

## Estimating a Smooth Common Transfer Function with a Panel of Time Series - Inflow of Larvae Cod as an Example

Elizabeth Hansen  
Department of Mathematics  
476 Morgan Hall, 1 University Circle  
Macomb, IL 61455, USA  
EA-Hansen@wiu.edu

Kung-Sik Chan  
Department of Statistics and Actuarial Science  
Schaeffer Hall 263, University of Iowa  
IA 52242, USA  
kung-sik-chan@uiowa.edu

Nils Chr. Stenseth  
Centre for Ecological and Evolutionary Synthesis (CEES)  
University of Oslo, P.O. Box 1066, Blindern  
N-0316 Oslo, Norway  
n.c.stenseth@bio.uio.no

### Abstract

The annual response variable in an ecological monitoring study often relates linearly to the weighted cumulative effect of some daily covariate, after adjusting for other annual covariates. Here we consider the problem of non-parametrically estimating the weights involved in computing the aforementioned cumulative effect, with a panel of short and contemporaneously correlated time series whose responses share the common cumulative effect of a daily covariate. The sequence of (unknown) daily weights constitutes the so-called transfer function. Specifically, we consider the problem of estimating a smooth common transfer function shared by a panel of short time series that are contemporaneously correlated. We propose an estimation scheme using a likelihood approach that penalizes the roughness of the common transfer function. We illustrate the proposed method with a simulation study and a biological example of indirectly estimating the spawning date distribution of North Sea cod.

**Keywords:** Cod-spawning-date distribution, (generalized) cross-validation, Seemingly unrelated regression, Multimodality.

### 1. Introduction

Recent genetic analysis by Knutsen, Andre., Jorde, Skogen, Thuroczy, and Stenseth (2004), suggested that the young (half-year old) cod sampled in some fjords in the Skagerrak, Norway, resembled adult cod in the North Sea in year 2001 but less so in year 2000. It was, furthermore, found that in 2001 when the sampled young cod of Skagerrak were genetically similar to the adult cod of North Sea, there was higher than average inflow of sea current from the North Sea to the Skagerrak, but not so in 2000 when the resemblance

switched to the local adult cod. Thus, Knutsen et al. (2004) suggested the hypothesis that the North Sea cod stock might have contributed to the local cod population in the Skagerrak via transportation of cod eggs by sea current from North Sea into the Skagerrak. Stenseth, Jorde, Chan, Hansen, Knutsen, Andre, Skogen, and Lekve (2006) tested this hypothesis using a long-term monitoring data on the (adjusted) annual counts of young cod, the (annual) spawning biomass of North Sea cod and daily inflow of sea current from North Sea to Skagerrak. It is believed that the cod spawn, or breed, in the months of March and April, but it is not known specifically when the majority of the spawning took place. Stenseth et al. (2006) computed the average daily inflow (from North Sea to the Skagerrak) over several windows of 2-week period between March and April, and tested the transportation hypothesis using a regression model with a covariate that is the product of average sea influx times log spawning biomass (of the North Sea cod), a proxy for the transportable amount of cod eggs, the coefficient of which is non-zero under the transportation hypothesis and zero otherwise. Stenseth et al. (2006) found that the transportation hypothesis is consistent with the data, with stronger, significant result when the mean inflow is computed over the second half of March. Clearly, which two-week period over which the mean inflow is computed is critical as the test can be made more powerful by aligning the period with the main period when the cod spawned.

Here, we propose to study the problem from a different perspective. Instead of searching for an optimal window for averaging the daily inflow, we consider the distribution of the cod spawning date. Let  $S$  be the day counted starting from the beginning of March of each year, when a randomly selected adult cod spawns. Let  $\psi_j$  be the probability that  $S = j$ . The daily contribution of North Sea cod to the Skagerrak is postulated to additively contribute, on the logarithmic scale, to the young cod counts by an amount proportional to  $\psi_j c_{t,j} b_t$  where  $b_t$  is the log spawning biomass in the  $t$ th year and  $c_{t,j}$  is the mean inflow on the  $j$ th day (counted starting March 1st) of the  $t$ th year; for simplicity of notation, the proportional constant is absorbed into  $\psi_j$  so that they need not sum to 1. In other words, the total annual North-Sea-cod contribution equaled  $\sum_{j=1}^{61} \psi_j c_{t,j} b_t$ , under the transportation hypothesis. See Figure 1 for time plots of the daily sea current and annual spawning biomass data.

Let  $n_{t,s}^0$  be the logarithm of the number of young cod caught in fjord  $s$  in year  $t$ , after adjusted for the intra-specific and the inter-specific effects, as well as the environmental effects on the local cod in the Skagerrak (Chan, Stenseth, Kittilsen, Gjosaeter, Lekve, Smith, Tveite, and Danielssen 2003a, Chan, Stenseth, Lekve, and Gjosaeter 2003b, Stenseth et al. 2006). Specifically, the adjustment is based on an auto-regressive moving-average exogenous-variable (ARMAX) model of auto-regressive moving-average order (2,2), for the logarithmically transformed 0-group cod abundance series. The exogenous variables include abundance of co-existing species (mainly adult Pollock, *Pollachius pollacius*), environmental factors such as water temperature and North Atlantic Oscillation, experimental larvae releases, as well as the effect of an extensive algal bloom in 1988. The  $n_{t,s}^0$  are residuals from the model fitted to data over the period from 1945 to 1997.

We confine the analysis reported below to eight fjords in the Southern Norway (Figure 2), over the period from 1971 to 1997 over which we have complete data. These eight fjords are reported in the earlier analysis to admit significant transportation effects. We now state the model.

$$n_{t,s}^0 = \kappa_s + \zeta b_t + \sum_{j=1}^{61} \psi_j c_{t,j} b_t + e_{t,s}; t = 1, \dots, 26; s = 1, \dots, 8. \quad (1.1)$$

The  $\kappa_s$ 's can be interpreted as the fjord-specific effect on the cod population, and they may be expected to be close to zero because  $n_{t,s}^0$  are part of the longer residual series from an ARMAX model fitted to data from 1945 to 1997. The term  $\psi_j c_{t,j} b_t$  is a proxy for the amount of cod larvae spawned on day  $j$  of year  $t$  that were transported by current to the local fjord  $s$ . The eight fjords (Figure 2) are located in a relatively small region in Skagerrak, so we adopt the first-order approximation that they received the same amount of contributions from the North Sea cod on the average. (This assumption may be relaxed in a number of ways, e.g. a mixed-effect approach, and will be pursued elsewhere.) The term  $\zeta b_t$  can be interpreted as the contributions of the North Sea adult cod by directly swimming to the Skagerrak and spawning there.

