Working Document for WKPELA 2016

A statistical model for estimating fish stock parameters accounting for errors in data: Applications to data for Norwegian Spring Spawning herring.

By

Sondre Aanes

Norwegian Computing Center, Oslo, Norway

Abstract

A model template based on a state space model and structural time series models for fish stock assessment is described. Input to analytical stock assessment models are generally obtained from sample surveys and are thus prone to errors. Analysis of sample survey data show complex error structures in the data available for the models, which the model itself not is able to realistically quantify without prior knowledge. The model described here can utilize the sampling distributions derived from analysis of sample survey data by giving appropriate weights to data points. Simulations show that it can provide improved estimates of stock parameters with better representation of uncertainty. The model is applied to NSS herring with multiple data sources and compared to estimates obtained by other models. The model gives similar estimates of stock parameters compared to other available estimates and appear to give realistic measures of uncertainty given the quality of the available data.

Introduction

Analytical fish stock assessments rely on estimates of catch at age and abundance indices at age. Usually, these estimates are available as point estimates at age and time and are applied as is in various stock assessment models. Traditionally, the uncertainty in catch at age has been ignored (VPA type of models), whereas errors in data are assigned solely to abundance indices in various methods for tuning the VPA. Other methods allow also the catch data to be uncertain, and are often referred to as statistical assessment models. A few examples of statistical models are found in Gudmundsson (1994), Quinn and Deriso (1999), Aanes et al (2007), Gudmundsson and Gunnlaugsson (2012), Nielsen and Berg (2014) (and references therein). Accepting that the data represents noisy observations from a hidden population dynamical system has made the state space representation of the system tractable. This representation includes a model for the observations that explicitly formulate properties of the available data. Other methods such as ADAPT (see Quinn and Deriso (1999), and references therein) also includes observation models to control for errors in observations. For example, ADAPT can apply weight to each datum which correspond to the inverse of the estimated variance. Usually, the observation model is formulated in a simple way including parameters that oftentimes is estimated by the model. This is perhaps reflecting that little is generally known about the data available for modelling.

Data used in assessment are generally based on data from multistage sampling. It is well known from sampling theory that multistage cluster sampling may greatly inflate the variance in estimates of population characteristics as compared with simple random sampling of the same number of elementary units (specimens of fish in this paper) from a population (e.g., Sarndal et al. 1992, Lehtonen and Pahkinen 2004). However, this fact has often been ignored in the field of fisheries research (Nelson 2014). Analysis of sample data has revealed that errors in data (both catch and abundance indices) are significant with complex error structures (Aanes and Pennington 2003, Hirst et al 2012, Aanes and Vølstad 2015) including data for NSS herring (Stenevik et al 2015), which emphasize the importance of specifying the observation model appropriately. Estimates of stock parameters do obviously depend on the data and its quality, but the actual effect is difficult to quantify and interpret when the errors not are explicitly accounted for. It is therefore reason to believe that proper use of observations in any stock assessment model would improve the estimates of stock parameters.

In this document we first describe a general time series model for the observational data available for NSS herring which is inspired by the model of described in Gudmundsson (1994). We use this as the basis due to its appealing features concerning utilizing potential time dependencies in the processes. We extend parts of the model to increase the generality, and in particular the observation model is extended. It is proposed to rely on analysis of sample data to parameterize the error structure which effectively provides weight to the observational data.

The process model is extended by allowing the Random Walks in the fishing mortalities, as formulated in Gudmundsson (1994), be replace by AR(1) models. As the time series models for fishing mortality produce very imprecise predictions, the estimate of F in the assessment year (when estimates of catch at age not is available), it is shown how predictions of total catches (e.g. TACs) can be utilized to improve the estimates.

This model have familiarities with SAM (Nielsen and Berg, 2014), but is also different. This is briefly discussed.

This model is nonlinear and high dimensional which offers some challenges in specification of the likelihood for the data and in maximizing the likelihood. To resolve these issues the R library TMB (Template Model Builder) is applied (Kristensen 2014, https://github.com/kaskr/adcomp/wiki). The methods in this library relies on being able to specify the joint likelihood for the observational data as well as the latent processes (which is a simpler task than specifying the marginal likelihood for the data) as it uses efficient numerical methods for both finding the marginal likelihood and for handling with the nonlinearity (Skaug and Fournier, 2006).

The highly variable recruitment of NSS herring results in strong year-classes that can be traced for a long time (suggests choosing a high age for the plus-group) contrasted with weak year-classes which quickly disappear in both catches and surveys (suggest setting a low age for the plus-group). The practical consequences are that it is necessary to treat sequential observations along a cohort with the value of 0 properly, as well as finding the appropriate age for the plus-group. To answer this question, simulations based on empirical data, was used and this is dealt with in an Appendix.

By analyses of available sample survey data it is shown how the observation error models is constructed and how this will influence the weights that each data point is given.

We test the model focusing on the formulation and assumptions on observation errors by simulations based on empirical data for NSS herring. Then we fit the model to input-data for NSS herring and evaluate the results. The estimates are compared by a version of SAM which includes the same observation model as specified here as well as estimates from WGWIDE in 2015 (ICES, 2015).

