameters  $\rho_1, \dots, \rho_k$ , specifying the dependency structure of the stationary process  $\{\epsilon_i\}$ . Valid and relevant focus parameters include quantiles, regression coefficients, a specified lagged correlation, and various types of predictions and data-dependent functions, like the probability that the future HSI will be below a given threshold (say the 3.0 value reached in the year 1903), given the observed history of liver quality values.

Suppose there are candidate models  $M_1, \dots, M_k$ , leading to focus parameter estimates  $\hat{\mu}_1, \dots, \hat{\mu}_k$ , respectively. The underlying idea leading to the FIC is to estimate the mean squared error (MSE) of  $\hat{\mu}_j$  for each candidate model  $M_j$ , for  $j = 1, \dots, k$ , and prefer the model that achieves the smallest value. The MSE in question is

$$(7) \quad r_j = E(\hat{\mu}_j - \mu_{\text{true}})^2 = \text{Var } \hat{\mu}_j + \text{bias}(\hat{\mu}_j)^2$$

comprising the variance and the squared bias in relation to the true parameter value  $\mu_{\text{true}}$ . Thus, the FIC consists of finding ways of assessing, approximating, and then estimating the  $r_j$  for each candidate model, and the winning model is the one with smallest  $\hat{r}_j$ . How

this may be done depends on both the candidate models and the focus parameter, as well as on other characteristics of the entire modelling framework. The FIC apparatus hence leads to different types of formulae in different setups (see Claeskens and Hjort (2008, chapters 5 and 6) for a fuller discussion, illustrations, and generalizations). For time series models, as encountered when modelling the HSI series and its covariate series, certain complexities are involved (see Hermansen and Hjort 2015b). Importantly, the FIC may lead to different models being pinpointed as best for different foci. An FIC package for the statistical software package R ([www.r-project.org](http://www.r-project.org)) is currently under development; preliminary R scripts for FIC analysis can be obtained from this article’s corresponding author.

We will now apply the FIC strategy for the HSI study. We consider eight natural candidate models, corresponding to autoregressive orders 1, 2, 3, and 4, with or without a linear trend component over time. The focus function we consider first is

$$\mu_{\text{pred}}(\text{year}) = Z_{\text{year}} \quad \text{for year} > 2012$$

and more specifically we aim for this illustration at models that are good at predicting 1, 3, and 10 years into the future (i.e., 2013, 2015, 2022, as seen from year 2012, the currently available endpoint of the HSI bulk series). Carrying out the FIC step aids in understanding the short-, medium-, and longer-term mechanisms involved in the HSI process. The results are shown in Fig. 9. These FIC plots display predictions on the y axis and root-FIC scores (i.e., estimated root-mean-squared errors  $\hat{r}_j^{1/2}$ ) on the x axis.

For the 1-year-ahead predicted HSI value of 2013, we observe that essentially all models are in reasonable agreement (apart from the too simple model, which uses independent errors; see below). This is also reflected by the small differences in the estimated root-mean-squared error values. For such large sample sizes ( $n = 154$ ) and such relatively simple focus questions, we do not expect the quality of the estimates to deviate considerably (as long as all fitted models are all fairly reasonable).

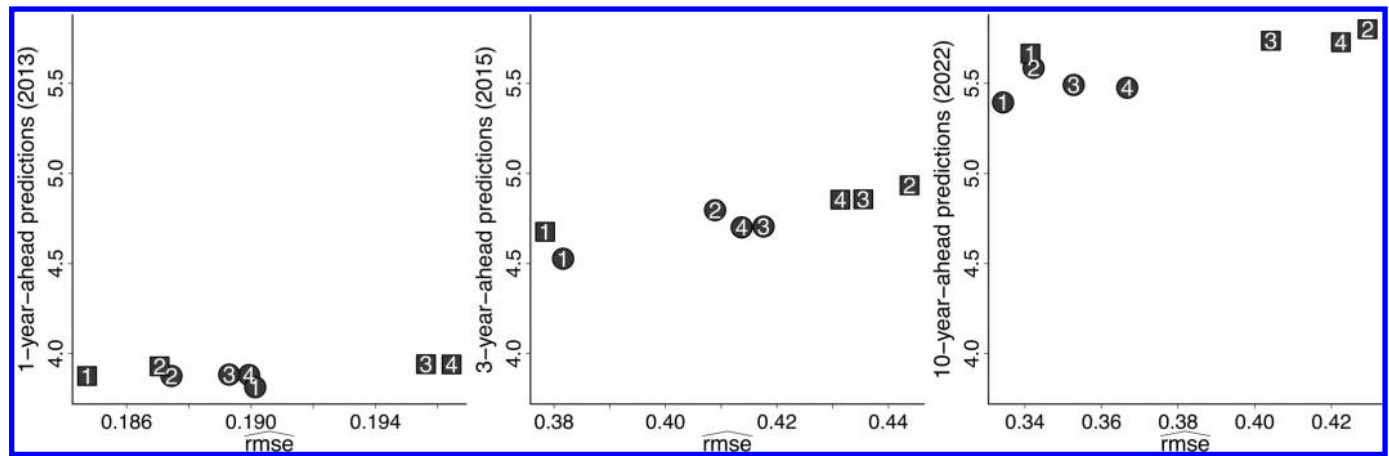
It is worth pointing out that the FIC judges the autoregressive model with linear trend to be the best model for predicting the liver quality 10 years ahead (here 2022). This is in contrast with observations made in the discussion of stationarity of HSI earlier, where the model with linear trend seemed to be inferior to the simpler stationary model with a constant trend. As a general observation, we see that as we try to predict further and further into the future, the models with a linear trend start to dominate among the “best” models. This suggests that there might be a (significant) linear effect needed to explain the long-term behaviour of the HSI. Further illustrations and discussion of the FIC are given below. The models with independent errors (referred to also as autoregressive models of order zero) are not included in the plot. These resulted in estimated root-mean-squared errors of more than 1.5, falling outside the natural scale spanned by the other and better models in Fig. 9.

There is even more room and need for FIC-type model building and selection tools when covariate information is taken into account, as we shall see in the following section. With covariates included, the number of natural candidate models also increases rapidly. In many cases, one would also wish to construct a model working well for several focus parameters conjointly, say for predicting ahead at different time spans or for analysing influence of a covariate for a certain range of covariate values. For such purposes, the FIC methods have been extended to an appropriate average-weighted FIC, called the AFIC (see Claeskens and Hjort 2008, chapters 6–7).

### Covarying factors and further models

In this section we briefly report on investigations on whether certain factors may be seen to be significantly correlated with the

**Fig. 9.** FIC plots for the predicted liver quality index, 1, 3, and 10 years from now, with FIC scores (i.e., estimated root-mean-squared error) on the x axes. The models are autoregressive models of orders 1–4 (indicated by the number inside each point), with (circles) and without (squares) a linear trend. The AIC and BIC, which do not differentiate between the different foci, both prefer the autoregressive model of order two without the linear trend.



HSI series. In cases where there is such an identifiable correlation, further considerations and analyses might be called for when it comes to determining what causes what. As pointed out in our Introduction, and as further detailed in Table 1, the explanatory variables we consider are (i) Kola temperatures, (ii) mean length distribution, (iii) fish mortality  $F$ , and (iv) food availability (ICES 2014), the ratio of biomass for capelin with spawning stock biomass for the NEA cod.

As a curiosity, since Hjort (1914, p. 186) found it necessary to briefly dismiss the hypothesis apparently put forward by Helland-Hansen and Nansen (1909, chapter VII.3) that the annual sunspot numbers could influence the liver quality of cod, we will also compare the sunspots series with the  $HSI_{bulk}$ . In addition to studying each of these and their respective connection to the Hjort series on their own, we will also use the forthcoming discussion to select some of the covariate series for a further combined analysis.