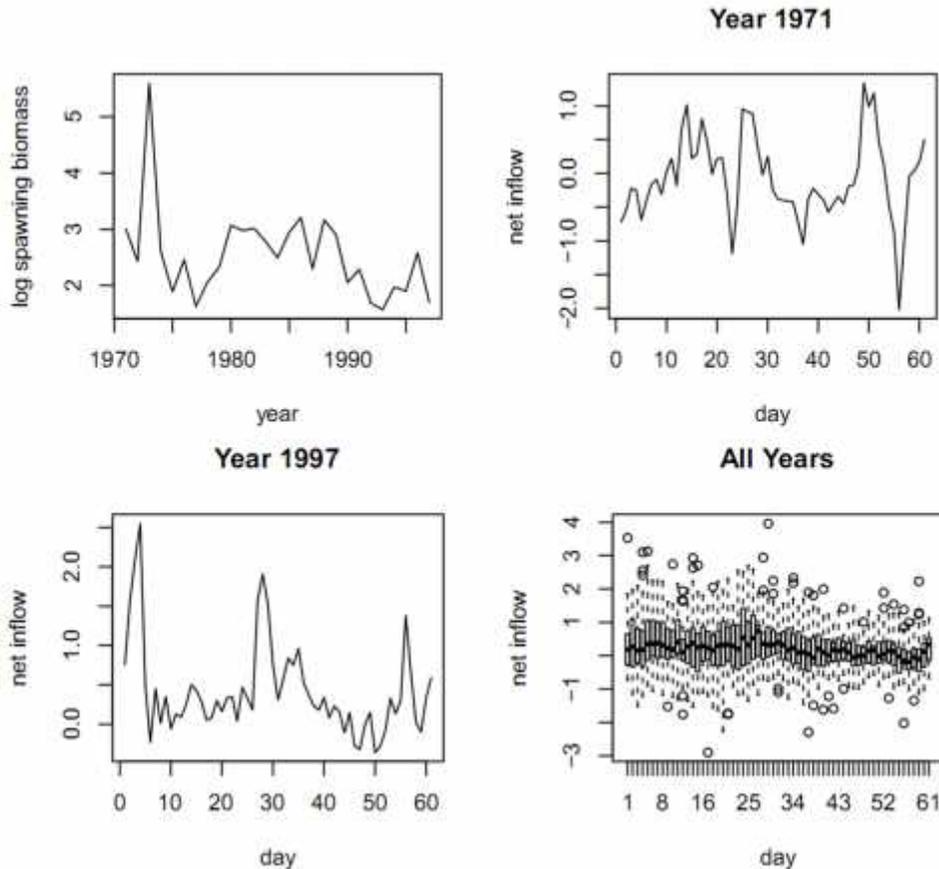


Figure 1. a) Time plot of the annual spawning biomass under the log transformation. b) Time plot of the daily net inflow into the Skagerrak in 1971. c) Time plot of the daily net inflow into the Skagerrak in 1997. d) Box plots of the daily net inflow into the Skagerrak for all years.

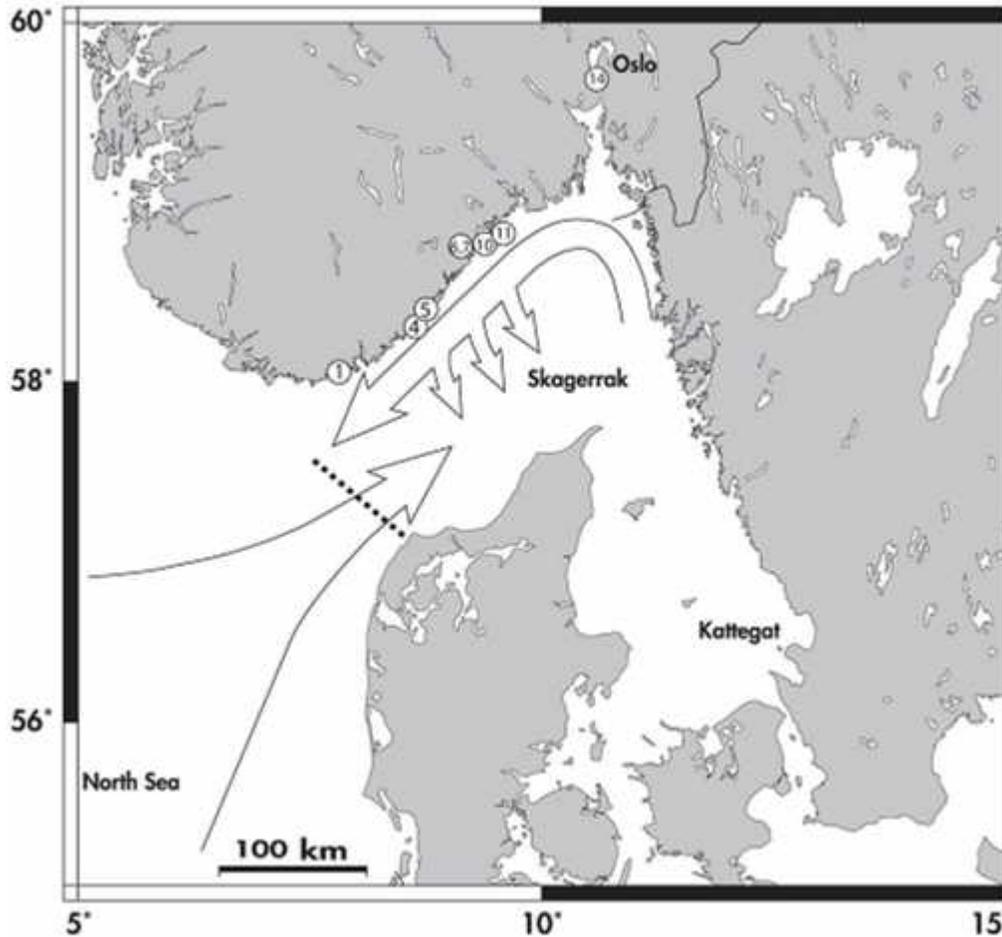


Figure 2. Map of the North Sea-Skagerrak area displaying the eight juvenile cod monitoring stations (numbered circles) on which the statistical analysis is focused, and predominant ocean currents (arrows).

Motivated by the above ecological problem, we first consider the following stochastic regression model describing how a scalar response depends on the aggregate effects of a covariate:

$$Y_t = \alpha^T W_t + \sum_{j=m_1}^{m_2} \psi_j X_{t,j} + e_t, t=1,2,\dots,T, \quad (1.2)$$

where  $Y_t$  are the responses,  $W_t$  and  $X_{t,j}$  are vector-valued and scalar-valued covariates, respectively,  $\alpha$  and  $\psi$ s are parameters and  $\{e_t\}$  is a sequence of independent and identically distributed random variables of zero mean and finite variance; the superscript  $T$  denotes the transpose. The sampling scheme is as follows. Both  $Y$  and  $W$  are measured over regular, basic sampling intervals, e.g. years. Each basic sampling interval is further sub-divided into, say  $M$ , equal intervals, e.g. days, over each of which  $X$  is measured. The measurement of  $X$  in the  $j$ th sub-interval of the  $t$ th basic sampling interval is denoted by  $X_{t,j}$ . The summation limits  $m_1$  and  $m_2$  are assumed to be known integers. The model defined by (1.2) is also known as a transfer-function model (Box, Jenkins, and Reinsel 1994 p. 371) or distributed-lag model (Almon 1965).