In addition to simulations, the estimates are evaluated by considering estimates of precision and by regular diagnostics such as residual plots, retrospective runs and the effect of applying different input data (e.g. single fleet runs).

In this document a time varying catchability or estimation of natural mortality, is not considered as any variability in catchability will be confounded with variability in mortality and observation error in abundance indices from surveys. Moreover, it will be shown that the analyzed data provide little support for estimation of both time varying catchability or level of natural mortality.

The analysis in this document is restricted to the time period 1988-2015 and ages 3-15 using input data as reported by WGWIDE in 2015 (ICES, 2015) including catch at age, fishery independent abundance indices at age from the surveys Norwegian acoustic survey on spawning grounds in February/March (Fleet 1), Norwegian acoustic survey in November/December (Fleet 2), Norwegian acoustic survey in January (Fleet 3) and International ecosystem survey in the Nordic Seas (Fleet 5). I will mostly refer to the different fishery independent survey series using the same names as WGWIDE (e.g. Fleet 1). The respective fishery independent surveys are described in the WGWIDE report 2015 (ICES, 2015) and the WGIPS report 2015 (ICES, 2015b) and in the coming report from WGIPS (not completed). Data from analyses of sample survey data will be used to parameterize models for observation error. Increasing the age- or time- range is in principle straight forward and the model can be used for predictions. This is not shown here, but is briefly discussed.

A statistical stock assessment model, XSAM

To be able to utilize potential dependencies in time as well as account for observational error we adopt the state space framework for building a statistical assessment model. A number of statistical assessment model exists, whereas perhaps the most known currently within ICES is SAM (Nielsen and Berg 2014, stockassessment.org), although a series of other models do exist. To not confuse this model with the SAM model we adopt the name XSAM to keep familiarity to a statistical model but flag that it is version X. We build on the model described by Gudmundsson (1994) to create a general and flexible template model that include other documented statistical assessment models as special cases. The main difference between the original formulation and this is 1) that we consider to replace the Random walks with AR(1) models and 2) that we expand on the observation models to be able to utilize the input data in a better way. We use the same notation as Gudmundsson (1994) as far as possible.

The process

The population size at age a at the beginning of year t is connected to the population size of the same cohort the next year through

$$N_{a+1,t+1} = N_{a,t}\exp\left(-F_{a,t} - M_{a,t}\right), \text{ for } a_{min} \leq a \leq A, 1 \leq t \leq T$$

where $Z_{a,t} = F_{a,t} + M_{a,t}$ is the mortality rate at age a in the time interval $(t, t+1)$ such that $\exp\left(-Z_{a,t}\right)$ is the proportion of the cohort surviving from t to $t+1$. The minimum and maximum ages considered are a_{min} and A, respectively, and the time (years) is indexed from 1 to T. Since data for ages above a threshold level A is aggregated into a plus-group, we consider modeling the aggregate as a dynamical pool:

$$N_{A^+,t+1} = N_{A-1,t}\exp\left(-F_{A-1,t} - M_{A-1,t}\right) + N_{A^+,t}\exp\left(-F_{A^+,t} - M_{A^+,t}\right)$$

The fishing mortality is modeled as a hierarchical latent process

$$\log\left(F_{a,t}\right) = \mu_{a,t}^{F} + \delta_{a,t}^{(1)} = U_{a,t} + V_t + \delta_{a,t}^{(1)}$$

where the mean $\mu_{a,t}^{F}$ is a separable model with selectivity $U_{a,t}$, and effort V_t. Deviations from the separable model are allowed by adding random disturbances $\left\{\delta_{a,t}^{(1)}\right\}_{a=a_{min},\ldots,A}$ to the mean. We model these as multivariate normal distributed

$$\left\{\delta_{a,t}^{(1)}\right\}_{a=a_{min},\ldots,A} \sim \mathrm{MVN}\left(\mathbf{0}, \mathbf{\Sigma}^{(1)}\right)$$

with diagonal entries of the covariance matrix $\sigma_{a,t}^{(1)2}$. Although, one might consider to put structures in $\mathbf{\Sigma}^{(1)}$ we will in this document restrict analysis to $\mathbf{\Sigma}^{(1)} = \sigma^{(1)2}\mathbf{I}$.

A separable may impose too strict structure on the fishing mortality and thus not fit all fisheries well. Therefore the selectivity is allowed to change over time and is modelled as a multivariate 1. order autoregressive process:

$$U_{a,t} = \alpha_{aU} + \beta_U U_{a,t-1} + \delta_{a,t}^{(2)}, \;\; a_{min} \leq a \leq a_m$$

and is set constant for ages older than a_m. Note that α_{aU} is age specific but β_U is constant across ages.

$$U_{a,t} = U_{a_m,t}, \quad a \geq a_m$$

with the constraint:

$$\sum_{a=1}^{a_m} U_{a,t} = 0$$

$$\delta_{a,t}^{(i)} \sim N(0, \sigma_{a,i}^2)$$

where the residual vector of the selectivity $\left\{\delta_{a,t}^{(2)}\right\}_{a=a_{min},\ldots,a_m-1}$ is multivariate normal distributed

$$\left\{\delta_{a,t}^{(2)}\right\}_{a=a_{min},\ldots,a_m-1} \sim \mathrm{MVN}\left(\mathbf{0}, \mathbf{\Sigma}^{(2)}\right)$$

with diagonal entries of the covariance matrix $\sigma_{a,t}^{(2)2}$. As for $\mathbf{\Sigma}^{(1)}$ we will in this document restrict analysis to $\mathbf{\Sigma}^{(2)} = \sigma^{(2)2}\mathbf{I}$.