#### (i) Kola temperatures

In Kjesbu et al. (2014a), the connection between the annual Kola temperature (1900–2012) and the  $HSI_{bulk}$  was studied. The two time series exhibit a covarying pattern lasting for many years, but the apparently strong relationship seems to have ended somewhere after 1960. We commented on this in connection with Figs. 6 and 8. In the following section, we will continue this discussion and also try to unveil where the separation takes place.

Instead of working with the annual mean temperature, we will use what we define as the mean winter temperature, averaging the monthly means from start of October (previous year) to start of March (current year). The data used are monthly means of Kola temperatures from 1921 to 2012. These winter months may carry more relevant information, since it is during this period the cod recuperates for spawning.

By studying the effect of taking different time lags (i.e., the relationship between the HSI for a given year and various combinations of preceding predictors, in this case the mean Kola winter temperature), we observe that the mean winter temperature of the previous winter carries more relevant information and also provides a more important signal than the effect from the current period. This is not evident if the raw correlations are used as a measure of “information quality”, however. We note that the raw correlation between last Kola winter temperature and  $HSI_{bulk}$  is  $-0.02$  and for the last annual temperature is  $-0.03$ . The corresponding correlation estimates with the current year are  $-0.07$  and  $-0.15$ , which might suggest that the annual temperature of

the same year is the most relevant. These correlations are all on the tiny side of the spectrum, however.

#### (ii) Length

The mean length series of NEA cod (1932–2012) is shown in Fig. 10. The raw correlation between the HSI and the length series is about 0.51, which is quite high and in fact is the largest among the explanatory variables considered here. This also suggests that it should be a good predictor, since the correlation is essentially a measure of linear relationship. Also, by carrying out analyses similar to that presented with the monitoring bridge plots of Fig. 6, we observe that as a predictor for HSI, the length mean for the same year carries the most information about the liver quality index. This is in contrast with the other series discussed here, where all are seen to carry the most information about the current status of the HSI if the value for the previous year is used as an explanatory variable.

There is growing evidence that fluctuations in expressions of body condition (e.g., HSI) and reproductive investment are not only dependent upon the current environmental situation, but that the influence also dates further back in time, in line with statements in Stearns (1992). This should in particular be true for the present capital breeding NEA cod undertaking long spawning migration, which may show high levels of omission of spawning (i.e., skipping; Skjæraasen et al. 2012).

Unfortunately, there is a notable gap of missing values (1973–1979) in the length series, which renders complete joint analysis impossible (we cannot use observations prior to 1980). To overcome this, we reconstruct the missing values using the same methodology presented in the discussion on combination of information; see the discussion below.

The observed length series (Fig. 10) indicates that there might be a change in the underlying model around 1965. From a statistical model building perspective, discontinuities or change points in the underlying model may have critical effects on the final analysis, inference, and the validity of the conclusions made. Here, the effect of the jump is perhaps further highlighted by the fact that there seems to be an increasing trend prior to the apparent regime shift around 1965. There are various tools constructed to find and work with models that have jump discontinuities (see, e.g., the jump information criterion JIC of Grønneberg et al. (2015)). Since we are not studying the length series on its own, but are mostly interested in the effect on the liver quality index and are therefore merely using it as a covariate, the potential discontinuity point (also compared with the overall variation) will not violate

**Fig. 10.** The mean length time series with reconstructed length values for the period 1973–1979. The observed data suggest that there might be a jump discontinuity or regime shift around the mid-1960s. The horizontal dashed lines show the estimated mean signal before (lower line) and after (upper line) the potential shift.

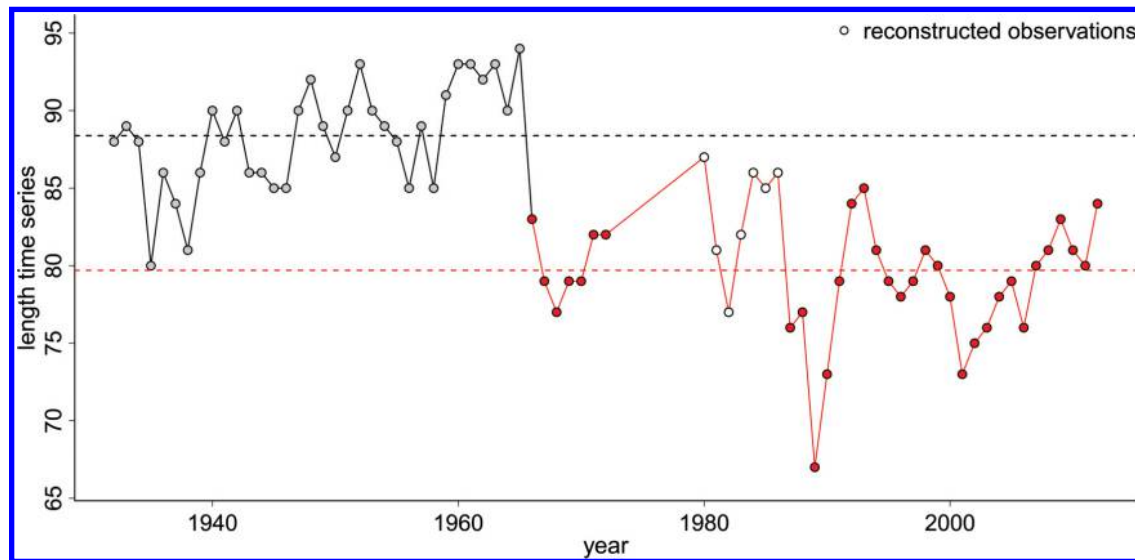
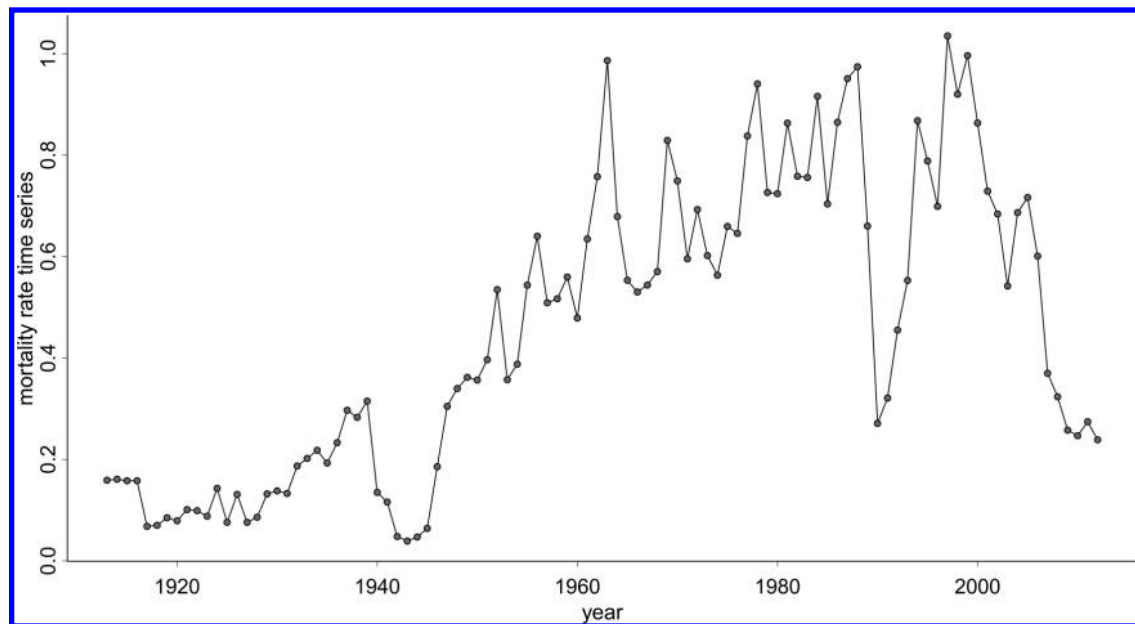


**Fig. 11.** The mortality rate time series. The long persistence in the series and almost linear relationship over time may signal long-range dependencies.



the underlying assumptions to an extent necessary for introducing more complex modelling tools.

