Estimation of harp seal pup production in the Greenland Sea using spatial analysis on aerial survey data

Arnt-Børre Salberg[*], Garry B. Stenson[†], Tore Haug[*], and Kjell T. Nilsen[*]

Abstract

From 14 March to 6 April 2002 aerial surveys were carried out in the Greenland Sea pack-ice to assess the pup production of the Greenland Sea population of harp seals *Pagophilus groenlandicus*. One fixed-wing twin-engined aircraft was used for reconnaissance flights and photographic strip transect surveys of the whelping patches once they had been located and identified. A helicopter assisted in the reconnaissance flights, and was used subsequently to fly visual strip transect surveys over the whelping patches. The helicopter was also used to collect data for estimating the distribution of births over time. Three harp seal breeding patches (A, B and C) were located and surveyed either visually and/or photographically. Using traditional strip-transect analysis, the total estimate of pup production, including the visual survey of Patch A, both visual and photographic surveys of Patch B, and photographic survey of Patch C, was calculated at 98 500 (SE = 16 800), giving a coefficient of variation for the survey of 17.9%. A new approach in analysis of this type of data, using a spatial analysis method, has been developed recently. It is the aim of this work to apply this new method to the data obtained in the 2002 photographic surveys (i.e., patches B and C), and to compare the results obtained (both point estimates and variances) with previous results obtained in the traditional strip-transect analysis.

---

[*] Institute of Marine Research, Tromsø, Norway, Email corresponding author: arntbs@imr.no

[†] Northwest Atlantic Fisheries Centre, Dept. Fisheries and Oceans, Science Branch

1 Introduction

Estimating abundance and monitoring changes in population size is critical for the management of harp seals and to understand their role in the North Atlantic ecosystem. Harp seals are the most abundant pinniped in the North Atlantic, where they are the focus of the largest marine mammal harvest in the world. Although the three populations have historically been exploited and managed separately, the combined total reported harvest (conducted by Canada, Greenland, Norway and Russia) in 2002 was approximately 450 000 animals (ICES, 2004). Thus, there is considerable interest in assessing the status and monitoring changes in abundance in all three populations in order to manage the respective harvests responsibly. In addition, knowledge of harp seal population size is one factor required in order to estimate the potential influence of this species on other marine organisms, including commercially important fish species.

From 14 March to 6 April 2002 aerial surveys were carried out in the Greenland Sea pack-ice to assess the pup production of the Greenland Sea population of harp seals. The method used to estimate the pup production was to count the number of seal on photographs taken along 1 nm separated transects (Haug et al., 2005/6). To extrapolate the counted number of pubs to the number of pubs in the whole patch, a conversion factor determined by dividing the transect interval by the transect width was used. The method does not take into account spatial seal density variability along the transects, but extrapolate the mean density along a transect to area between the transects.

In this paper we propose to model the expected seal density (or seal counts) in a patch as a function of spatial position using a Generalized Additive Model (GAM) (Hastie and Tibshirani, 1990; Wood and Augustin, 2002; Hedley et al., 1999; Augustin et al., 1998). The idea is to assume that the number of pubs counted from aerial photographs are negative binomial distributed but with different mean values. Then, by using thin-plate smoothing splines the GAM provides us with an estimate of the expected seal density at each spatial location in the patch.

2 Materials and methods

2.1 Reconnaissance surveys

Whelping concentrations were located using fixed-wing and helicopter reconnaissance surveys of areas historically used by harp and hooded seals in the Greenland Sea, mainly the pack ice areas along the eastern coast of Greenland between 67° 30'N and 74° 40'N (Fig. 1). Surveys were carried out between 14 March and 5 April 2002 at altitudes between 800-1000 ft. Reconnaissance flights using the fixed-wing aircraft were generally flown as repeated systematic east-west transects spaced 10 nm apart, from the ice edge in the east into the dense drift ice closer to the Greenland shore. Due to ice drift and variation in pupping dates (mid to late March, see Øritsland and Øien (1995)), most areas were surveyed repeatedly to minimize the chance of missing whelping concentrations. Color markers and VHF transmitters were deployed in major whelping concentrations to facilitate relocation and to monitor ice drift.