In the ecological application,  $Y_t$  is  $t$ -th yearly abundance of some local cod population in Southern Norway, after adjusting for the intra-specific dynamics and other known interventions,  $X_{t,j} = b_t c_{t,j}$  is proportional to the daily amount of eggs in day  $j$  of year  $y$  that could be transported into the Skagerrak, and its parameter  $\psi_j$  may be interpreted as proportional to the spawning probability on the  $j$ -th day, under the transportation hypothesis that cod eggs were carried by sea current into the Skagerrak from the North Sea, and zero otherwise.

The main interest is to estimate  $\psi_j$  as a function of  $j$ . Often, the functional form of  $\psi$  is unknown. Empirical parametric models such as the rational transfer function model and the Almon polynomial lag model are popular methods for estimating  $\psi$ , but they are less useful with complex functional forms. For example, in our biological application,  $\psi$  may be a multimodal function, in which case both the rational transfer function model and the Almon polynomial lag model require many parameters for providing an adequate description of  $\psi$ . Shiller (1973) introduces a nonparametric approach for estimating a smooth  $\psi$  function by postulating a smoothness prior on the second difference of  $\psi$ , but otherwise putting no constraints on  $\psi$ .

$$(1-B)^2 \psi_j = \eta_j, \tag{1.3}$$

where  $B$  is the backshift operator defined by  $B^k \psi_j = \psi_{j-k}$ , for  $k = 0, 1, 2, \dots$ , and  $\eta_j$  are iid normal with zero mean and variance  $\sigma_\eta^2 > 0$ . That is, a hierarchical model is employed. Shiller discussed both the use of a fully Bayesian analysis with non-informative priors as well as a sort of empirical Bayes approach where  $\sigma_\eta^2$  is specified by some rule of thumb. Using a smoothness prior on the second differences codifies the belief that the  $\psi$ 's have small "curvature" as a function of  $j$ . For example, in the extreme case that  $\sigma_\eta^2 = 0$ , the  $\psi$ 's fall on a straight line. See, also (Kitagawa and Gersch 1996, p.37). Note that the approach introduced by Shiller is similar to spline smoothing; see Wahba (1990), Wood (2000, 2006a) and Gu (2002). Indeed, Corradi (1977) showed that the estimator of Shiller's approach can be identified as some smoothing spline function on a suitably defined Hilbert space.

However, even the nonparametric approach fails if the number of data cases is small compared to  $m_2 - m_1 + 1$ , the number of  $X_{t,j}$ 's appearing in the model. This problem may be circumvented if there exist a panel of  $S$  time series that share the same transfer function so that information can be pooled across series for estimating  $\psi$ . In our application, we have time-series data from eight fjords in the Skagerrak. Here, we consider this situation so that the  $s$  th series is generated by the model:

$$Y_{t,s} = \alpha^T W_{t,s} + \sum_{j=m_1}^{m_2} \psi_j X_{t,j} + e_{t,s}, t = 1, 2, \dots, T, s = 1, 2, \dots, S, \tag{1.4}$$

where we note that the same  $X$ 's enter into the equation for each component series, but  $W$  and  $e$  may vary across series. For the panel data, the errors are often

contemporaneously correlated although they may be serially independent. Here, we "extend" Shiller's approach to a multivariate stochastic regression model with contemporaneously correlated errors that subsumes the common transfer function model defined by (1.4). However, our approach differs from Shiller's approach in that we use a penalized likelihood approach (Green and Silverman 1994 p. 5). Note that the model defined by (1.4) is a special case of the unit-rank regression model with the left singular vector of the coefficient matrix proportional to the vector whose elements all equal 1 (Reinsel and Velu 1998, Chapter 8). Thus, a generalization of (1.4) is to allow a general left singular vector, which relaxes the strong assumption of common transfer function. We will pursue this generalization elsewhere.

We now outline the organization of the rest of the paper. In section 2, we elaborate on the framework of a multivariate stochastic regression model that subsumes the common transfer function model. A cross-validation approach is outlined for estimating the smoothness parameter. Some large-sample properties of the estimator are derived. In section 4 we apply the proposed method to analyze the Skagerrak cod data. In particular, we estimate the probability density function of the egg spawning date of North Sea cod indirectly based on data on sea current, spawning biomass and adjusted counts of half-year old cod in eight fjords in Southern Norway. In section 5, a simulation study is reported where the simulation model is motivated by the real application. In the simulation study we investigate the empirical power of our approach for detecting multimodality in the  $\psi$  function. We briefly conclude in section 6.

## 2. A multivariate stochastic regression model

Consider the following general regression model with multivariate response and covariate.

$$\mathbf{Y}_t = X_t \mathbf{b} + \mathbf{e}_t; t = 1, \dots, T \tag{2.1}$$

where the dimension of  $\mathbf{Y}_t$  is  $S \times 1$ ,  $X_t$  is  $S \times k$ , and the coefficient vector  $\mathbf{b}$  is  $k \times 1$ . The  $\mathbf{e}_t$ 's are independent and identically distributed as normal with mean zero and variance-covariance matrix  $\Omega$ , and  $\mathbf{e}_t$  is independent of  $X_t$ . (Here, we restrict the errors to be normally distributed for convenience; extension to non-normality is straightforward but it will complicate the iterative estimation procedure below.) This model is rather general and subsumes the common transfer-function model defined by (1.4), upon letting

$$\mathbf{Y}_t = \begin{pmatrix} Y_{t,1} \\ Y_{t,2} \\ \vdots \\ Y_{t,S} \end{pmatrix}; X_t = \begin{pmatrix} W_{t,1}^T & X_{t,m_1} & X_{t,m_1+1} & \cdots & X_{t,m_2} \\ W_{t,2}^T & X_{t,m_1} & X_{t,m_1+1} & \cdots & X_{t,m_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_{t,S}^T & X_{t,m_1} & X_{t,m_1+1} & \cdots & X_{t,m_2} \end{pmatrix} \tag{2.2}$$

$\beta = (\alpha^T, \psi_{m_1}, \dots, \psi_{m_2})^T$ , and  $\mathbf{e}_t = (e_{t,1}, \dots, e_{t,S})^T$ , Consider the case that the dimension of  $\beta$  is high compared to the sample size resulting in near multicollinearity. The multicollinearity problem can be mitigated by exploiting some known "smoothness" property of  $\beta$ . Suppose that the roughness of  $\beta$  can be quantified by the Euclidean norm

of  $\eta = A\beta$  where  $A$  is a known  $m \times k$  matrix. (In the case of the common transfer function model defined by (1.4),  $A\beta$  is essentially the vector of second differences of  $\psi$ .) We can now construct a penalized log-likelihood where a quadratic penalty for  $A\beta$  is used:

$$\ell(\beta, \Omega; \sigma_\eta^2) = -\frac{T}{2} \log |\Omega| - \frac{1}{2} \sum_{t=1}^T (\mathbf{Y}_t - X_t \beta)^T \Omega^{-1} (\mathbf{Y}_t - X_t \beta) - \frac{1}{2} \frac{\beta^T A^T A \beta}{\sigma_\eta^2}, \quad (2.3)$$

where the coefficient  $\sigma_\eta^2 > 0$  quantifies the trade-off between badness of fit and roughness of the parameter. Here, for known  $\sigma_\eta^2$ , the penalized log-likelihood has the Bayesian interpretation that the components of  $\eta$  have joint prior independent and identical normal distribution of zero mean and variance  $\sigma_\eta^2$ . In practice,  $\sigma_\eta^2$  is unknown, and we propose to determine  $\sigma_\eta^2$  by the method of generalized cross-validation. Specifically, for each  $\sigma_\eta^2 > 0$ , a generalized cross-validation score denoted by  $GCV(\sigma_\eta^2)$  and whose definition is given below, is computed for the model with the parameters  $\beta$  and  $\Omega$  estimated by maximizing the corresponding penalized maximum likelihood function; the generalized cross-validatory estimator of  $\sigma_\eta^2$  is set to be  $\hat{\sigma}_\eta^2$ , the (global) minimizer of the GCV function. Finally, we estimate  $\beta$  and  $\Omega$  by their penalized maximum likelihood estimator (PMLE) with  $\sigma_\eta^2$  equal to  $\hat{\sigma}_\eta^2$ .