### (iii) Mortality rate

The third long series we consider is the mortality rate  $F$  for the NEA cod. It has a strong and almost linear persistence in time (see Fig. 11). The raw correlation with  $HSI_{bulk}$  was  $-0.19$ , which is quite high in absolute value compared with the others series we consider. It is, however, not clear whether the effect will be present after correcting for a linear trend. If analysed on its own, subtracting a linear trend makes the resulting residual series look like white noise. In our analyses, the role of the  $F$  series is to be used as one of several predictors for HSI, and it is unproblematic if the  $F$  series is not stationary over time. Studying relevant graphical

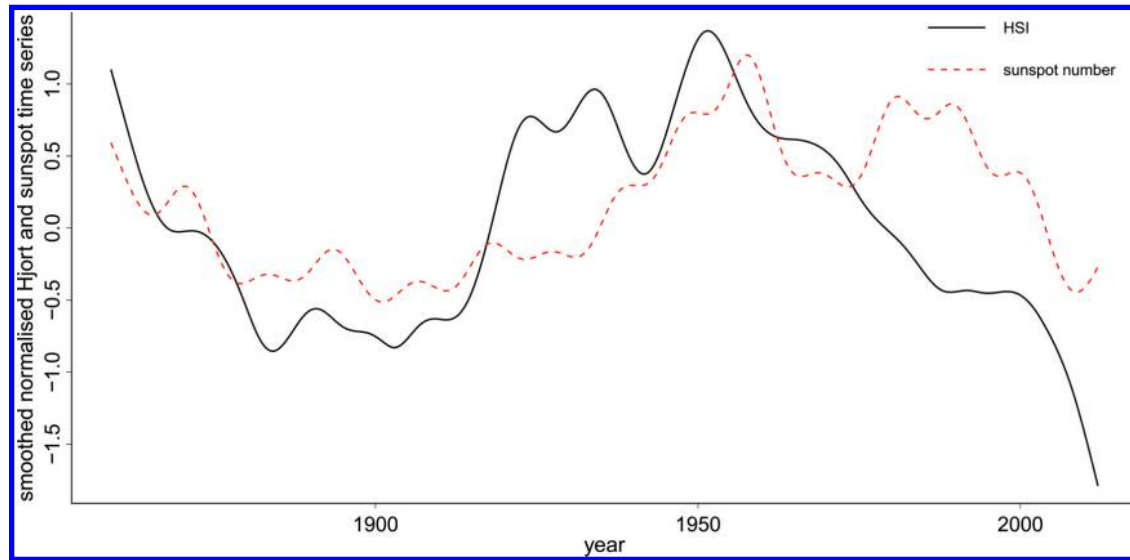
plots indicates again that last year's mortality rate is of more importance than that of the current year.

### (iv) Capelin index

As a proxy for the amount of available food, we will use the biomass ratio estimated from the years 1980 to 2012. From univariate analyses, we learn that a log-transform of this food proxy improves the prediction quality with respect to the HSI. This seems to be reasonable, since relative changes in the food supply is not the same when there is an insufficient amount of food as when in abundance. This also results in a raw correlation of size  $0.56$ , indicating a strong linear relationship and a reasonable predictive ability.

An autoregressive model of order three gives a good fit to the capelin series. This suggests using three distinct lags as predictors

**Fig. 12.** The Hjort HSI<sub>bulk</sub> time series (solid line) with annual sunspot numbers (dashed line) for 1880 to 2012, both standardized to have mean zero and unit standard deviation. The estimated correlation is 0.11. This and related tests and graphs show little or no signs of any underlying relationships.



for the HSI. Including more than one lag, however (i.e., the amount of food last year), does not improve on the model. This is a reminder that preliminary results associated with univariate analyses may be misleading when building joint multivariate models; see the discussion below.

#### (v) Annual sunspot numbers

The multitasking national hero, Fridtjof Nansen (polar explorer, athlete, writer, artist, diplomat, winner of the Nobel Peace Prize), was also a broadly oriented scientist, publishing in zoology and oceanography. With physical oceanographer Bjørn Helland-Hansen, he published the ambitious *The Norwegian Sea: Its Physiological Oceanography, Based Upon the Norwegian Researches 1900–1904* (Helland-Hansen and Nansen 1909), which also recorded various time series pertaining to deep-sea temperatures, salinity, density, seasonal variations, the Polar currents, etc. The authors also studied how oceanographic and other parameters could be related to the growth and spawning of food-fishes and indeed also examined the liver index. They also appear to argue that such measurements are related to and perhaps causally influenced by “the periods of Sun-spots” (chapter VII.3). Such ideas of the annual sunspots exerting influence on the earth’s climate and biology had also been examined and speculated over by other scholars, such as Ljungman (1879) (see the account of Smith (1994, chapter 5) reviewing this historic period, and also Lindquist (2002); Yndestad (2009)).

Hjort (1914, p. 186) was, however, sceptical towards such viewpoints. He obtained the sunspot numbers for the period 1880–1911 in question from Otto Pettersson, displaying them along with the liver index series in his Fig. 116, and commented, “they do not, however, by any means coincide”, going on to state his disagreement: “The only warrantable conclusion would seem to be, that no relation can be shown to exist between the two phenomena, in any case not for the present, nor in the way suggested by Helland-Hansen and Nansen.” Incidentally, Jølle (2011, Part 4), in his illuminating biography of Nansen, records and comments on several clashes between Nansen and Hjort, regarding matters of both research administration, how to conduct science, and how to communicate research questions and findings to the general audience.

A hundred years later we may complement the sunspot analysis above, examining the two time series in question for 1859 to 2012 (Fig. 12). We find no clear statistical relation and side with Hjort over Nansen.

#### Multivariate models and multiple covariates

Let  $t_i$  denote year  $i$ , for convenience here taken as calendar year minus 1980, and let further  $x_{1,i}$ ,  $x_{2,i}$ ,  $x_{3,i}$ ,  $x_{4,i}$  be year  $i$  observations for the four series discussed above (i.e., mean winter Kola temperature, total length, mortality rate, and the capelin index as a proxy for the food availability, respectively). With the univariate analyses above in mind, a natural candidate for a joint model for the liver quality index  $z_i$  is now

$$(8) \quad z_i = \beta_0 + \beta_{\text{year}} t_i + \beta_{\text{kola}} x'_{1,i-1} + \beta_{\text{length}} x'_{2,i} + \beta_{\text{mortality}} x'_{3,i-1} + \beta_{\text{capelin}} x'_{4,i-1} + \varepsilon_i$$

where the  $\varepsilon_i$  form a stationary Gaussian autoregressive time series of order two and where we use  $x'_{1,i-1}$  etc. to denote centred variables (so  $x'_{2,i}$  denotes  $x_{2,i}$  minus the mean of these values, and so forth). This gives  $\beta_0$  interpretation as overall mean of the HSI. The above amounts to a rather short series with only 32 complete observations in that the capelin series only ranges from 1981 to 2012. We therefore supplement model 8 with a second model, where we bypass the capelin index and can use 81 complete observations with  $(x_1, x_2, x_3)$  in place (note that the mean length is imputed from 1973 to 1979).