Note that if $\alpha_{aU} = 0\ \forall a$ and $\beta_U = 1$ then this time series model is a random walk and is equivalent to the original formulation of selectivity in Gudmundsson (1994).

The effort V_t includes the latent process Y_t

$$V_t = Y_t + \delta_t^{(3)}$$

where $\delta_t^{(3)} \sim N(0, \sigma_3^2)$ adds random noise to the underlying AR(1) process

$$Y_t = \alpha_Y + \beta_Y Y_{t-1} + \delta_t^{(4)}$$

where $\delta_t^{(4)} \sim N(0, \sigma_4^2)$ such that the effort is a random process where the AR(1) model produce permanent variations in the effort whereas $\delta_t^{(3)}$ just produce deviation from the mean.

This formulation is general and a few special cases are worth noticing:

1) If $\sigma_{a,2}^2 = 0\ \forall a$, and $\alpha_{aU} = 0\ \forall a$ and $\beta_U = 1$, then $U_{a,t} = U_a$, i.e. the selectivity constant over time, and U_a is parameters and not latent variables.
2) If $\sigma_{a,1}^2 = 0\ \forall a$ and $\sigma_3^2 = \sigma_4^2 = 0$, and $\beta_Y = 0$ such that $V_t = \alpha_Y\ \forall t$, and if $\sigma_{a,1}^2 = 0\ \forall a$, then this model is similar to the model for fishing mortality described in Nielsen and Bergh (2014, i.e. the SAM model).

The estimates of intercept and slope in AR(1) models are usually heavily correlated. However, the intercept doesn't have to be estimated since it is given by the slope and mean of the process. To see this consider the AR(1) process $y_i = \alpha + \beta y_{i-1}, i = 1, \ldots, n$. Then the sum over the time series $\sum_{i=2}^{n} y_i = \sum_{i=2}^{n}(\alpha + \beta y_{i-1}) = \sum_{i=1}^{n-1}(\alpha + \beta y_i) = (n-1)\alpha + \beta \sum_{i=1}^{n-1} y_i$. Such that

$$\alpha = \frac{\sum_{i=2}^{n} y_i - \beta \sum_{i=1}^{n-1} y_i}{n-1} \approx \bar{y} - \beta\bar{y} = \bar{y}(1-\beta)$$

Spawning stock biomass is calculated as $SSB(t) = \sum_{a=1}^{A} N_{a,t}\, P_{a,t} w_{a,t}^{s}$, where $P_{a,t}$ and $w_{a,t}^{s}$ are the proportion mature and mean weight in stock at age a and time t, respectively. In this document $P_{a,t}$ and

$w_{a,t}^{s}$ are assumed to be known without any error such that any variability in these variables is not accounted for.

A comment to the SAM model

The SAM model, as described in Nielsen and Berg (2012), conceptually differs from the model outlined here since it adds process error to the cohort equation $N_{a+1,t+1} = N_{a,t}\exp\left(-F_{a,t} - M_{a,t} + \varepsilon_{a,t}\right)$ where $\varepsilon_{a,t} \sim N(0, \sigma_P^2)$. The total mortality may then be written as $Z_{a,t} = F_{a,t} + M_{a,t} + \varepsilon_{a,t}$ such that the error term accounts for any deviation from the model values of total mortality dictated by $F_{a,t}$ and $M_{a,t}$. Since this error term has mean 0 this means that some of the random values will be positive such that the total mortality is reduced. This either means a reduction in natural mortality or that migration is occurring such that "mortality" may even be negative. These may in principle be reasonable assumptions, but the interpretation of this value may be difficult to grasp and will be confounded with observation errors. In this document it will be focused on the situation where this practically is set to 0, as a large value indicates misspecification of the model (i.e. variability in catchability, mortalities and observations not accounted for by the model), and instead we let errors in specification of the process be transferred to uncertainty in parameters for the processes. An extended approach is found in Aanes et al. (2007) where it is attempted to explicitly model the dynamics in total mortality as $M_{a,t}$ itself is modelled as a process.

Observation model

The catches are related to the population through the catch equation

$$C_{a,t} = \frac{F_{a,t}}{Z_{a,t}}\left(1 - e^{-Z_{a,t}}\right)N_{a,t}$$

Since the catch at age data available are estimates based on sample survey programs it is necessary to account for the errors in the estimates and we relate the vector of estimated catch at age in year t, $\widehat{\boldsymbol{C}}_t = \left\{\hat{C}_{a,t}\right\}_{a=a_{min},\ldots,A}$ to the true value by the observation model

$$\widehat{\boldsymbol{C}}_t = \boldsymbol{C}_t\exp(\boldsymbol{\varepsilon}_t^c)$$

Where $\boldsymbol{\varepsilon}_t^c$ are residuals at the log scale of the observations representing the observation or sampling errors. Empirical data suggest that the sampling errors may be approximated by a log normal distribution and hence

$$\boldsymbol{\varepsilon}_t^c \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_t^c)$$

where the sampling error is characterized by the covariance matrix $\boldsymbol{\Sigma}_t^c$.