2.2 Estimates of abundance

2.2.1 Photographic surveys

Fixed-wing aerial photographic surveys were flown using a PA31 Piper Navajo fitted with the gyro mounted Leica RC 30 camera with 15,3 cm lens and AGFA PAN 200 aerographic black-and-white film. The surveys were mainly conducted at an altitude of 191m (includes the entire patch B), but due to low ceilings most transects were carried out at lower altitudes (some as low as 138.5m) in patch C. To avoid variations along transects, altitudes were

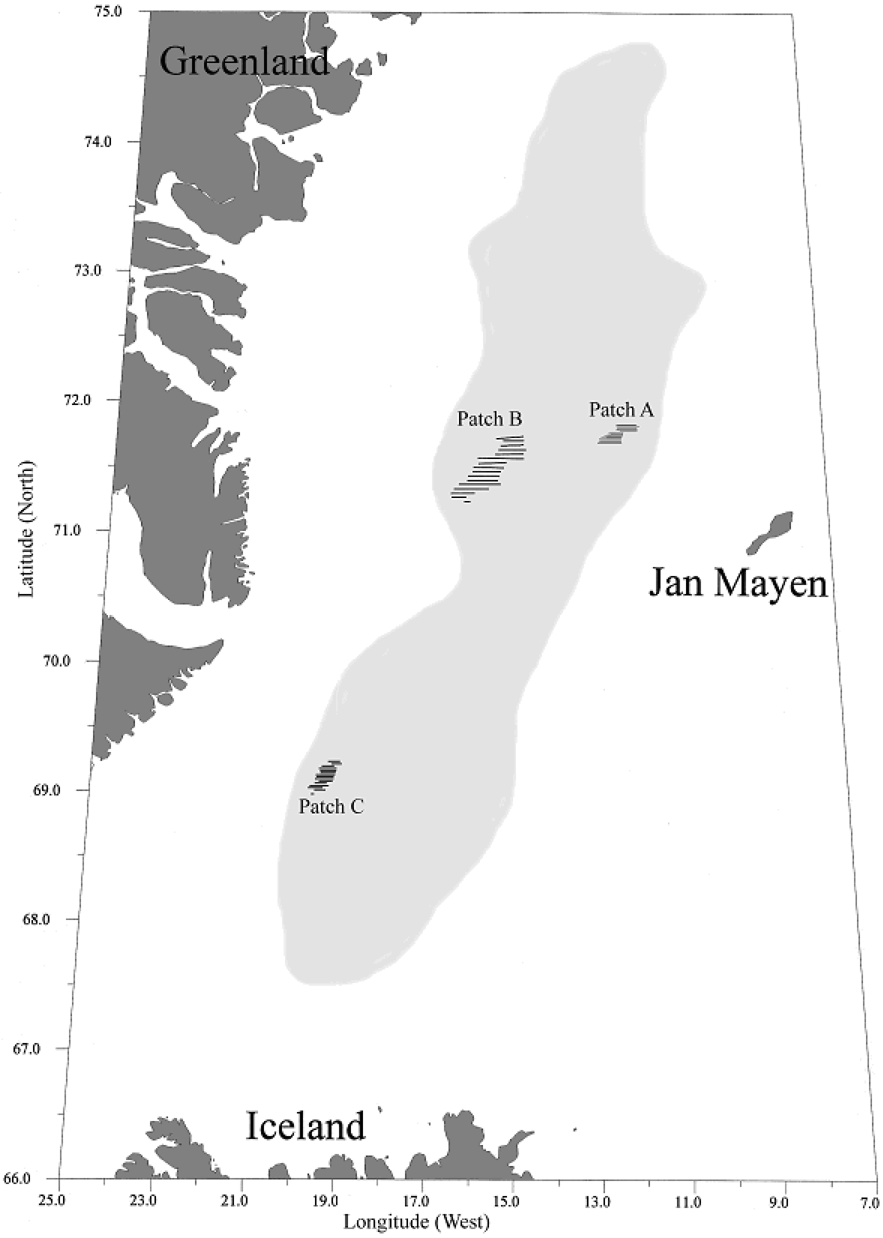

Figure 1: Survey area in the Greenland sea with three seal patches (A, B, and C): Shaded area indicate where fixed-wing reconnaissance surveys where flown.

monitored continuously during the entire photographic survey. The images covered areas varying from 284.1 × 284.1m to 206.2 × 206.2m per photo at altitudes of 191m and 138.5m, respectively. Each transect was allocated coverage according to flying altitude. Photos were taken along each transect at time intervals separated sufficiently to avoid overlap. The camera was turned on when seals were observed on a transect line, turned off if open water occurred for an extended period along a transect, and turned on when ice was encountered again. The photography on a transect line was finished when no seals were observed. Correct altitude and transect spacing were maintained using radar altimeter and a satellite navigation system (GPS).

2.2.2 Photographic counts

Positive prints were examined by two readers. Each frame was examined using an illuminated hand-lens (7-8X magnification). Readers examined a common series of photographs and compared seals identified with a reader with extensive previous experience. Once the cues used to identify seals were consistent among readers, all photos were read once. For each photograph the number and position of all pups were recorded on a clear acetate overlay.

After all photographs were read, the readers re-read a series of their photographs in sequence to determine if identifications had improved over the course of the readings (i.e. the 'learning curve'). Photos were read until the second readings were consistently within 1% of the first. The original readings were replaced with the second readings up to this point. Additional photos were read subsequently to ensure that the first and second reading were consistent.