As a first analysis, we fit the data to the regression framework with autoregressive errors of order two to investigate the implied predictive quality (see Table 3). A common measure of how well data fit a model is the coefficient of determination, typically denoted  $R^2$ , with a fine-tuned version  $R^2_{\text{adj}}$ . For the full model above, we obtain  $R^2 = 0.59$  ( $R^2_{\text{adj}} = 0.48$ ). This is in a sense not that representative for the entire series, since the capelin series only covers the years 1981–2012. For the second analysis, which bypasses the capelin index, we obtain the more promising values of  $R^2 = 0.74$  ( $R^2_{\text{adj}} = 0.72$ ). The baseline autoregressive model (with intercept and linear trend) achieves  $R^2 = 0.59$  and ( $R^2_{\text{adj}} = 0.58$ ). This shows that including covariates is a genuine improvement in model quality.

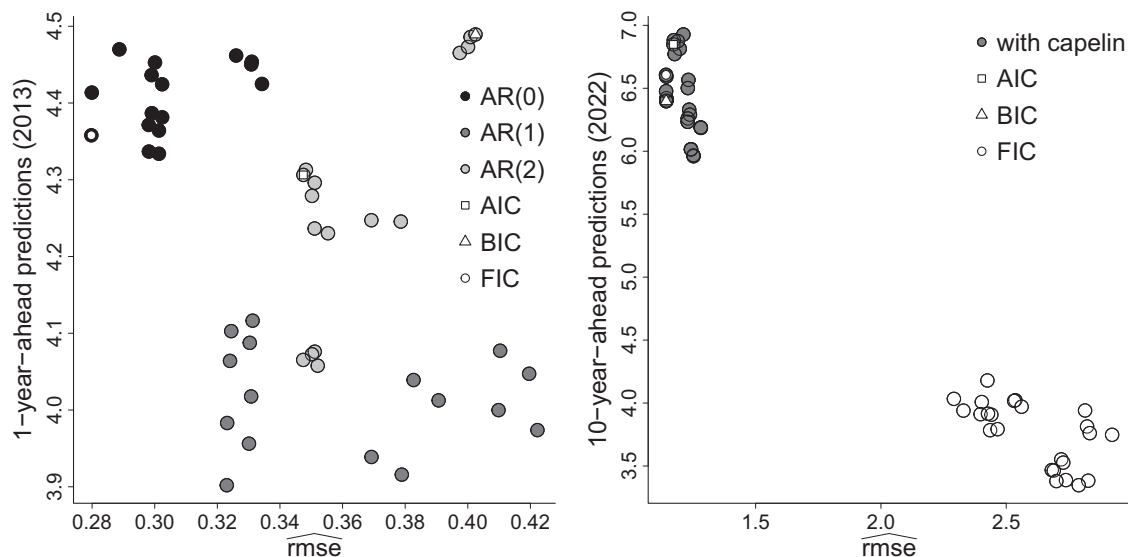
From Table 3 we see that in our joint model all covariates, except the length, are important for the HSI. The fish mortality rate  $F$  and the capelin, our proxy for the food availability, are the most important covariates. This fits quite well with our univariate analysis above. It is perhaps surprising that the total length

**Table 3.** Estimates, standard errors, Wald ratios of estimates by standard errors, and  $p$  values for the parameters in the model given by eq. 8, fitted using maximum likelihood estimation, with data from 1981 to 2012 (left part) and also from 1932 to 2012 (right part) for the dataset that does not include the capelin series.

|                            | 1981–2012 |       |        |           | 1932–2012 |       |        |           |
|----------------------------|-----------|-------|--------|-----------|-----------|-------|--------|-----------|
|                            | Estimate  | SD    | Ratio  | $p$ value | Estimate  | SD    | Ratio  | $p$ value |
| $\beta_0$                  | 5.395     | 0.304 | —      | —         | 7.724     | 0.292 | —      | —         |
| $\beta_{\text{year}}$      | −0.029    | 0.016 | −1.830 | 0.040*    | −0.041    | 0.007 | −6.115 | 0.000**   |
| $\beta_{\text{Kola}}$      | 0.317     | 0.291 | 1.090  | 0.144     | 0.256     | 0.152 | 1.683  | 0.048*    |
| $\beta_{\text{length}}$    | 0.021     | 0.027 | 0.751  | 0.230     | 0.021     | 0.020 | 1.070  | 0.144     |
| $\beta_{\text{mortality}}$ | 1.399     | 0.600 | 2.333  | 0.014*    | 0.974     | 0.541 | 1.800  | 0.038*    |
| $\beta_{\text{capelin}}$   | 0.239     | 0.092 | 2.581  | 0.008**   | —         | —     | —      | —         |
| $\rho_1$                   | 0.375     | 0.168 | 2.224  | 0.018*    | 0.798     | 0.106 | 7.519  | 0.000**   |
| $\rho_2$                   | −0.512    | 0.154 | −3.332 | 0.001**   | −0.293    | 0.115 | −2.548 | 0.006**   |
| $\sigma$                   | 0.542     | 0.068 | —      | —         | 0.581     | 0.046 | —      | —         |

Note: \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ .

**Fig. 13.** For predicting the HSI for next year (2013, left panel), the FIC prefers the smallest model with no predictors and independence. This seems reasonable, since among the “best” models this is clearly the simplest option. It indicates that the FIC, in case of a reasonable tie among the candidate models, often selects the model with the lowest complexity, which leads to less variation in the estimates and also to potentially more precise inference. For prediction 10 years ahead (2022, right panel), the FIC selects the independent model using only the capelin series, suggesting that the food availability is one of the main signals in the long-term effect of the HSI.



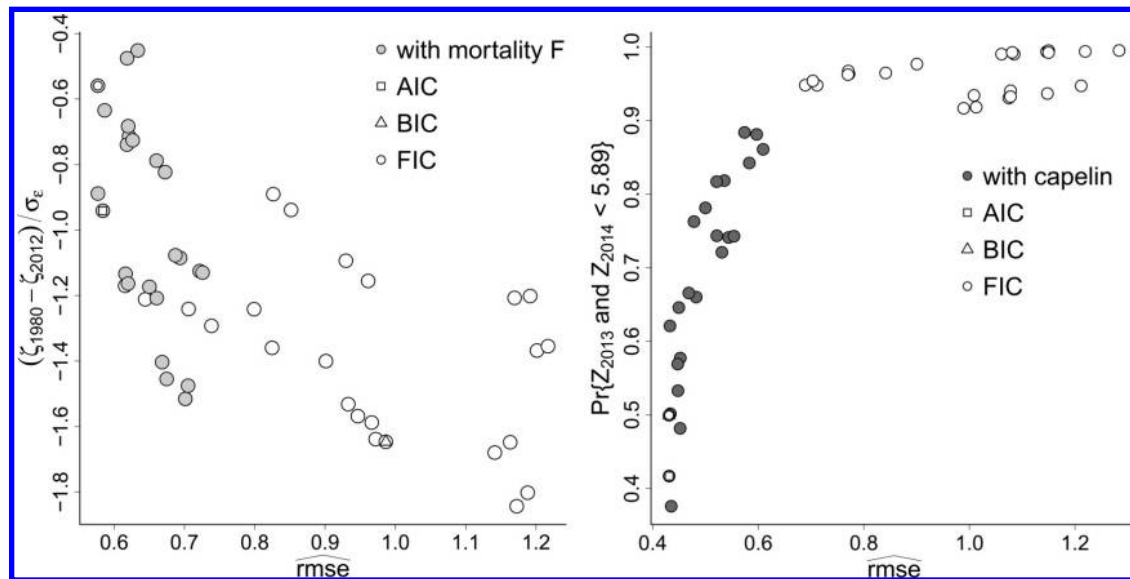
variable is not more vital here (separately, it has a strong correlation with the HSI); this happens partly because it is well explained by the other predictors. The effect does not seem to be a product of the short series (with 31 complete samples) used for the analysis in Table 3. More important, however, is the fact that  $\beta_{\text{year}}$  is estimated to have a negative slope and small  $p$  values (for both year groupings of Table 3). With a somewhat strict interpretation, this means that the liver quality index is currently decreasing with time. This is not an artefact of the three successive low HSI<sub>bulk</sub> values at the end (2010–2012) of the series (values for which we have already expressed mild scepticism; see the discussion related to Fig. 4 above). Indeed, we obtain very similar results when carrying out certain robustified analyses (not reported on in detail here).