The estimated survey indices at age $\mathbf{I}_t = \left\{I_{a,t}\right\}_{a=a_{min},\ldots,A}$ are assumed to be proportional to the abundance at the time of the survey with age specific proportionality constants $\mathbf{q} = \{q_a\}_{a=a_{min},\ldots,A}$ such that

$$\mathbf{I}_t = \mathbf{q}\mathbf{N}_t\exp(-\delta_t\mathbf{Z}_t)\exp(\boldsymbol{\varepsilon}_t^I)$$

where $\exp(-\delta_t \mathbf{Z}_t)$ adjusts for the mortality at the time of the survey as δ_t is the fraction of the year passed at the time of the survey, and $\boldsymbol{\varepsilon}_t^I$ are residuals at the log scale of the observations representing the sampling errors in the survey. With a similar argument as for the catch data we assume

$$\boldsymbol{\varepsilon}_t^I \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_t^I)$$

such that the estimated relative abundance is log normal distributed.

For both catch and survey data, the most common assumption appears to be the simplest form that the sampling error is log normal distributed, time invariant and the errors are independent and identically distributed with equal variance for all ages at the log scale. Using catch data as an example, this is obtained by setting $\Sigma_t^c = \sigma_c^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix such that the only parameter involved is σ_c^2, and consequently all observations are given equal weight in the likelihood function. Note that this assumption implies that the variance is proportional to the square of the mean at the original scale. Usually, σ_c^2 is estimated by fitting the model to data without using prior knowledge about sampling variability. It is the inverse of the covariance matrices for the observation models that dictates how the data entering the model should be given weight relative to each other when fitting the model to data in a likelihood setting. For the example $\Sigma_t^c = \sigma_c^2 \mathbf{I}$ this means that all data across ages and time are given equal weight, namely $1/\sigma_c^2$.

Empirical data on observation errors for both catches and abundance indices suggest that the form of the covariance matrices generally are complex and depend on both the sampling intensity and sampling design (e.g. Aanes and Vølstad 2015, Hirst et al. 2012, Hrafnakellson and Stefánsson 2004) and this is also the case for the herring data (see the section *Data* in this document). This challenges the most common and simplest form for the observation error. The covariance matrices must to a large degree be specified a priori as it is not possible to estimate a complex form of the covariance matrices with any reliability (see also Gudmundsson 1994). However, adequate analysis of the survey sample data largely informs these structures. To obtain flexibility and realism we therefore expand the formulation the following way (omitting the time index for convenience)

$$\boldsymbol{\Sigma}^{\boldsymbol{O}} = \left(\sqrt{\mathbf{h}} \cdot \boldsymbol{\sigma}\right)\left(\sqrt{\mathbf{h}} \cdot \boldsymbol{\sigma}\right)^{\boldsymbol{t}} \mathbf{R}$$

Where $\mathbf{h} = (h_1, \dots, h_A)^t$ is a scaling factor of the variances, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_A)^t$ the standard deviations (i.e. standard errors) and $\mathbf{R}$ is the $A \times A$ dimensional correlation matrix for the observations. The scaling factors $\mathbf{h}$ and standard errors $\boldsymbol{\sigma}$ are generally not possible to disentangle, but is included to cover cases where available estimates of standard errors can be regarded as relative, i.e. known up to a scaling constant (see example below). When fitting the model, the values of $\boldsymbol{\sigma}$ and $\mathbf{R}$ are predetermined, whereas $\mathbf{h}$ can optionally be estimated. This formulation gives flexibility concerning which elements of $\boldsymbol{\Sigma}^{\boldsymbol{O}}$ to be estimated and which to be fixed. Examples are:

1. The common assumption of uncorrelated observations with a unknown common variance is obtained by setting $\boldsymbol{\sigma} = (1, \dots, 1)^t$, $\mathbf{R} = \mathbf{I}$ (the identity matrix) and $h_i \equiv h \ \forall i$, i.e. estimate h.
2. If $\boldsymbol{\Sigma}$ is actually known, (through $\boldsymbol{\sigma}$ and $\mathbf{R}$), then this is obtained by fixing $h_i \equiv 1 \ \forall i$.
3. If $\boldsymbol{\Sigma}$ is known up to a scaling constant, estimate $h_i \equiv h \ \forall i$. This is relevant if

a. Random variability in catchability in surveys (e.g. if $q_{a,t} = q_a e^{\varepsilon_{a,t}}$, where $\varepsilon_{a,t} \sim N(0, \sigma_q^2)$) will add to the variance in the estimates caused by the actual sample survey, i.e. $h_i > 1$
b. If the estimation of catch at age is conditioned on the reported total catch, the variability should be increased if the reported total catch is subject to variability.