To implement the proposed estimation scheme, we first elaborate how to compute the PMLE with a fixed  $\sigma_\eta^2$ . In this case, the objective function (2.3) can be maximized by iteratively alternating the updating of  $\Omega$  and  $\beta$  as follows. For fixed  $\beta$ , it can be shown that the objective function is maximized with  $\Omega$  equal to

$$\hat{\Omega}(\beta) = \frac{1}{T} \sum_{t=1}^T (\mathbf{Y}_t - X_t^T \beta)(\mathbf{Y}_t - X_t^T \beta)^T. \quad (2.4)$$

On the other hand, for fixed  $\Omega$ , the objective function is maximized at

$$\hat{\beta}(\Omega) = (A^T A / \sigma_\eta^2 + \sum_t X_t^T \Omega^{-1} X_t)^{-1} \sum_t X_t^T \Omega^{-1} Y_t. \quad (2.5)$$

The PMLE of  $\beta$  and  $\Omega$  is then obtained by iterating (2.4) and (2.5) until some stopping criterion is satisfied, e.g., when the relative change in the parameter estimates or the objective function, defined in (2.3), is smaller than some prespecified tolerance level. The iteration can be started by initializing  $\Omega$  as the identity matrix. The iterative procedure bears resemblance to the method of iteratively seemingly unrelated regression technique, see Zellner (1962) and (Hamilton 1994, p. 315). Indeed, it reduces to iteratively seemingly unrelated regression in the absence of penalty.

Next, we extend the cross-validatory score and its variant, the generalized cross-validatory score, for correlated data; see, also, Han and Gu (2008). The basic idea consists of defining the CV or (GCV) in terms of their counterparts for the decorrelated

data. For a fixed  $\sigma_\eta^2$ , write  $\hat{\Omega}$  and  $\hat{\beta}$  for the corresponding PMLEs. Let  $\mathbf{Z}$  be the vector obtained by stacking up the  $\mathbf{Z}_t = \hat{\Omega}^{-1/2} \mathbf{Y}_t$ 's,  $\tilde{\mathbf{X}}$  the corresponding design matrix found by stacking the  $\hat{\Omega}^{-1/2} X_t$ 's, and  $\hat{\mathbf{Z}}$  the fitted values so that

$$\hat{\mathbf{Z}} = \tilde{\mathbf{X}}\hat{\beta} = H\mathbf{Z}, \tag{2.6}$$

where the matrix  $H = [h_{ij}]$  is the hat matrix equal to  $\tilde{\mathbf{X}}W$  where  $W$  is implicitly defined by (2.5) in the vector form  $\hat{\beta} = W\mathbf{Z}$ . The incorporation of  $\mathbf{Z}$  is necessary to the calculation of CV and GCV to account for the contemporaneous correlation. Let  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_{TS})^T$  be the result of stacking the  $\mathbf{Z}_t$ , where  $Z_i$  are scalars. The value of  $CV = CV(\sigma_\eta^2)$  is defined by the following formula:

$$CV(\sigma_\eta^2) = \sum_{i=1}^{TS} \frac{1}{TS} \frac{(Z_i - \hat{Z}_i)^2}{(1 - h_{ii})^2}. \tag{2.7}$$

GCV can also be defined similarly by replacing  $1 - h_{ii}$  by the average.

$$GCV(\sigma_\eta^2) = \sum_{i=1}^{TS} \frac{1}{TS} \frac{(Z_i - \hat{Z}_i)^2}{\left(1 - \frac{1}{TS} \sum_{j=1}^{TS} h_{jj}\right)^2}. \tag{2.8}$$

In principle,  $\sigma_\eta^2$  can be determined by minimizing the CV or GCV. However, either optimization problem is very computationally intensive. Instead, a computationally more efficient approach consists of (i) updating  $\beta$  and  $\sigma_\eta^2$  jointly, (ii) updating  $\Omega$ , and (iii) repeating (i) and (ii) until convergence; c.f. the method of generalized additive model (GAM) fitting as detailed in Wood (2006a). Since step (ii) admits a closed-form solution, We need only elaborate step (i). Let  $\Omega$  equal to some iterate and be fixed. Re-parametrize  $\sigma_\eta^2$  by the parameter  $\lambda^2 = 1/\sigma_\eta^2$ . Maximization of the objective function (2.3) is then equivalent to minimizing the following expression.

$$\begin{aligned} & \sum_{t=1}^T (\mathbf{Y}_t - X_t\beta)^T \Omega^{-1} (\mathbf{Y}_t - X_t\beta) + \frac{\beta^T A^T A\beta}{\sigma_\eta^2} \\ &= \sum_{t=1}^T \left\{ \Omega^{-\frac{1}{2}} \mathbf{Y}_t - \Omega^{-\frac{1}{2}} X_t\beta \right\}^T \left\{ \Omega^{-\frac{1}{2}} \mathbf{Y}_t - \Omega^{-\frac{1}{2}} X_t\beta \right\} \\ & \quad + \lambda^2 \beta^T A^T A\beta. \end{aligned} \tag{2.9}$$

Importantly, (2.9) equals the penalized sum of squares for the classical regression problem with uncorrelated data where the response vector consists of stacking up  $\Omega^{-1/2} \mathbf{Y}_t$ 's and the design matrix obtained by stacking up  $\Omega^{-1/2} X_t$ 's, with the regression coefficient vector subject to the quadratic penalty  $\lambda^2 \beta^T A^T A\beta$ . Within the framework of penalized least squares, we can apply the efficient algorithm developed by Wood (2000) for determining  $\lambda^2$  by GCV and the corresponding penalized estimator for  $\beta$ . Step (ii)

can be implemented via the magic function of the mgcv library of R (Wood 2004), which is done in the simulation and data analysis reported below. The function magic is a quick algorithm to find the smoothing parameter that minimized the GCV function.