Based on the univariate and joint analyses presented in Table 3, we do not believe that all covariates (i.e., year, Kola, length, mortality, and capelin) are of the same importance. Moreover, the full model, comprising all five predictors and four additional modelling parameters ( $\beta_0$  and three for the autoregressive part), is also a bit too rich for reliable estimation, with only 32 samples. Traditionally, one would have to do some preselection, or model selection, to obtain a subset of predictors to use in the final analysis.

There are several possibilities, with classical approaches including AIC and BIC and the so-called forward, backward, or all subsets options. For those possibilities, only significant covariates reach the final model, typically evaluated via  $p$  values needing to be less than a threshold. This is not always the best approach, however, since a single model is typically not best for all purposes, and the relative importance of the different predictors or the model complexity may depend on what we are trying to answer (predictions, threshold probabilities, or estimation of underlying structural changes). For these reasons, the focused model selection strategy (FIC) presented above may easily lead to the best solution.

We have already discussed how the FIC can be used to find the model best suited for making predictions, and above we showed how models of different complexity were preferred for different numbers of time steps (years) into the future (see Fig. 9). There is a variety of further focus functions that can help the researcher to select the model best suited for the particular problem at hand. Using the short series discussed in Table 3, we are again attempting to predict the HSI for the future in Fig. 13, and in Fig. 14 we consider two additional foci:

**Fig. 14.** For  $\mu_{\text{slope}}$  (left panel) from eq. 9, the FIC agrees with the AIC and prefers the autoregressive model of order two with the mortality rate  $F$  present in most of the best models. For  $\mu_{\text{threshold}}$  (right panel), the best model according to the FIC is the independent model that includes all covariates except the Kola temperature; note that the capelin index is included in most of the top models and that BIC agrees with FIC on which model is best.



$$(9) \quad \begin{aligned} \mu_{\text{slope}} &= (\xi_{1980} - \xi_{2000}) / \sigma_{\epsilon} \\ \mu_{\text{threshold}} &= \Pr\{Z_{2013} \text{ and } Z_{2014} \text{ smaller than } 5.89\} \end{aligned}$$

Here  $\xi_{1980}$  and  $\xi_{2000}$  are the expected levels of the HSI at years 1980 and 2000, respectively, with different formulae applying for these in different models, and 5.89 is the overall mean of the HSI<sub>bulk</sub> series.

The AIC prefers the autoregressive model of order two with both mortality rate and the food availability proxy capelin. The BIC selects the autoregressive model of order two with only the capelin index. Again, the AIC and BIC suggest their models regardless of any focus parameters, whereas the FIC takes these into account and suggests different top models for different purposes.

Note that by using the robustified HSI series from Fig. 4 (also including the capelin), the overall picture is somewhat changed. The prediction for 2013 increases to 5.1 (based on FIC, whereas the AIC and BIC suggest that it will be as high as 5.3). Also, the threshold probability is reduced to 0.22 where the model selected by AIC gives an estimate of 0.25, and BIC suggests a model leading to an estimated probability as low as 0.12. In general, using the FIC we are able to reach reliable conclusions via the most relevant model selection process. This is also borne out of analyses summarized in Table 4. These pertain to estimated sequential out-of-sample 1-, 3-, and 10-year-ahead HSI prediction algorithms and their associated RMSE (root-mean-squared error). These algorithms for predicting the HSI stem from using models selected by the three criteria AIC, BIC, and FIC, asked to find the best candidate among autoregressive models of orders up to order three. On average, the FIC selects models better for the prediction task. The differences for this particular illustration are small, however. We have carried out similar experiments for simulated series and have seen the same pattern emerge as here.

### Combining information across data sources

For various substantive sciences, it is important to be able to pool information sources together, leading to proper combination or meta-analyses for the more crucial research questions. Typical scenarios met (e.g., in the biomedical sphere) involve similarly structured experiments, carried out by different research groups, after which a meta-analysis properly combines the sum-

**Table 4.** Estimated sequential out-of-sample 1-, 3-, and 10-year-ahead prediction RMSE (root-mean-squared error) when predicting the HSI for the models selected by the three criteria FIC, AIC, and BIC.

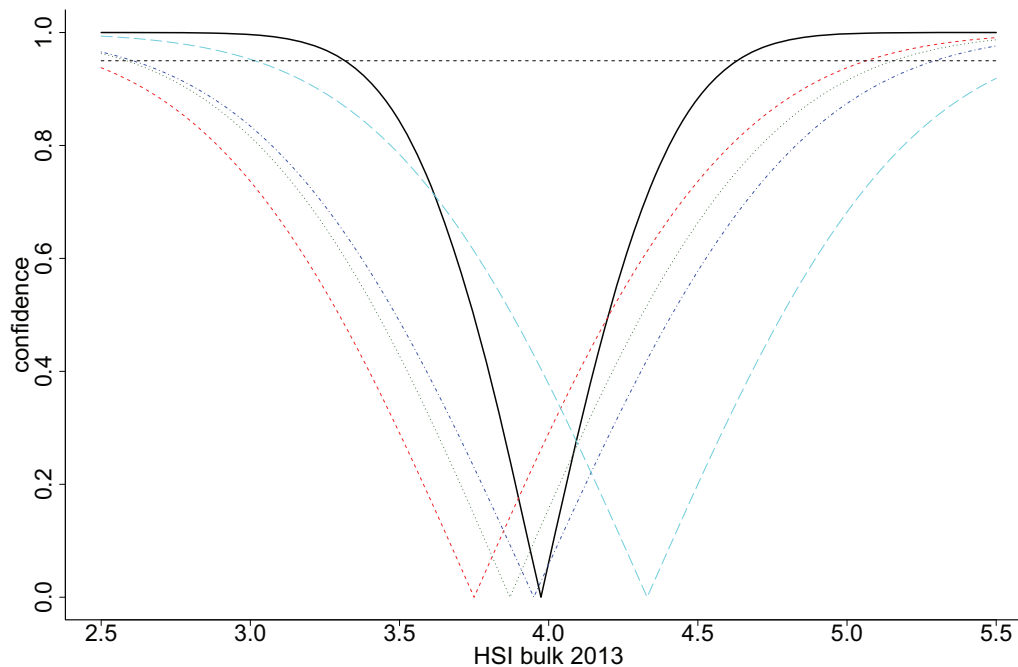
| Criterion | 2000–2012    |      |      | 1960–2012    |      |      |
|-----------|--------------|------|------|--------------|------|------|
|           | h-year-ahead |      |      | h-year-ahead |      |      |
| FIC       | 1            | 3    | 10   | 1            | 3    | 10   |
| FIC       | 0.87         | 1.01 | 1.17 | 1.23         | 1.36 | 1.05 |
| AIC       | 1.01         | 1.10 | 1.60 | 1.23         | 1.36 | 1.00 |
| BIC       | 1.01         | 1.08 | 1.49 | 1.29         | 1.37 | 1.10 |

**Note:** Data for 2002–2012 uses all covariates, as in the model of eq. 8 for the years 1981–2012, and data for 1960–2012 use the longer series without capelin for the years 1932–2012; see also Table 3. In the short series, the FIC selects models that are on average better, in terms of RMSE. The differences are small for this illustration, however; similar patterns emerge for higher- and lower-order autoregressive models.

mary statistics across these individual studies. Increasingly, such meta-questions arise also in other fields and in situations where the information sources are rather more diverse. A case in point is Myers (2001), who analysed a compilation of over 700 populations of fish, in a framework of several multivariate time series. Other and yet more complex meta-studies could involve field studies, measurements from satellites, mark and recapture data, biological evolutionary theories, etc. Traditional meta-analysis methods would often not cope with the associated data summary problems, also since the individual data summaries themselves would differ in format and level of precision.