4. A version of the first example appear to be rather common: Create groups of variances to be estimated to reduce the number of parameters to estimate, e.g. $\mathbf{h} = (h_1, h_2 \dots, h_2, h_3)^t$ which implies estimating one variance for the first age group one common variance for all age groups except the last which is estimated separately (see Nielsen and Berg 2014 for an example).

Note that the use of the scaling constant is to reflect whether we have complete knowledge about the observation error structure, or if it is known up to a scaling constant, which may be relevant for the reasons outlined under point 3.

In this document we will focus on three versions of specification of observation error: 1) Applying iid errors for each data source will be referred to as *observation model 0*, 2) applying empirical values of standard errors, but set correlation to 0, will be referred to as *observation model 1*, and 3) applying empirical values of standard errors and correlations, will be referred to as *observation model 3*.

Fitting the model to data

The likelihood function is first defined by the simultaneous distribution of both random effects and observations. The inference is based on the marginal likelihood for the observations, i.e. the random effects are integrated out of the likelihood. This is achieved by La Place approximation of the integral (Skaug and Fournier, 2006). These steps are carried out using TMB (Template Model Builder, Kristensen 2014, https://github.com/kaskr/adcomp/wiki) which is an R package for fitting models including latent variables (e.g. random effects) to data. This method essentially does the same as ADMB (http://www.admb-project.org/) which the SAM model is based on (Nielsen and Berg 2014, http://www.stockassessment.org) which uses automatic derivation to fit non-linear statistical models. This means that the full likelihood function must be specified in a C-template, whereas TMB finds the derivatives of the likelihood function which is utilized by the various optimizers available in R (in this document I have used the R function nlminb to do the optimization).

Note that since the marginal likelihood value for the fitted models is available it implies that model selection can be aided by the AIC criterion.

For this model, the latent variables are

1. Fishing mortalities $\{F_{a,t}\}_{a=1,\dots,A,\ \ t=1,\dots,T}$ with dimension $A \times T$.
2. If $\sigma_{a,2}^2 > 0$, then selectivity parameters are latent variables $\{U_{a,t}\}_{a=1,\dots,a_m,\ \ t=1,\dots,T}$ with dimension $(a_m - 1) \times T$ (due to constraint), otherwise U_a are treated as fixed parameters, as the mean is given by a separable model.
3. If $\sigma_3^2 > 0$, then the latent effort vector is latent $\{V_t\}_{t=1,\dots,T}$ with dimension $1 \times T$.
4. The effort vector $\{Y_t\}_{t=1,\dots,T}$ with dimension $1 \times T$.

5. If recruitment is defined as a process (see *Initial values* below): $\{R_t = N_{1,t}\}_{t=1,\dots,T}$ with dimension $1 \times T$.

All other variables: initial values, possibly recruits and age specific selectivity, time series parameters for selectivity and/or effort (except for random walks in which they are predetermined), and all variances are fixed parameters to be estimated.

Initial values

The initial values of the model include the abundance the abundance of the recruiting age $\{N_{a_{min},t}\}_{t=1,\dots,T}$ and the abundance at age the first year $\{N_{a,1}\}_{a=a_{min}+1,\dots,A}$. In the SAM model which includes the process error all population sizes, including initial values are latent variables which are estimated. In XSAM where we set the process error to 0, the initial abundances $\{N_{a,1}\}_{a=a_{min}+1,\dots,A}$ are treated as fixed parameters to be estimated.

Gudmundsson determine initial abundances $\{N_{a,1}\}_{a=a_{min}+1,\dots,A}$ by solving the catch equation for $N_{a,1}$ assuming cacthes are given without error the first year such that $N_{a,1} = C_{a,1}\, Z_{a,1}/F_{a,1}(1 - e^{-Z_{a,1}})$ (Gudmundsson 1994). In reality, we have an observation of the catch and the mortality parameters are estimated. To acknowledge that the actual catch predicted estimated in the model, and thus a random variable, we treat the initial abundances as parameters to be estimated.

For the recruits we consider two approaches: 1) to treat the recruiting ages recruiting age $\{N_{a_{min},t}\}_{t=1,\dots,T}$ as fixed parameters to be estimated and 2) modelling $N_{a_{min},t}$ as a process $N_{a_{min},t} = f(\theta)e^{\varepsilon_{rt}^2}$, where the function $f(\theta)$ described the deterministic relationship in the stock recruitment function depending on parameters θ. In the general framework I have implemented various forms of recruitment functions includes Ricker type, Beverton-Holt type, AR(1), Random walk, and the simple $f(\theta) = N_r$, i.e. constant recruitment. Since the recruitment processes for marine fishes generally are poorly known (e.g. Subbey et al. 2014) which also includes NSS herring, modelling recruitment has been beyond the scope of this study, but comparing the simple form $f(\theta) = N_r$ and to treat $\{N_{a_{min},t}\}_{t=1,\dots,T}$ as fixed parameters to be estimated, these two approaches yields practically the same results. However, convergence is faster by modelling recruitment as a process and since it automatically provides a prediction of recruitment, I have chosen to treat recruitment as a process with $f(\theta) = N_r$, where N_r represents mean recruitment and the variance in recruitment σ_r^2 (from $\varepsilon_{rt}^2 \sim N(0, \sigma_r^2)$) are parameters to be estimated determining the latent states of recruiting abundances.