The minimum GCV approach for estimating the smoothing parameter can be justified as follows. For fixed  $\lambda^2$  and  $\Omega$ , Denote the optimal estimator of  $\beta$  minimizing (2.3) by  $\beta(\lambda^2, \Omega)$ . For selecting a model with correlated data, we may use the Kullback-Leibler loss function defined by the formula

$$K(\lambda^2, \Omega) = E_0(\log(f(\mathbf{Y} | \mathbf{X}; \beta(\lambda^2, \Omega), \Omega) / f(\mathbf{Y} | \mathbf{X}; \beta_0, \Omega_0)))$$

where  $(\underline{\mathbf{X}}, \underline{\mathbf{Y}})$  are independent of  $(\mathbf{X}, \mathbf{Y})$  but have identical joint distribution,  $\beta_0$  and  $\Omega_0$  denote the true parameter values and  $E_0(\cdot)$  denotes taking the expectation of the enclosed expression under the true model. It follows from Theorem 4.2 of Han and Gu (2008) that the proposed minimum GCV criterion is asymptotically equivalent to minimizing the Kullback-Leibler loss, under some regularity conditions detailed in Han and Gu (2008) and the additional condition that  $tr(H) = o_p(T)$ , uniformly in  $\lambda$  and  $\Omega$ . (This essentially follows from the asymptotic equivalence of (3.4) and (3.6) there, under the additional trace condition.) The preceding trace condition is, however, generally valid in our setting since it can be checked that  $tr(H) = tr((A^T A / \sigma_\eta^2 + \sum_{t=1}^T X_t^T \Omega^{-1} X_t)^{-1} \sum_{t=1}^T X_t^T \Omega^{-1} X_t) \leq k$ , if the inverse exists, which happens almost surely, for  $T$  sufficiently large, whenever the Law of Large Numbers hold:  $\sum_{t=1}^T X_t^T \Omega^{-1} X_t / T \rightarrow$  some fixed positive definite  $k \times k$  matrix.

### 3. Computing confidence bands

Conditional on the covariates and the estimated smoothing parameter, it follows from (2.5) that the asymptotic covariance matrix of  $\hat{\beta}$  is approximately given by

$$(A^T A / \hat{\sigma}_\eta^2 + \sum_t X_t^T \hat{\Omega}^{-1} X_t)^{-1} (\sum_t X_t^T \hat{\Omega}^{-1} X_t) (A^T A / \hat{\sigma}_\eta^2 + \sum_t X_t^T \hat{\Omega}^{-1} X_t)^{-1}. \quad (3.1)$$

Along with the preceding frequentist approach to estimating an individual confidence band for  $\beta$ , other methods of estimation are possible. In Section 3.1 we will consider the Bayesian approach of computing a confidence band (Wood 2006a). In Section 3.2 we will elaborate two approaches to computing the confidence band using bootstrap approaches (Efron and Tibshirani 1998, p. 105).

#### 3.1 Bayesian approach

Using a Bayesian approach, the smoothness in  $\beta$  may be formulated in terms of putting a prior on  $\beta$  as

$$f_\beta(\beta) \propto e^{-\frac{\lambda^2}{2} \beta^T A^T A \beta}.$$

This prior is typically improper because  $A$  may not be of full rank. Now the conditional distribution of  $\mathbf{Y}$  given the parameter  $\beta$  can then be stated as

$$f(\mathbf{Y} | \beta) \propto e^{-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)^T \Omega^{-1}(\mathbf{Y}-\mathbf{X}\beta)}$$

So now we can compute the posterior distribution  $f(\beta | \mathbf{Y})$ .

$$f(\mathbf{Y} | \beta) \propto e^{-\frac{1}{2}\{-2\beta^T \mathbf{X}^T \Omega^{-1} \mathbf{Y} + \beta^T (\mathbf{X}^T \Omega^{-1} \mathbf{X} + \lambda^2 A^T A) \beta\}}$$

This posterior distribution turns out to be a normal distribution with variance-covariance matrix

$$\left( \lambda^2 A^T A + \sum_t X_t^T \hat{\Omega}^{-1} X_t \right)^{-1}. \tag{3.2}$$

We can then use the square root of the diagonal elements of (3.2) as standard errors for  $\beta$ , which can then be used for constructing confidence bands for  $\beta$ . The confidence band can then be computed by using the square roots of the diagonals of (3.2) as standard errors. For the case of spline function estimation, the Bayesian confidence band enjoys the property that its across-function-coverage rate is close to the nominal level despite ignoring the variability in the smoothing parameter, see (Wahba 1990, p. 69).

### 3.2 Bootstrap confidence bands

In order to compute the bootstrap confidence bands, we will have to use the fitted values of  $\hat{\mathbf{Y}}_t$  to compute the estimated residuals.

$$\hat{\varepsilon}_t = \mathbf{Y}_t - \hat{\mathbf{Y}}_t, t = 1, \dots, T \tag{3.3}$$

These  $\hat{\varepsilon}_t$ 's would keep the property that they approximately preserve the underlying contemporaneous correlation. From that, we can repeat the following steps  $N$  times.

1. Take a random sample of  $T$   $\hat{\varepsilon}_t$ 's with replacement, where  $\hat{\varepsilon}_t$  are defined in Equation (3.3). These sampled values will be called  $\varepsilon_1^*, \varepsilon_2^*, \dots, \varepsilon_T^*$ .
2. Compute the bootstrap response vector  $\mathbf{Y}_t^*$  as

$$\mathbf{Y}_t^* = \hat{\mathbf{Y}}_t + \varepsilon_t^* \tag{3.4}$$

3. Fit  $\mathbf{Y}_t^*$  from Equation (3.4) onto  $X_t$  with  $\Omega$  and  $\lambda$  fixed at the values estimated from the original analysis to obtain  $\beta^*$ .

From these  $N$  bootstrapped samples, we can construct  $(1-\alpha)\times 100\%$  (individual) confidence band for  $\beta$ ; specifically, for each  $\beta_i$ , the  $(\alpha/2)\times 100$ th percentile and the  $(1-\alpha/2)\times 100$ th percentile of  $\beta_i^*$  are the endpoints of the individual confidence interval (Efron and Tibshirani 1998, Section 13.3).

Similarly, we can define steps for computing a parametric bootstrap confidence band for  $\beta$ . Repeat the following  $N$  times.

1. Simulate  $\varepsilon_t^*, t = 1, \dots, T$  from  $N(0, \hat{\Omega})$ , where  $\hat{\Omega}$  is the variance-covariance matrix estimated from the original analysis.
2. Compute the bootstrap response vector  $\mathbf{Y}_t^*$  as

$$\mathbf{Y}_t^* = \widehat{\mathbf{Y}}_t + \varepsilon_t^* \tag{3.5}$$

3. Fit  $\mathbf{Y}_t^*$  from Equation (3.5) onto  $X_t$  with  $\Omega$  and  $\lambda$  fixed at the values estimated from the original analysis to obtain  $\beta^*$ .

Parametric bootstrapping is suitable in situations where the form of the underlying error distribution is known. On the other hand, the idea of bootstrapping is to be able to compute standard errors without restricting ourselves to a known distribution. So nonparametric bootstrap standard errors would be preferred when an assumption on the error distributions is uncertain and is considered more robust than parametric bootstrap confidence bands. It should be noted that all methods discussed so far do not account for the variations in the smoothing parameters, i.e. the confidence bands so constructed are conditional on the estimated smoothing parameter; however, see (Wood 2006a, Section 4.9) for a related bootstrap approach that accounts for the uncertainty in the smoothing parameters.