2.2.3 Conventional survey analysis

Both visual and photographic surveys were based on a systematic sampling design with a single random start and a sampling unit of a transect of variable length.

The estimated number of pups for the ith survey is given by

$$N_i^{conv} = k_i \sum_{j=1}^{J_i} x_i, \tag{1}$$

where J_i is the number of transects in the ith survey, k_i is a weighting factor for the ith survey determined by dividing the transect interval by the transect width, x_j is the number of pups on the jth transect. The estimates of the number of pups along a transect could not be corrected for areas that were not surveyed.

The estimates of error variance V_i^{conv}, based on serial differences between transects (Kingsley et al., 1985), were calculated as

$$V_i^{conv} = \frac{k_i(k_i-1)J_i}{2(J_i-1)} \sum_{j=1}^{J_i-1} (x_j - x_{j+1})^2. \tag{2}$$

2.2.4 Proposed survey analysis

The data were analyzed using the spatial modeling methods based on Generalized Additive Models (GAMs). The proposed technique is similar to those proposed by Hedley et al. (1999), but applies a 2D smoothing function. The counted number of pubs of the ith photograph is

$$N_i = A_i b_i, \tag{3}$$

where A_i is the area covered by the ith photograph, and b_i is the density of pubs in photograph i.

Even thought we are dealing with count data, a Poisson error distribution will not be appropriate because of the over-dispersion. We therefore assume that the data are negative binomial distributed (Thurston et al., 2000). The negative binomial distribution model of the count data is defined by

$$P(N_i = n_i|\mu_i, k) = \left(\frac{\Gamma(N_i + k)}{\Gamma(k)\Gamma(N_i + 1)}\right)\left(\frac{\mu_i}{k + \mu_i}\right)^{N_i}\left(\frac{k}{k + \mu_i}\right)^{k}, \tag{4}$$

where $\mu_i = \mathrm{E}\{N_i\}$ and k is a shape parameter. If k is known, the negative binomial distribution would be in the exponential family of distributions. For a given $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^T$, the log likelihood (assuming statistically independent data) of the shape parameter k is (Thurston et al., 2000)

$$\ell(k; \boldsymbol{\mu}) = n[k \log k - \log \Gamma(k)] + \sum_{i=1}^{n}[\log \Gamma(N_i + k) - (N_i + k)\log(k + \mu_i)] + d(N, \boldsymbol{\mu}), \tag{5}$$

for some function $d(N, \boldsymbol{\mu})$.