Estimating abundance and fishing mortality in the intermediate year

At the time of the working groups, the basis for the quota advice is the estimated abundance in the beginning of next year (i.e. the size of the fishable stock which forms the basis for the next quota). This means that the assessment must predict the stock ~1 year ahead, depending on the timing of the compilation of the survey results and assessment working groups. In the assessment for herring we have some survey indices for the current year available at the time of WGWIDE, whereas the catch is still unknown. Effectively this means that we do not have direct information about fishing mortality in the intermediate year. Time series models of fishing mortality allow prediction of fishing mortality and thus a prediction of the catch. However, the prediction of fishing mortality, based on a time series model, is very imprecise so that its usefulness for prediction can be questioned. In reality we do have some information about the catch in the assessment year, i.e. the assumed TAC, which not is utilized in the traditional use of the time series models (Gudmundsson 1994, Aanes et al 2007, Nielsen and Berg 2014). For NSS herring the relationship between the catch predicted by the working group in the assessment year correspond very well with the reported catch in retrospect with no systematic biases (see Table 1).

Table 1. Catch prediction made by WGWIDE in the assessment year and the actual reported catch. The prediction and reported catches are proportional with a slope not statistically different from 1. The variability over the period corresponds to a relative standard error of 5.4%. The predictions made by WGWIDE are collected from the respective annual WG reports.

Year	Catch prediction	Reported catch
1996	1400	1220
1997	1500	1427
1998	1302	1223
1999	1302	1235
2000	1250	1207
2001	850	766
2002	850	808
2003	710	790
2004	825	794
2005	1000	1003
2006	967	969
2007	1280	1267
2008	1518	1546
2009	1643	1687
2010	1483	1457
2011	988	993
2012	833	826
2013	692	685
2014	436	461
2015	328*	--

*Catch advice according to ICES is 328.

In the various working groups somewhat different approaches have been used to deal with estimation in the intermediate year. In AFWG, they backshift the survey results to be able to effectively utilize the most recent survey results (i.e. provides an estimate of abundance January 1. in the assessment year) in pair with the catch data and predict the stock based on an assumption on the total catch and the fishing pattern. In WGWIDE, the traditional approach have been to assume that the fishing pattern is equal to last year's estimate applied on a prediction of the total reported catch $\hat{\hat{W}}_y$ (a double hat is used to illustrate that the reported catch may deviate from the actual catch, and the prediction is thus of the reported catch and not actual catch), and use the survey information as is. The prediction of the total catch is usually based on assuming the total catch equals the agreed TAC (catch constraint) or the catch that would be obtained by some fishing mortality based on last years estimated F or the average over the most recent years (F-status quo). WGWIDE have usually used catch constraint to predict the reported total catch.

This means that the estimated abundance at the end of the assessment year (or the intermediate year), or in the beginning of the next year, which forms the basis for the quota advice for next year, must be based on a prediction of the catch at age in the assessment year T. In Appendix A1 it is shown how we can derive the distributional properties of the predicted total catch, which includes sampling variability of catch at age as well as the prediction error of total catch, and thus defines a component in the likelihood function which can be utilized for estimation of the parameters in the model. As it will be shown, the effect of this is to improve the estimates of fishing mortality in the assessment year.

Input data

Sampling error and data weighting

Analysis of sample data according to their respective designs is necessary to estimate both abundance indices and catch at age for commercial landings with reliable estimates of the variability. This crucial for assessing quality of the estimates such that expensive sampling programs can be optimized, and should provide valuable input to stock assessment. The method used for estimating Norwegian catch at age is described in Hirst et al (2012) and is implemented in the ECA software used at IMR for estimation of catch at age and was presented for WGWIDE in 2015 (Salthaug and Aanes 2015). Similarly, estimation of abundance indices based on analysis of sample data is implemented in the StoX software at IMR with methods presented at WGIPS 2015 (ICES, 2015b,), and 2016 (in prep). Common to the approaches are that they provide sampling distributions of the estimates such that standard errors and covariance structures are available. An overview of the respective fishery independent survey series, including key information as samples sizes and coverage is provided in ICES (2015, 2015b).

An example of key features for the estimates of Norwegian catch at age, abundance indices from International ecosystem survey in the Nordic Seas in May (Fleet 5), and from the Norwegian acoustic survey on spawning grounds in February/March (Fleet 1) is shown in Figure 1. A summary of the analyzed sample data is provided in Appendix 2, Figures A2.1-A2.3. The precision in the catch data, measured by its Relative Standard Error (standard error by mean, RSE) is typically around 10% for the most abundant ages (5-10) in the catch, whereas it increases drastically to more than 50% for less abundant ages in the catch. The rather low precision for the catch data, particularly for older ages suggest that fishing mortality cannot be expected to be estimated precisely, particularly for older ages.

The precision in abundance indices from the May survey (fleet 5, Figure 1, Figure A2.2) is somewhat lower (RSE~15-20% for the most abundant ages in the survey), and less abundant ages has lower precision than the more abundant ages for the catch estimates. The precision in abundance indices from the spawning ground survey (fleet 1, Figure 1, Figure A2.3) is variable but generally lower than for Fleet 5. Also these estimates have lower precision for less abundant ages than the more abundant ages for Fleet 1. To ease the interpretation of the RSE as a measure of precision recall that it can be translated into an approximate 95% confidence interval by the mean $\pm$ 2 times RSE of the mean, which means that RSE of 50% means $\pm$ 100% of the point estimate which means that the estimate is very imprecise.