#### **4. Inflow of larvae cod as an example**

We now analyze the cod data using the methods developed in the preceding two sections. In estimating (1.1), we impose the following constraints:

$$\psi_i - 2\psi_{i-1} + \psi_{i-2} = \eta_i; i = 3, \dots, D,$$

as well as end constraints:

$$\begin{aligned} \psi_1 &= \eta_1 \\ \psi_2 - 2\psi_1 &= \eta_2 \\ -2\psi_D + \psi_{D-1} &= \eta_{D+1} \\ \psi_D &= \eta_{D+2}, \end{aligned}$$

to ensure that the  $\psi$  function estimates are smooth across the boundaries beyond which they are zero, where  $D = 61$ . These end constraints merely incorporate the prior assumption that  $\psi$  is zero beyond March and April, and maintain the constraint of small roughness across the boundaries. We estimated the model using the method proposed in Section 2. We can compare the model defined by (1.1) with two other models, namely, the model obtained by suppressing the common transfer function (no transportation effects) and the model that only keeps the intercepts in (1.1) (no transportation effects and no direct contributions from North Sea cod). The GCV for the model (1.1) equals 0.6066, whereas the GCVs of the latter two competing models equal 1.049 and 1.038, respectively. Thus, based on GCV, the common transfer-function model is preferred, which lends support to the transportation hypothesis.

Figure 3 plots the  $\psi$  function where the central solid line is the estimated curve and the other two dashed lines enclose the individual 95% confidence limits. For the frequentist confidence band, there is a significant spike in the spawning that begins on March 1st and ends March 24th. For the Bayesian confidence band, the spike is between March 16th and March 21st. The 90% nonparametric bootstrap confidence band indicates an interval of March 12th to March 29th, and the 90% parametric bootstrap confidence band indicates March 18th to March 27th. Certainly all of these approaches are asserting that the spawning is occurring in and around the second half of March. The  $\psi$  estimates sum to

0.1089, so that systematically increasing the daily transportable cod eggs by 1 unit leads to about 11% increase in local cod abundance in Southern Norway, on the average. Table 3.1 reports the rest of the parameter estimates. In particular, all intercept terms are non-significant, and so is the coefficient estimate of  $\zeta$ .

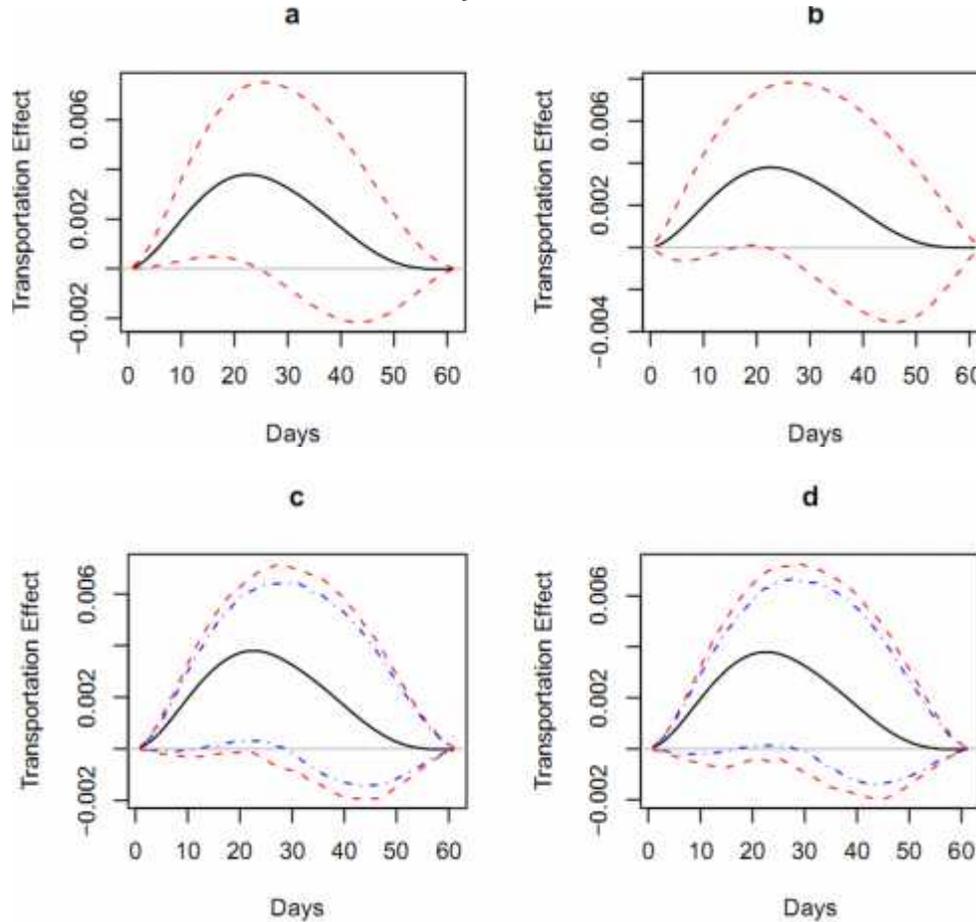


Figure 3. The plot of  $b \psi_j$  versus day  $j$  for the North Sea cod – the central solid line in each graph; day 0 corresponds to March 1. The confidence bands are shown for a) frequentist (dashed line is 95%), b) Bayesian (dashed line is 95%), c) nonparametric bootstrap (dashed line is 95%, dotted-dashed line is 90%), and d) parametric bootstrap (dashed line is 95%, dotted-dashed line is 90%) methods.

Model checking procedures were done on our data by examining the residuals. To save space, diagnostics figures are not shown, but we note that the residual against the fitted value plot suggests no residual nonlinearity and that the residuals have constant variance. The time plot of the residuals (not shown) displays no temporal patterns. The normal Q-Q plot of the residuals appears mostly straight. These model diagnostics suggest that the model assumptions seem reasonable.

In conclusion, there is clear evidence that sea current transported the North Sea cod eggs to the Skagerrak, mainly over the second half of March. Furthermore, the data suggest that the North Sea adult cod did not swim to the Skagerrak to spawn there. These conclusions are consistent with the previous findings that is obtained by assuming

constant  $\psi$  over 2-week periods (Stenseth et al. 2006). However, our new method allows for much more refined conclusions of great importance to the field of marine ecology.