The canonical link function for the negative binomial distirbution is $\eta_i = \log\{\mu_i/(\mu_i + k)\}$, but has the disadvantage that η_i must be negative (Thurston et al., 2000). We therefore apply the log link, i.e.

$$\mu_i = \mathrm{E}\{N_i\} = \exp\left[\log(A_i) + \theta_0 + S(z_{i1}, z_{i2}))\right], \tag{6}$$

where the offset variable A_i is the area of the ith photograph, θ_0 is an offset parameter to be estimated, and $S(\cdot, \cdot)$ is a smoothing function of the spatial covariables. We will assume that the smoothing function is a thin-plate smoothing spline (Green and Silverman, 1994; Wood, 2003). Note that the expected seal density is then given by

$$\mathrm{E}\{b_i\} = \exp\left[\theta_0 + S(z_{i1}, z_{i2}))\right], \tag{7}$$

and since it is often more preferable to work with the density of seals rather than the exact count, the smoothing function $S(\cdot, \cdot)$ and the parameter θ_0 will be adjusted to model the expected seal density.

Under the assumption of exponential family data, the GAM may be fitted to the data using the local scoring algorithm (Hastie and Tibshirani, 1990). However, since the negative binomial distribution is only included in the exponential family for known k, the local scoring algorithm may only be used to fit the GAM for fixed k. Furthermore, for known μ_is the shape parameter may be estimated as the argument that maximizes the log likelihood function in Eq. (5). After the unknown μs and shape parameter k are estimated, the count data are modeled as negative binomial distributed data, but with different mean values corresponding to different photographs.

2.2.5 Abundance estimation

Once the model has been chosen, we may use the model to predict the seal density at any location in the patch. Hence, the GAM provides a smooth expected seal density surface over survey area. To estimate the total abundance in the patch, we integrate (numerically) the expected seal density surface over space (Augustin et al., 1998).

The method used to estimate the parameters involved in the GAM model are given in the Appendix.

3 Results

3.1 Patch B

A survey of Patch B (occupying an area between 70° 52'N - 71° 25'N and 14° 44'W - 16° 38'W) was successfully completed 29 March (Fig. (?)). Twenty transects were flown in an