The inevitable cluster sampling for most surveys for fish along with length stratified sampling of ages (for other species) for both survey and catch generally result in a complex correlation structure where a positive correlation often is found for neighboring ages (c.f. Hrafnkelsson and Stefánsson 2004, Aanes and Vølstad 2015) . This is also the case for the data from the acoustic surveys and the correlation structures for estimates of catch at age and abundance at age are shown in Figure 1 and Figures A2.1-A2.3. The mean correlation by distance in age over the years for the May survey shown in Figure A2.4 where the correlation is ~0.5 for neighboring ages and drops to 0 for ages that are ~10 years apart, remarkably stable for all the years. The mean correlation for the spawning ground survey is somewhat lower but also shows correlation over a wide age range (Figure A2.5). The correlation structure for the catch at age estimates for 2008 shown here (Figure 1) are much less pronounced as the correlation drops to 0 for ages close to each other.

The implication of the strong positive correlation is that the amount of information in the estimates is drastically reduced as neighboring ages effectively contain the same information about the abundances, resulting in a reduction of effective sample size for the survey. It should be noted that this represent a candidate to appear as a year effect in the survey since an over- or under- estimate will apply for a series of neighboring ages. Positive correlations of this magnitude are therefore expected to have an impact on the assessment and emphasize the need for appropriate weighing of data used in assessment models.

Observation errors are often assumed to follow a normal distribution at the log scale, or log-normal at the original scale. Analysis of the empirical distributions of the sample distributions of the estimates supports this assumption since we cannot reject the hypothesis that the data (e.g. estimated catch in numbers at age a) do follow a normal distribution on the log scale, and we therefore use the lognormal distribution to approximate the sampling distribution for the data. As already mentioned, the most common assumption is that sampling errors are independent and identically distributed (i.e. the variances are equal across ages). As a consequence, it is assumed that the variance is proportional to the square of the mean while the RSE is constant independent of the mean. In a statistical modeling framework this means that log values of observations would be given the same weight. The empirical findings shown here clearly contrasts this assumption as it show that observations should be given unequal weight, where most weight should be given the most abundant ages both in survey and catch, and less weight to the less abundant ages which in practice means the young and old ones. In this way, analysis of the sample data for both landings and survey provides an objective way of providing appropriate weights to the data both within and across data sources. Figure 2 show the resulting relative weights for estimates of catch at age and abundance indices at age for fleet 5: It is evident that catch at age will be given larger weight than the abundance indices from fleet 5, but it also shows how strong year-classes are given more weight than weak year-classes. Applying these weights in any assessment model utilizing weights, we can deduce that the estimates will be dominated by information from strong year-classes and the catch data.

Analysis of the sample data has not been available for all surveys in all years such that the covariance matrices are not readily available for all data. However, we are able to model the survey variability based on established theory and the empirical data which includes the actual samples sizes (see Appendix A3), and we are therefore able to replace the covariance matrices with estimated covariance matrices based on the available point estimates and knowledge about annual sample sizes, or by assumptions based on average estimates.

Examining the effect of observation error structures on estimates

Although it may seem obvious that observation models should be parametrized correctly, such that observational data used in assessment models are given the appropriate weight, it is difficult to assess the consequence of violating assumptions concerning data properties. We therefore evaluate this by simulation such that we can compare the effect of various assumptions concerning observation errors. A model for a “herring” like population is described in Appendix A4 and is used to simulate realistic population trajectories for herring. Then we parameterize the observation models according to empirical data by establishing standard errors for catch at age and abundance indices at age using the model described in Appendix A3 fitted to data for Norwegian catch at age and the abundance indices from the May Survey. Note that this implies a standard error that varies according to the mean such that the error varies with catch size and abundance size. Having realistic population and observation models we can use them to simulate sets of observations with similar features as the real data. For each replicate we fit the following models for two 2 scenarios