**Table 1: Estimates of model parameters for the Skagerrak cod.**

Parameter	Estimate	SE
$\kappa_1$	-0.22	0.26
$\kappa_2$	-0.21	0.27
$\kappa_3$	-0.42	0.33
$\kappa_4$	-0.20	0.22
$\kappa_5$	-0.02	0.24
$\kappa_6$	-0.00	0.26
$\kappa_7$	-0.13	0.27
$\kappa_8$	-0.35	0.26
$\zeta$	0.12	0.08

### 5. Simulation

We investigate the empirical performance of the proposed method by simulations. We will study two matters: the number of significant modes detected and the error involved in the estimates. The simulation model is motivated by the cod example:

$$\mathbf{Y}_t = \boldsymbol{\kappa} + \zeta \mathbf{b}_t + \boldsymbol{\psi}^T \mathbf{c}_t \mathbf{b}_t + \mathbf{e}_t; t = 1, \dots, T, \tag{5.1}$$

where  $\mathbf{Y}_t$  is the response vector of dimension  $S \times 1$ ,  $\boldsymbol{\kappa}$ ,  $\zeta$ , and  $\boldsymbol{\psi}$  are parameters, and  $\mathbf{b}_t$  and  $\mathbf{c}_t$  are covariates, where the dimension of the vectors  $\boldsymbol{\psi}$  and  $\mathbf{c}_t$  is  $D \times 1$  for both. The error vector  $\mathbf{e}_t$  has a multivariate normal distribution with mean zero and variance-covariance matrix  $\sigma_e^2 P$ , where  $P$  is a correlation matrix.

In the examples below,  $T = 50$ ,  $S = 3$ ,  $D = 61$ ,  $\boldsymbol{\kappa} = [1, 0, -1]^T$ , and  $\zeta = 1$ . The value of  $\mathbf{b}_t$  is determined by taking a random number from a normal distribution with mean zero and standard deviation one, while the value of  $\mathbf{c}_t$  are determined by creating a vector of sixty-one random numbers from a normal distribution with zero mean and unit variance. There will be two levels of the error standard deviation: either 0.05 or 0.1. We will use a correlation matrix that is compound symmetric, meaning that

$$P = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

where there are four levels of  $\rho$ :  $-0.2$ ,  $0.2$ ,  $0.5$ , and  $0.9$ . The  $\psi_j$  equals the probability density function at  $j$  of an equal mixture of two normal distributions, namely  $N(30 - \Delta, 9)$  and  $N(30 + \Delta, 4)$ ,  $\Delta = 10, 5, 0$ , making three levels of the  $\psi$  function. Hence

$\psi$  has two modes that are separated by either 20 units, 10 units, or 0 units (thus making one mode). Each case was simulated 1,000 times. The plots of the three different sets of  $\psi$ 's used can be found in Figure 4.

Simulation results are summarized in Figure 5. Recall the simulation was used to study two aspects of the problem: the number of significant modes detected (based on 95% confidence band) and the error involved in the estimates. Recent works have made a comparison between the Bayesian and frequentist approaches for the case of curves with continuous arguments (Wood 2006b), which we also do here with the confidence bands computed via either the frequentist approach, i.e. using Formula (3.1), or the Bayesian approach. In regard to the first aspect, Figure 5 shows that the simulation catches unimodality and multimodality fairly well when the correlation is high and the error variance is low. (Note that we count the number of modes only for the  $\psi_j$ 's that are significantly different from zero, hence there could be no mode in the curve if none of the  $\psi_j$ 's are significant.) The Bayesian confidence bands tend to catch unimodality and bimodality better than the frequentist confidence bands when the correlation is closer to zero. It seems that the two methods give similar results with a correlation around 0.5. The frequentist confidence bands tend to catch unimodality and bimodality better than the Bayesian confidence bands when the correlation is around 0.9. As far as the second aspect, the mean absolute deviation and mean deviation (unreported) were small as compared to the maximum value of the  $\psi$ 's being estimated. Their standard deviations were small as well and depended proportionately on the error variance.

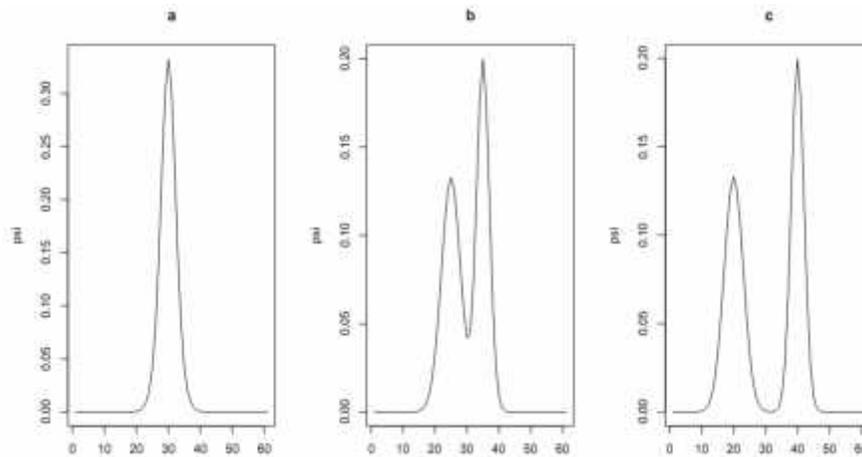


Figure 4. Plots of  $\psi_j$  versus  $j$  for the simulation model;  $\psi_j$  is the probability density function of the an equal mixture of  $N(30 - \Delta, 9)$  and  $N(30 + \Delta, 4)$ , where  $\Delta = 0, 5, 10$ , from left to right.