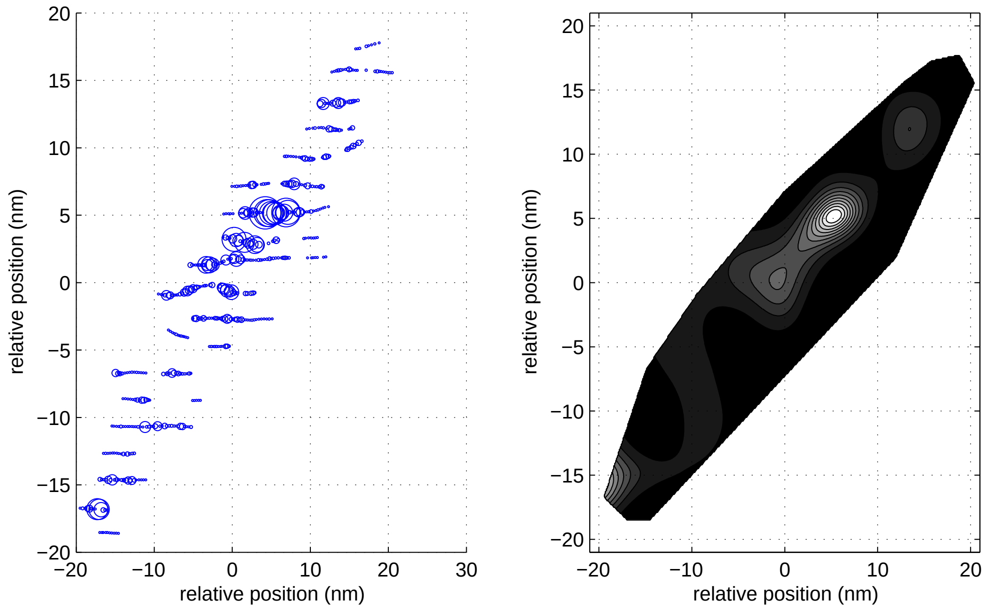

Figure 2: **Left:** Relative sample locations of Patch B of seal density data, with symbol sizes proportional to the row density estimate at the sampling location. **Right:** Contour plot of the expected seal density of Patch B.

east-west direction, spaced 2 nautical miles apart. A total of 5 220 pups was counted on the 521 exposures obtained.

Correcting for reader errors, but not for pups born outside the photographs along a transect, pup production using **conventional survey analysis** was estimated to be 66 545 (SE = 13 534).

Using the **proposed method** the pub production (not correcting for readers errors) was estimated to be 149 760. That is almost twice as using the conventional survey analysis. Fig. 2 (left) show the relative sample locations of Patch B of seal density data, with symbol sizes proportional to the row density estimate at the sampling location. The right panel show the contour plot of the expected seal density of Patch B obtained from the GAM.

3.2 Patch C

The harp seal whelping Patch C was surveyed with photographic strip transects on 6 April in relatively difficult weather conditions (Fig. 1). However, 14 east-west transects, spaced 1 nautical mile apart, were flown over the whelping patch which covered an area between 69° 01'N - 69° 14'N and 19° 06'W - 19°51'W. A total of 321 exposures were taken, on these 1 282 pups were counted. Including the correction for reader errors, but not for pups born outside of photographs along a transect, a total of 11 166 (SE= 1 202) were estimated to have been born using the **conventional survey analysis**.

Using the **proposed method** the pub production (not correcting for readers errors) was estimated to be 22 700. Hence, we obtained an estimate that is more than twice as high as using the conventional survey analysis. Fig. 3 (left) show the relative sample locations of Patch C of seal density data, with symbol sizes proportional to the row density estimate at the sampling location. The right panel show the contour plot of the expected seal density of Patch C obtained from the GAM.

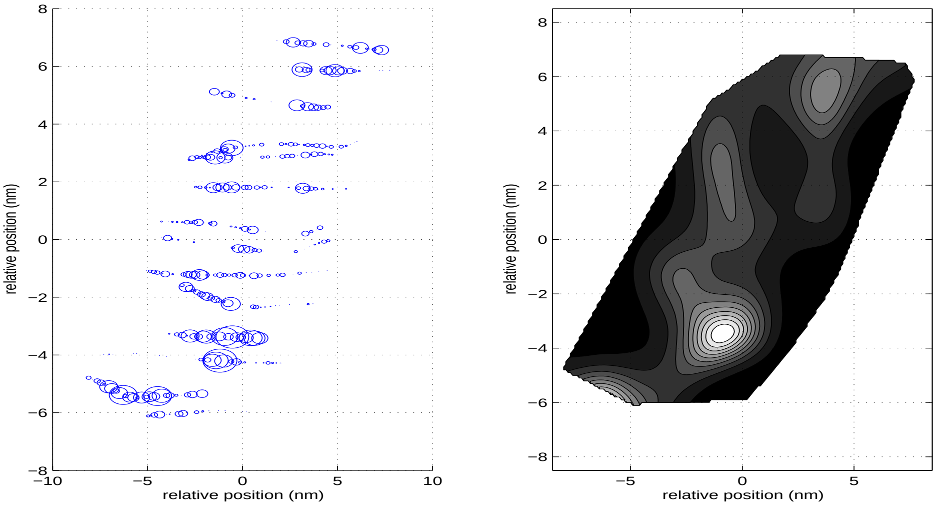

Figure 3: **Left:** Relative sample locations of Patch C of seal density data, with symbol sizes proportional to the row density estimate at the sampling location. **Right:** Contour plot of the expected seal density of Patch C.