A powerful general framework for such modern meta-analyses involves that of “confidence distributions” (Xie and Singh 2013; Schweder and Hjort 2002, 2016). A confidence distribution for a parameter, say  $\theta$ , based on data, say  $y$ , is a cumulative distribution function  $C(\theta, y)$ , with the property that it spans all confidence intervals; thus  $[C^{-1}(0.05, y), C^{-1}(0.95, y)]$  is a 90% confidence interval, etc. The confidence distributions are prior-free and thus principally different from Bayesian calculations, but they have the same conceptual and operational aim, that of providing probability distributions for focus parameters given the data. See the above references for broad discussions and illustrations. It is also useful to work with the associated “confidence curves”:  $cc(\theta, y) = [1 - 2C(\theta, y)]$ . These have the property that setting  $cc(\theta, y)$  equal to a

**Fig. 15.** Confidence curves for predicting the value of  $HSI_{bulk}$  for the year 2013, based on separate analyses of HSI against Kola winter temperature (short dashed line), length data (dotted line), mortality figures (dot-dashed line), and capelin score (long dashed line), along with the optimal confidence curve combination (solid line) of these sources. The y axis indicates degree of confidence, with the 0.95 confidence line plotted (horizontal dotted line).



given confidence level, say 0.95, leads to a lower and an upper solution point, spanning the confidence interval in question. Also, the confidence curve “points to” the median confidence estimate,  $\hat{\theta} = C^{-1}(0.5, y)$ , which is sometimes but not always identical to the associated maximum likelihood estimate.

Figure 15 displays such confidence curves for predicting  $HSI_{bulk}$  for the year 2013 (the year after the current last year of the HSI series) for each of the separate studies related to Kola winter temperature, mortality rate, capelin index, and length data. A fifth and tighter confidence curve is the one representing the combined information, as explained now.

The general meta-analysis idea associated with confidence distributions is as follows. For a parameter of primary interest, like the trend parameter  $\beta_1 = \beta_{year}$  of eq. 5 reflecting a potential change of the HSI over time, each separate source of relevant information leads to a confidence distribution for the parameter, say  $C_1(\beta_1), \dots, C_k(\beta_1)$ . Each of these studies might be a complex statistical affair, regarding model building, interpretation, data gathering, and operational practicalities, and might also involve various other statistical parameters along the way. The confidence distributions may be converted to profiled log-likelihood functions, say  $\ell_1(\beta_1), \dots, \ell_k(\beta_1)$ , in ways developed and illustrated in Schweder and Hjort (2016). This leads to the intended meta-analysis by adding up the log-likelihood pieces and reverting the result to a confidence distribution again.

We give two illustrations here. The first has actually already been pointed to, when Fig. 4 illustrated improved estimates of the HSI bulk for 1997 to 2012, combining the observed  $HSI_{bulk}$  numbers with appropriately transformed versions of the  $HSI_{ind}$  numbers from the Lofoten survey. The resulting “combined information curve” (the black line in the figure mentioned) has emerged via the above recipe, involving assessment of each contributing curve’s precision level. Our second illustration is Fig. 15, for estimating HSI for 2013 based on other data sources. Here the solid black curve gives the proper optimal combination of the four other confidence curves, those based on separate analyses of Kola winter temperature, length data, mortality, and capelin index. 95% confidence intervals may be read off for the

separate analyses as well as for the combined one. These precision intervals are [2.44, 5.06] for the Kola winter temperatures, [2.58, 5.16] for the length data, [2.62, 5.29] for the mortality, [3.02, 5.64] for the capelin index, and finally the much shorter interval [3.32, 4.63] for the method that optimally combines these pieces of information. The point estimate for HSI for 2013, using the four covariates, is 3.97.

### Supplementary remarks

We end our paper with a list of concluding comments, some pointing to the fruitfulness of further investigations, regarding both methodology and the examination of the Hjort index and its related factors.

#### A. To explain or to predict?

Breiman (2001) makes the case that there is a Snow-like “two cultures” aspect of modern statistical science, with one camp eager to fit data to make good classifications and predictions, while the other focuses more on the finer details of models to calibrate understanding of say underlying biological mechanisms. The “to explain or to predict” question (cf. Shmueli 2010) is also pertinent when it comes to the many aspects of understanding the Hjort liver quality index. It is certainly of intrinsic value to understand a determining aspect of how the Kola winter temperature influences the liver quality, even though this might not shed light on its own on how the size of the population develops over the next 20 years. As argued in Claeskens and Hjort (2003, 2008), it is not a paradox that one model does a good job with “explanation” with another one is better at “looking ahead”. The FIC methods for model selection aid both aims.

#### B. Influential factors, changes, fluctuations: deeper issues

Regarding the “to explain or to predict” dimensions pointed to in the previous remark, we suggest that Johan Hjort expressed deep interest in both. His work reflects energetic fascination with both the finer aspects and details of how nature works and how insights lead to assessment of the future, from life in the ocean to human endeavours and industries associated with it. He was also

profoundly interested in the nature and business of variation and fluctuations, in and by themselves, also in his more philosophical writings (e.g., Hjort 1920, chapter III), seeing also tentative consequences for the sociology of human beings.

He was also clearly interested in fluctuations and in their interplay and correlations with other phenomena, long before his 1914 book. In the other classic *The Depths of the Ocean* with Sir John Murray (Hjort and Murray 1912), he studied percentage of fat in sprats caught off the Norwegian west coast in different months, comparing these with mean temperature of the surface of the sea off Bergen in each month of the year and comments: "The fat-contents of the sprat increase during summer, when there is a rise in temperature, while both decrease towards the end of the year; it follows from this that the growth of the fish must be influenced by the prevailing temperatures in different waters." The remark captures the spirit of constructive curiosity that we still need to learn more.

The following quotation is also apt for our limited efforts and successes in understanding how other factors influence the lives, quantity, and quality of the skrei (from Hjort 1938, chapter VI, the chapter on the methods of correlation and experiment): "The cause of our anxiety is the critical fear of an overestimation of the method, or rather the conclusions drawn from its application. The correlations which we call adaptation are certainly of far too great a complexity for any postulate of causal connections, for the organism and environment are two enormously complicated conceptions. The understanding of the causal connections between them demands first of all a prolonged critical analysis like that which Gregor Mendel introduced into the biological study of the external characters of the plant, but an analysis of this kind is possible only on the basis of previous experience created by studies of correlations."

### C. Model averaging

Sometimes it is fruitful to not merely point to one winning statistical model (and to discard all its competitors), but to keep several good models on board. Model averaging is the term for keeping several model-based estimates, with the final analysis being a mixture across these. The weights in question, given to the different models, might be based on FIC scores, so that the best-looking models have higher weights than those not scoring well for the purpose at hand. This is sometimes particularly fruitful in prediction settings (Hjort and Claeskens 2003; Claeskens and Hjort 2008; Hansen 2008; Cheng and Hansen 2015).

### D. Modelling HSI as a mixture

Naturally measuring the per-fish liver index for a number of specimens takes certain laborious and precision demanding efforts. We learn from the detailed analysis reported on above, involving the bivariate gamma model (eq. 3), that the per-fish and bulk liver indexes are strongly related and that inference about the former may be reached based on the latter, these involving easier measurements. We point out that for other populations, and specifically for smaller fish, the correlation might be smaller and the difference between the two indices larger. This is reflected in the fact that  $E(X/EY)$  is not a good approximation to  $E(X/Y)$  when the variance of  $Y$  is moderate or big and in particular when  $Y$  may be small with reasonable probability.