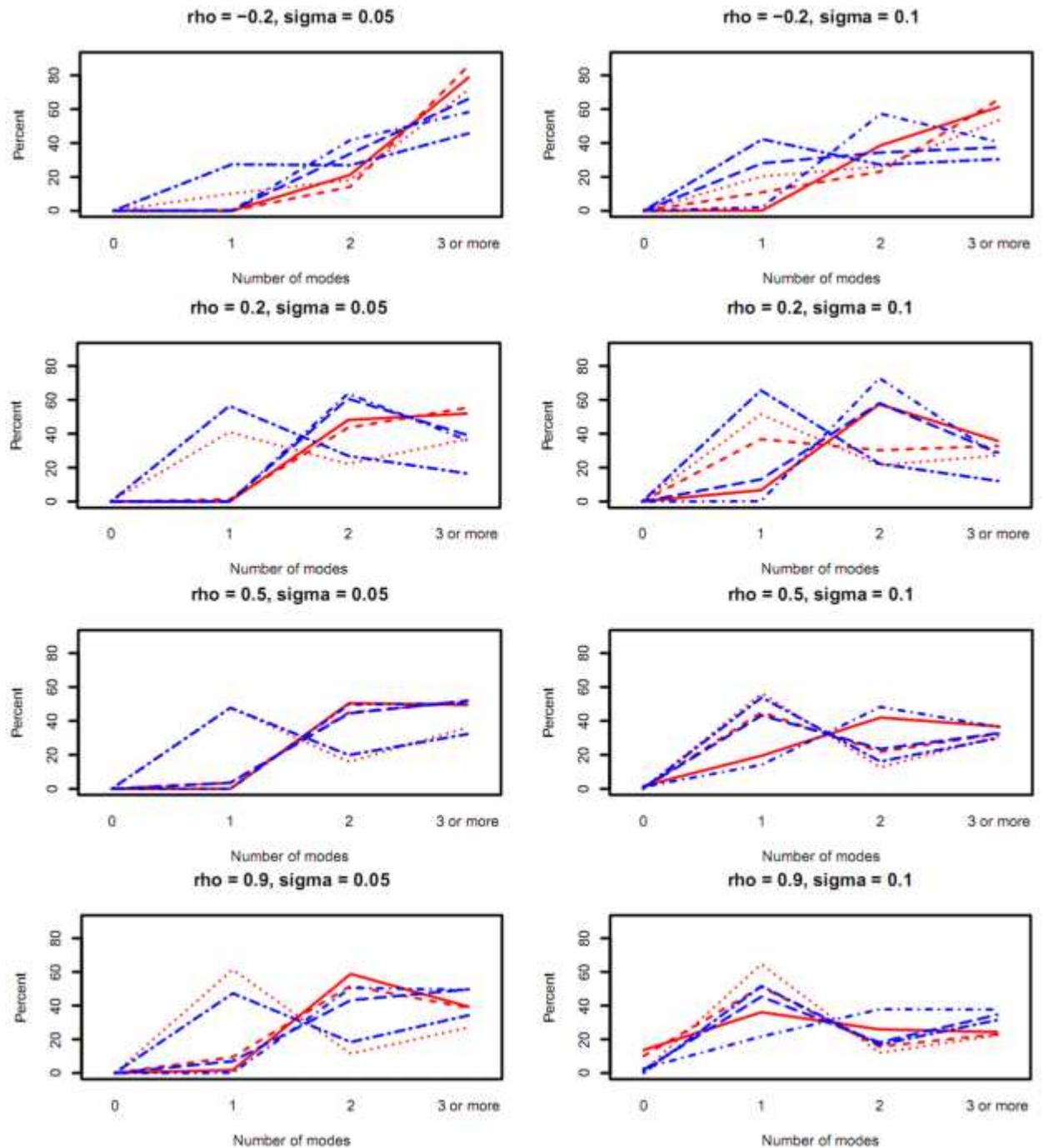


Figure 5. Percents of 0, 1, 2, 3 or more modes detected. Red dotted lines: frequentist,  $\Delta = 0$ ; red dashed lines: frequentist,  $\Delta = 5$ ; red solid lines: frequentist,  $\Delta = 10$ ; blue twodash lines: Bayesian,  $\Delta = 0$ ; blue longdash lines: Bayesian,  $\Delta = 5$ ; blue dashdot lines: Bayesian,  $\Delta = 10$ .

## 6. Conclusion

We have demonstrated the potential usefulness of the method of penalized likelihood for estimating a smooth common transfer function with a panel of short, contemporaneously correlated time series. As illustrated with our marine example, the new method provides

refined conclusion within the field of marine ecology, particularly with reference to how different populations of a marine fish species are interlinked through larvae inflow. As such, our results are of direct relevance for studies on the ecological effects of climate change (see, e.g., Stenseth, Myrseth, Ottersen, Hurrell, Chan, and Lima (2002)).

There are a few interesting future research problems. First, it is of interest to work out the case of non-normal errors in greater details. Second, the common transfer function assumption is a strong one. A more flexible approach is to incorporate random-effects in the transfer function model.

### **Acknowledgements**

We are grateful to Chong Gu for valuable suggestions. We thank Halvor Knutsen for assistance in making Figure 2. The empirical work referred to above was done in collaboration with colleagues at the Institute of Marine Research, Flødevigen IMR, Lysekil and Tjärnö Marine Biological Laboratory; that work was funded by the Norwegian Research Council, the Swedish Council for the Environment and the Nordic Council of Ministers supported this work. Work reported in this paper was supported financially by the Norwegian Research Council, the University of Oslo and the National Science Foundation (NSF DMS-0934617).

### **References**

1. Almon, S. (1965), "The distributed lag between capital appropriations and expenditures," *Econometrica*, 33, 178 – 196.
2. Box, G., Jenkins, G., and Reinsel, G. (1994), *Time Series Analysis, Forecasting and Control*, San Fransisco, California, USA: Pearson Education.
3. Chan, K.-S., Stenseth, N. C., Kittilsen, M., Gjosæter, J., Lekve, K., Smith, T., Tveite, S., and Danielssen, D. (2003a), "Assessing the effectiveness of releasing cod larvae for stock improvement: using long-term beach-seine monitoring data to resolve a century-long controversy," *Ecological Applications*, 13, 3 – 22.
4. Chan, K.-S., Stenseth, N. C., Lekve, K., and Gjosæter, J. (2003b) "Modeling pulse disturbance impact on cod population dynamics; the 1988 algae bloom of Skagerrak, Norway," *Ecological Monographs*, 73, 151 – 171.
5. Corradi, C. (1977), "Smooth distributed lag estimators and smoothing spline functions in Hilbert space," *Journal of Econometrics*, 5, 211 – 219.
6. Efron, B. and Tibshirani, R. (1998), *An Introduction to the Bootstrap*, Boca Raton: Chapman & Hall/CRC.
7. Green, P. and Silverman, B. (1994), *Nonparametric Regression and Generalized Linear Models*, London, England: Chapman and Hall.
8. Gu, C. (2002), *Smoothing Spline ANOVA Models*, New York, USA: Springer-Verlag.
9. Hamilton, J. D. (1994), *Time Series Analysis*, Princeton, New Jersey, USA: Princeton University Press.

10. Han, C. and Gu, C. (2008), "Optimal smoothing with correlated data," *Sankhyā: The Indian Journal of Statistics*, 70-A, 38–72.
11. Kitagawa, G. and Gersch, W. (1996), *Smoothness Priors Analysis of Time Series*, New York, USA: Springer-Verlag.
12. Knutsen, H., André, C., Jorde, P., Skogen, M., Thuroczy, E., and Stenseth, N. C. (2004), "Transport of north sea cod larva into the Skagerrak coastal populations," *Proceedings of the Royal Society of London, B*, 271, 1337–1344.
13. Reinsel, G. and Velu, P. (1998), *Multivariate reduced-rank regression: theory and applications*, New York: Springer-Verlag.
14. Shiller, R. (1973), "A distributed lag estimator derived from smoothness priors," *Econometrica*, 41, 775 – 788.
15. Stenseth, N., Jorde, P., Chan, K., Hansen, E., Knutsen, H., André, C., Skogen, M., and Lekve, K. (2006), "Ecological and genetic impact of Atlantic cod larval drift in the Skagerrak," *Proceedings of the Royal Society of London, B*, 273, 1085–1092.
16. Stenseth, N. C., Mysterud, A., Ottersen, G., Hurrell, J., Chan, K.-S., and Lima, M. (2002), "Ecological effects of climate fluctuations," *Science*, 297, 1292–1296.
17. Wahba, G. (1990), "Spline models for observational data," Philadelphia: SIAM.
18. Wood, S. (2000), "Modelling and smoothing parameter estimation with multiple quadratic penalties," *Journal of Royal Statistical Society, B62*, 413–428.
19. Wood, S. (2004), "Stable and efficient multiple smoothing parameter estimation for generalized additive models," *Journal of the American Statistical Association*, 99, 673–686.
20. Wood, S. (2006a), *Generalized Additive Models: An Introduction with R*, Chapman & Hall/ CRC.
21. Wood, S. (2006b), "On confidence intervals for generalized additive models based on penalized regression splines," *Australian & New Zealand Journal of Statistics*, 48, 445–464.
22. Zellner, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *Journal of the American Statistical Association*, 57, 348–368.