4 Discussion

By taking into account spatial inhomogeneity in the density of seal models, we showed that the estimated number of seal pubs changed dramatically, compared to the estimates obtained in (Haug et al., 2005/6). This is due mainly because the total area in patch using the proposed method is larger than the area used in (Haug et al., 2005/6).

However, the method possess some uncertainties. These includes: Which distribution to base the GAM model on. We have chosen the negative binomial distribution, mainly because it allows over-disbursed data. However, other densities may also be used. What are the correct number of degrees-of-freedom? In our paper we have applied the generalized cross-validation method (Wood and Augustin, 2002) to choose the number of degrees-of-freedom, but we may also apply the more computer intensive (standard) cross-validation method.

Note that we do not give an estimate of the abundance since we do no not take into account the temporal distribution of births, neither reading errors of the readers.

Appendix

To estimate the μ_is in the GAM model and the shape parameter k we apply the *alternate profile likelihood algorithm* proposed by Thurston et al. (2000), which alternates between holding k fixed and updating the μ_is through the local scoring algorithm, and holding the μ_is fixed and updating the maximization of Eq. (5).

The algorithm may be summarized as (Thurston et al., 2000):

Alternating profile likelihood algorithm

1. Initialize $\widehat{k}$.

2. Compute $\widehat{\eta}_i = \theta_0 + S(z_{i1}, z_{i2})$, a fitted GAM, using the local scoring algorithm assuming a negative binomial likelihood with $k = \widehat{k}$, and set $\widehat{\mu}_i = e^{\widehat{\eta}_i}$

3. Update k according to $\widehat{k} = \arg\max_k \ell(k, \widehat{\boldsymbol{\mu}})$

4. Repeat steps 2 and 3 until convergence.

Local scoring algorithm

1. Initialize $\widehat{\theta}_0 = \log \overline{b}$.

2. Set $Z = \widehat{\eta}_i + (b - e^{\widehat{\eta}_i})/e^{\widehat{\eta}_i}$, $w_i = ke^{\widehat{\eta}_i}/(e^{\widehat{\eta}_i} + k)$, and $\widehat{\theta}_0 = \overline{Z}$ and fit the weighted additive model using thin-plate smoothing splines with dependent variable Z and independent variables z_1 and z_2.

3. Repeat step 2 until convergence.

Thin-plate smoothing spline estimation

The thin-plate smoothing spline is calculated according to the method suggested by Wood (2003). Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ denote the spatial position where the photographs are taken, and let $\mathbf{y} = [b_1, \ldots, b_n]^T$ be a vector of the observed seal densities. Thin plate smoothing splines are used to find the function S minimizing (Green and Silverman, 1994; Wood, 2003)

$$\|\mathbf{y} - \mathbf{s}\|^2 + \lambda J(S), \tag{8}$$

where $\mathbf{s} = [S(\mathbf{x_1}), \ldots, S(\mathbf{x_n})]^T$, $J(S)$ is a penalty functional measuring the wiggliness of S and λ is a smoothing parameter that controls the trade-off between data-fitting and smoothness of S. For 2D smooting spliens the penalty is defined as

$$J = \iint \left(\frac{\partial^2 S}{\partial x_1^2} \right)^2 + \iint \left(\frac{\partial^2 S}{\partial x_1 \partial x_2} \right)^2 + \iint \left(\frac{\partial^2 S}{\partial x_2^2} \right)^2 dx_1 dx_2. \tag{9}$$