The HSI is an overall number for the liver quality of a stock of fish (in a given year), averaged over subclasses of age, sex, and perhaps other identifiable categories, like year-class. For larger specimens, the female cod typically have a higher HSI than the males (e.g., Krivobok and Tokareva 1973; Karlsen et al. 2005). We may represent this as  $HSI = \sum_{strata} w(u) HSI(u)$ , with  $HSI(u)$  denoting the index for stratum  $u$  and  $w(u)$  the relative frequency of this stratum among all strata. Thus, a different way of measuring or modelling the overall HSI is via separate modelling of the frequencies, say across age and sex, along with separate modelling or

sampling for these groups. These considerations also tell us that the HSI number is a complex quantity that may vary from one year to the next via a number of reasons. In particular, the HSI may change over time not because the individual specimens change their liver sizes but because the demographical characteristics change (e.g., having a higher proportion of older fish).

### E. Time-varying autoregressive modelling via locally stationary processes

The time series models we have used in this paper are of the type "trend function linear in covariates plus low-order autocorrelated noise". These are effective, not difficult to work with, and reasonably robust, in that moderate deviations from the model used will not severely disturb inferences. It is nevertheless fruitful to attempt other variations and perhaps more sophisticated time series models (e.g., involving a standard deviation function  $\sigma_t$  varying with year  $t$ ). Similarly, some of the bridge monitoring model checks we have carried out for the HSI and related data series indicate that model parameters may change over time. A pertinent class of models able to cope with such aspects is that worked with in Dahlhaus (1997), the time-varying autoregressive model (tvAR). These are defined by  $Y_t + \sum_{j=1}^p \alpha_j(t/n)Y_{t-j} = \sigma(t/n)\varepsilon_t$ , for  $t = 1, \dots, n$ , where the  $\varepsilon_t$  are independent and standard normal. When the  $\sigma(\cdot)$  and  $\alpha_j(\cdot)$  functions are constant, we are back to the familiar ground of autoregressive time series. We have tested these models for the HSI bulk series, using order  $p = 2$  above, which suggest that the  $\sigma(\cdot)$  function is not constant over time (as indicated in Fig. 7), whereas the  $\alpha_1(\cdot)$  and  $\alpha_2(\cdot)$  coefficient functions are approximately constant. This provides evidence that the dependency structure does not change significantly over time.

### Acknowledgements

We are grateful to Jennifer Devine and Jon Egil Skjæraasen at the Hjort Centre for Marine Ecosystems Dynamics and the Institute of Marine Research (IMR, Bergen), and also to Bjarte Bogstad at the IMR, for making relevant datasets available to us as well as for fruitful discussions. The Kola temperature time series stems from systematic work at the Polar Research Institute of Marine Fisheries and Oceanography (PINRO, Murmansk) reported in Boitsov et al. (2012), with the data kindly provided by Randi Ingvaldsen at the IMR, who communicates with PINRO in these regards. The authors are also grateful for detailed and constructive comments on an earlier version from an associate editor and three reviewers. G.H.H. and N.L.H. are also indebted to the Research Council of Norway for partial funding of the 5-year project FocuStat (Focus Driven Statistical Inference With Complex Data).

### References

- Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike*. Springer. pp. 199–213.
- Boitsov, V.D., Karsakov, A.L., and Trofimov, A.G. 2012. Atlantic water temperature and climate in the Barents Sea, 2000–2009. *ICES J. Mar. Sci.* **69**: 833–840. doi:10.1093/icesjms/fss075.
- Brander, K.M. 2010. Impacts of climate change on fisheries. *J. Mar. Syst.* **79**: 389–402. doi:10.1016/j.jmarsys.2008.12.015.
- Breiman, L. 2001. Statistical modeling: the two cultures. [With discussion and rejoinder.] *Stat. Sci.* **16**: 199–215. doi:10.1214/ss/1009213726.
- Brillinger, D.R. 1975. *Time series: data analysis and theory*. Holt, Rinehart and Winston.
- Brockwell, P.J., and Davis, R. 1991. *Time series: theory and methods*. Springer.
- Cheng, X., and Hansen, B.E. 2015. Forecasting with factor-augmented regression: a frequentist model averaging approach. *J. Econ.* **186**: 280–293. doi:10.1016/j.jeconom.2015.02.010.
- Claeskens, G., and Hjort, N.L. 2003. The focused information criterion. [With discussion and a rejoinder.] *J. Am. Stat. Assoc.* **98**: 900–916. doi:10.1198/016214503000000819.
- Claeskens, G., and Hjort, N.L. 2008. *Model selection and model averaging*. Cambridge University Press, Cambridge.
- Dahlhaus, R. 1997. Fitting time series models to nonstationary processes. *Ann. Stat.* **15**: 1–37. doi:10.1214/aos/1034276620.
- Dickey, D.A., and Fuller, W.A. 1979. Distribution of the estimators for autore-