The function minimizing (8) has the form

$$S(\mathbf{x}) = \sum_{i=1}^{n} \delta_i \eta(\|\mathbf{x} - \mathbf{x}_i\|) + \sum_{j=1}^{3} \alpha_j \phi_j(\mathbf{x}), \tag{10}$$

where $\eta(r) = (1/16\pi) r^2 \log(r^2)$, $\phi_1(\mathbf{x}) = 1$, $\phi_2(\mathbf{x}) = x_1$, and $\phi_2(\mathbf{x}) = x_2$. The parameter vectors and $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$ are unknown vectors to be estimated subject to the constraint $\mathbf{T}^T \boldsymbol{\delta} = \mathbf{0}$ and $T_{ij} = \phi_j(\mathbf{x})$. By defining the matrix $\mathbf{E}$ with $E_{ij} = \eta(\|\mathbf{x}_i - \mathbf{x}_j\|)$, the spline fitting problem becomes (Wood, 2003)

$$\text{minimize } \|\mathbf{y} - \mathbf{E}\boldsymbol{\delta} - \mathbf{T}\boldsymbol{\alpha}\|^2 + \lambda \boldsymbol{\delta}^T \mathbf{E} \boldsymbol{\delta}, \text{ subject to } \mathbf{T}^T \boldsymbol{\delta} = \mathbf{0}, \tag{11}$$

with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$.

To select the smoothing parameter λ we used the generalized cross-validation method (see e.g. (Wood, 2000)).

In this paper we consider the low-rank smoothers suggested by Wood (2003). The smoothers are constructed by considering a low rank approximation of $\mathbf{E}$ based on eigen-decomposition. Wood (2003) provide a full description of the method.

References

N. H. Augustin, D. L. Borchers, E. D. Clarke, S. T. Buckland, and M. Walsh. Spatiotemporal modelling for the annual egg production method of stock assessment using generalized additive models. *Can. J. Fish. Aquat. Sci.*, 55:2608–2621, 1998.

P. J. Green and B. W. Silverman. *Nonparameteric Regression and Generalized Linear Models*. Chapman and Hall, London, 1994.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. Chapman & Hall, Boca Raton, 1990.

T. Haug, G. B. Stenson, P. J. Corkeron, and K. T. Nilssen. Accepted for publication ICES Journal of Marine Sciences, 2005/6.

S. L. Hedley, S. T. Buckland, and D. L. Borchers. Spatial modelling from line transect data. *J. Cetacean Res. Manage.*, 1(3):255–264, Dec. 1999.

M. C. S. Kingsley, I. Stirling, and W. Calvert. The distribution and abundance of seals in the canadian high arctic. *Can J. Fish. Aquat. Sci.*, 42:1189–1210, 1985.

T. Øritsland and N. Øien. Aerial surveys of harp and hooded seal pubs in the Greenland Sea pack ice. In *Whales, Seals, Fish and Man*, pages 77–87. Elsivier Science B. V., Amsterdam, 1995.

S. W. Thurston, M. P. Wand, and J. K. Wiencke. Negative binomial addtive models. *Biometrics*, 56:139–144, Mar. 2000.

S. N. Wood. Modelling and smoothing parameter estimation with multiple quadratic penalities. *J. R. Statist. Soc. B*, 62(2):413–428, 2000.

S. N. Wood. Thin plate regression splines. *J. R. Statist. Soc. B*, 65(1):95–114, 2003.

S. N. Wood and N. H. Augustin. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, 157: 157–177, 2002.

Acknowledgements

Many thanks goes to Louise Burt at University of St Andrews, Scotland, for providing me references, and discussing ideas on how to estimate the number of seal pubs using a GAM framework. Many thanks goes also to Alf Harbitz at Institute of Marine Research, Tromsø, Norway for discussions on geostatistic.