- gressive time series with a unit root. *J. Am. Stat. Assoc.* **74**: 427–431. doi:[10.2307/2286348](https://doi.org/10.2307/2286348).
- Einstein, A. 1934. On the method of theoretical physics. *Philos. Sci.* **1**: 163–169. doi:[10.1086/286316](https://doi.org/10.1086/286316).
- Frigessi, A., and Hjort, N.L. 2002. Statistical models and methods for discontinuous phenomena. [Introduction to the special issue on this topic.] *J. Nonparam. Stat.* **14**: 1–6. doi:[10.1080/10485250211393](https://doi.org/10.1080/10485250211393).
- Grønneberg, S., Hermansen, G.H., and Hjort, N.L. 2015. The jump information criterion. Technical report, Department of Mathematics, University of Oslo.
- Hansen, B.E. 2008. Least-squares forecast averaging. *J. Econ.* **146**: 342–350. doi:[10.1016/j.jeconom.2008.08.022](https://doi.org/10.1016/j.jeconom.2008.08.022).
- Helland-Hansen, B., and Nansen, F. 1909. The Norwegian Sea: its physiological oceanography, based upon the Norwegian Researches 1900–1904. Report on Norwegian Fishery and Marine-Investigations, Kristiania.
- Hermansen, G.H., and Hjort, N.L. 2014. A new approach to Akaike's information criterion and model selection issues in stationary Gaussian time series. Technical report, University of Oslo and Norwegian Computing Centre.
- Hermansen, G.H., and Hjort, N.L. 2015a. Bernshtein-von Mises theorems for nonparametric function analysis via locally constant modelling: a unified approach. *J. Stat. Plann. Infer.* **166**: 138–157. doi:[10.1016/j.jspi.2015.02.007](https://doi.org/10.1016/j.jspi.2015.02.007).
- Hermansen, G.H., and Hjort, N.L. 2015b. Focused information criteria for time series. [Submitted for publication.]
- Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Kluwer Academic Publishers, Dordrecht.
- Hjort, J. 1914. Fluctuations in the Great Fisheries of Northern Europe, viewed in the light of biological research. Conseil Permanent International Pour l'Exploration de la Mer, Copenhagen.
- Hjort, J. 1920. The unity of science. Gyldendal, London.
- Hjort, J. 1938. The human value of biology. Harvard University Press, Cambridge, Mass.
- Hjort, J. and Murray, J. 1912. The depths of the ocean. Macmillan, London.
- Hjort, N.L., and Claeskens, G. 2003. Frequentist model average estimators. [With discussion and a rejoinder.] *J. Am. Stat. Assoc.* **98**: 879–899. doi:[10.1198/016214503000000828](https://doi.org/10.1198/016214503000000828).
- Hjort, N.L., and Koning, A.J. 2002. Tests for constancy of model parameters over time. *J. Nonparam. Stat.* **14**: 113–132. doi:[10.1080/10485250211394](https://doi.org/10.1080/10485250211394).
- Hylen, A., Nakken, O., and Nedreaas, K. 2008. Northeast Arctic cod: fisheries, life history, stock fluctuations and management. In *Norwegian spring-spawning herring & Northeast Arctic cod*. Tapir Academic Press, Trondheim, Norway. pp. 83–118.
- ICES. 2014. Report of the Arctic Fisheries Working Group (AFWG), 23–29 April 2014, Lisbon, Portugal. Technical report, ICES, Copenhagen.
- Jølle, H.D. 2011. Nansen: Oppdageren. [The Explorer.] Gyldendal, Oslo.
- Jørgensen, T. 1990. Long-term changes in age at sexual maturity of Northeast Arctic cod (*Gadus morhua* L.). *J. Cons. Int. Explor. Mer.* **46**: 235–248. doi:[10.1093/icesjms/46.3.235](https://doi.org/10.1093/icesjms/46.3.235).
- Karlsen, Ø., Holm, J.C., and Kjesbu, O.S. 2005. Effects of periodic starvation on reproductive investment in first-time spawning Atlantic cod (*Gadus morhua* L.). *Aquaculture*, **133**: 159–170. doi:[10.1016/0044-8486\(94\)00399-9](https://doi.org/10.1016/0044-8486(94)00399-9).
- King, J.R. 2005. Report of the study group on fisheries and ecosystem responses to recent regime shifts. Technical Report 28, PISCES.
- Kjesbu, O.S., Witthames, P.R., Solemdal, P., and Greer, Walker, M. 1998. Temporal variations in the fecundity of Arcto-Norwegian cod (*Gadus morhua*) in response to natural changes in food and temperature. *J. Sea Res.* **40**: 303–321. doi:[10.1016/S1385-1101\(98\)00029-X](https://doi.org/10.1016/S1385-1101(98)00029-X).
- Kjesbu, O.S., Righton, D., Krüger-Johnsen, M., Thorsen, A., Michalsen, K., Fonn, M., and Witthames, P.R. 2010. Thermal dynamics of ovarian maturation in Atlantic cod (*Gadus morhua*). *Can. J. Fish. Aquat. Sci.* **67**(4): 605–625. doi:[10.1139/F10-011](https://doi.org/10.1139/F10-011).
- Kjesbu, O.S., Opdal, A.F., Korsbrekke, K., Devine, J.A., and Skjæraasen, J.E. 2014a. Making use of Johan Hjort's 'unknown' legacy: reconstruction of a 150-year coastal time-series on northeast Arctic cod (*Gadus morhua*) liver data reveals long-term trends in energy allocation patterns. *ICES J. Mar. Sci.* **71**: 2053–2063. doi:[10.1093/icesjms/fsu030](https://doi.org/10.1093/icesjms/fsu030).
- Kjesbu, O.S., Bogstad, B., Devine, J.A., Gjøsaeter, H., Howell, D., Ingvaldsen, R.B., Nash, R.D.M.N., and Skjæraasen, J.E. 2014b. Synergies between climate and management for Atlantic cod fisheries at high latitudes. *Proc. Natl. Acad. Sci. U.S.A.* **111**: 3478–3483. doi:[10.1073/pnas.1316342111](https://doi.org/10.1073/pnas.1316342111). PMID:24550465.
- Krivobok, M.N., and Tokareva, G.I. 1973. Dynamics of weight variations of the body and individual organs of Baltic cod during the maturation of gonads. Fisheries Research Board of Canada [Translation Series No. 2722].
- Lindquist, A. 2002. Herring periods of Bohusian: a cross-sectorial approach. *ICES Mar. Sci. Symp.* **215**: 343–351.
- Ljungman, A.V. 1879. Bidrag til lösningen av frågan om de stora sillfiskenas sekuläre periodicitet. *Tidsk. Fisk.* **5**: 257–268.
- Marshall, C.T., Yaragina, N.A., Lambert, Y., and Kjesbu, O.S. 1999. Total lipid energy as a proxy for total egg production by fish stocks. *Nature*, **402**: 288–290. doi:[10.1038/46272](https://doi.org/10.1038/46272).
- Mjanger, H., Hestenes, K., Olsen, E., Svendsen, B.V., and de lange Wenneck, T. 2006. Manual for sampling of fish and crustaceans. Version 1.0. Technical report, Institute of Marine Research, Bergen.
- Myers, R.A. 2001. Stock and recruitment: generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. *ICES J. Mar. Sci.* **58**: 937–951. doi:[10.1006/jmsc.2001.1109](https://doi.org/10.1006/jmsc.2001.1109).
- Sandeman, L.R., Yaragina, N.A., and Marshall, C.T. 2008. Factors contributing to inter- and intra-annual variation in condition of cod (*Gadus morhua*) in the Barents Sea. *J. Anim. Ecol.* **77**: 725–734. doi:[10.1111/j.1365-2656.2008.01388.x](https://doi.org/10.1111/j.1365-2656.2008.01388.x). PMID:18384351.
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* **6**: 461–464. doi:[10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136).
- Schweder, T., and Hjort, N.L. 2002. Confidence and likelihood. *Scand. J. Stat.* **29**: 309–332. doi:[10.1111/1467-9469.00285](https://doi.org/10.1111/1467-9469.00285).
- Schweder, T., and Hjort, N.L. 2016. Confidence, likelihood, probability. Cambridge University Press, Cambridge.
- Shmueli, G. 2010. To explain or to predict? *Stat. Sci.* **25**: 289–310. doi:[10.1214/10-STS330](https://doi.org/10.1214/10-STS330).
- Skjæraasen, J.E., Nash, R.D.M., Korsbrekke, K., Fonn, M., Nilsen, T., Kennedy, J., Nedreaas, K.H., Thorsen, A., Witthames, P.R., Geffen, A.J., Høie, H., and Kjesbu, O.S. 2012. Frequent skipped spawning in the world's largest cod population. *Proc. Natl. Acad. Sci. U.S.A.* **109**: 8995–8999. doi:[10.1073/pnas.1200223109](https://doi.org/10.1073/pnas.1200223109). PMID:22615381.
- Smith, T.D. 1994. Scaling fisheries: the science of measuring the effects of fishing 1855–1955. Cambridge University Press, Cambridge.
- Stearns, S.C. 1992. The evolution of life histories. Oxford University Press, Oxford.
- Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., and Working Group I Technical Support Unit. (Editors). 2013. The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, IPCC.
- Tyler, C.R., and Sumpter, J.P. 1996. Oocyte growth and development in teleosts. *Rev. Fish Biol. Fish.* **6**: 287–318. doi:[10.1007/BF00122584](https://doi.org/10.1007/BF00122584).
- Vasilakopoulos, P., and Marshall, C.T. 2015. Resilience and tipping points of an exploited fish population over six decades. *Glob. Change Biol.* **21**: 1834–1847. doi:[10.1111/gcb.12845](https://doi.org/10.1111/gcb.12845).
- Xie, M., and Singh, K. 2013. Confidence distribution, the frequentist distribution estimator of a parameter: a review. [With discussion and a rejoinder.] *Int. Stat. Rev.* **81**: 3–39. doi:[10.1111/insr.12000](https://doi.org/10.1111/insr.12000).
- Yndestad, H. 2009. The influence of long tides on ecosystem dynamics in the Barents Sea. *Deep-Sea Res. II Top. Stud. Oceanogr.* **56**: 2108–2116. doi:[10.1016/j.dsr2.2008.11.022](https://doi.org/10.1016/j.dsr2.2008.11.022).