1. Uncorrelated error
    a. Applying the usual assumption of iid, i.e. assuming $\mathbf{\Sigma}_{\mathrm{t}}^{\mathrm{c}} = \sigma_c^2\mathbf{I}$ and $\Sigma_{\mathrm{t}}^{I} = \sigma_I^2\mathbf{I}$ where σ_c^2 and σ_I^2 are estimated (i.e. observation models 0).
    b. Applying actual standard errors but assume unknown scaling, i.e. $\mathbf{\Sigma}_t = \mathrm{h}\boldsymbol{\sigma}_t\ \boldsymbol{\sigma}_t^{\boldsymbol{t}}\mathbf{I}$ (omitting the index for data source) and estimate h for each data source (observation models 1).
    c. Applying the actual covariance matrix by setting $\mathrm{h} = 1$ for both sources, i.e. not estimating any parameters for the observation models (observation models 1).
2. Correlated error according to average correlation matrices for catch at age (Norwegian catch at age) and survey indices (May survey) at age as shown above
    a. Applying the usual assumption of iid, i.e. assuming $\mathbf{\Sigma}_{\mathrm{t}}^{\mathrm{c}} = \sigma_c^2\mathbf{I}$ and $\Sigma_{\mathrm{t}}^{I} = \sigma_I^2\mathbf{I}$ where σ_c^2 and σ_I^2 are estimated (observation models 0).
    b. Applying actual standard errors but assume unknown scaling, i.e. $\mathbf{\Sigma}_t = \mathrm{h}\boldsymbol{\sigma}_t\ \boldsymbol{\sigma}_t^{\boldsymbol{t}}\mathbf{I}$ (omitting the index for data source) and estimate h for each data source but ignoring the correlation structure (observation models 1).
    c. Applying actual standard errors, and correlation structure, but assume unknown scaling, i.e. $\mathbf{\Sigma}_t = \mathrm{h}\boldsymbol{\sigma}_t\ \boldsymbol{\sigma}_t^{\boldsymbol{t}}\mathbf{R}$ (omitting the index for data source) and estimate h for each data source but ignoring the correlation structure (observation models 2).
    d. Applying the actual covariance matrix, including the correlation structure, by setting $\mathrm{h} = 1$ for both sources, i.e. not estimating any parameters for the observation model (observation models 2).

This is repeated a large number of times to be able to extract estimates of mean relative error, relative standard error and coverage to be able to evaluate consequences of assumptions.

Results

Aided by the simulation experiment described in Appendix 4 it is decided to omit all 0 values in data, as the effect on inference will be marginal for the proportion of 0 values that are present in the data sets used here. The alternative of adding a small constant to the data will be detrimental for inference. Note that omitting 0's implies that the corresponding rows and columns of the covariance matrices for observation error also need to be removed.

It is found that the dynamic pool model works satisfactory (Appendix 4) and will thus be used in the rest of this document. It should be noted that increasing the plus-group (lowering the A+ age) might result in reduced precision, but will depend on the error structure in the available data as large uncertainties for older ages itself suggest grouping them into a plus-group.

The effect of observation error structures

The effect of assuming errors are identical and independently distributed, when the variance in reality depend on the mean (on the log scale), is to introduce a bias in the estimates of SSB (negative bias) and average fishing mortality (positive bias), and to reduce the precision (Figure 3). For SSB the increased variability does not ensure that the confidence interval matches the correct level as it misses the actual SSB in 70% of the time whereas the coverage for fishing mortality appears adequate. Informing the observation model with the variances from sample data reduces bias, increases precision and ensures appropriate coverage (although somewhat below nominal level for SSB). The effect of using the correct variance structure, but with an unknown scaling constant for each data source (catch and index) is to reduce precision compared to using the correct value, while the bias and coverage are similar.

When data are correlated according to the empirical data, we find the same result as for uncorrelated data concerning applying the assumption on errors being identical and independently distributed (Figure 4). The effect of ignoring the correlation (i.e. wrongly assume independence) is to overestimate precision such that the coverage of the estimates becomes too low, particularly for SSB. Accounting for the correlation structure slightly reduces the precision but sufficiently so such that the coverage becomes correct also for SSB. The effect of knowing the scaling constant for the observation error for the two data sources is the same as for uncorrelated errors; to reduce uncertainty.

Fitting XSAM to input-data for NSS herring

Identifying appropriate configuration of the model and a first fit: a single fleet run using catch at age and the May Survey (Fleet 5)

First we fit XSAM to data for the time period 1988-2015 ages 3-15 including estimates of catch at age and abundance indices estimated from the International ecosystem survey in the Nordic Seas in May (Fleet 5) only (available from 1996) as reported by WGWIDE in 2015 (ICES 2015). We include age 3 (according to the proposal in Salthaug and Johnsen 2014), and set the catchability constant for ages above 11 as by WGWIDE. The plus-group is set to age 15. It is not possible to estimate σ_3^2 (the variance component for the latent process of effort) based on these data as it tends towards 0 without obtaining convergence. This means that the data suggest that this component may be irrelevant and it is therefore set to 0, meaning that one level in the hierarchy of the model for fishing mortality is removed, but the effort is still modelled as a time series model. This applies for all other data and configuration of the model to be presented later. When parametrizing the XSAM model it is necessary to determine $a_m < A$, the lowest age for which selectivity is constant. For a time varying selectivity it is necessary to set $a_m = 11$ as the model cannot determine the variance in the model for time varying selectivity for older ages. Setting $a_m = 11$, there is some support for a time varying selectivity (AIC=1170.5 compared to AIC=1160.5 for constant selectivity) although the estimates of SSB and F hardly change. This should not come as a surprise as the available data for age 11 and older is very imprecise and therefore does not contain information on this parameter. For constant selection $a_m = 14$, but the estimates are not sensitive values of a_m above 11. We do find an improvement by modelling the time series components of selectivity and effort as AR(1) models compared to Random Walks. This implies estimating β_U for selectivity and β_Y for effort. The actual estimates of these parameters are significantly positive and smaller than 1 for the results reported in this document.

First, we show the result where we utilize the catch prediction for the total catch in 2015 (assumed unknown iid errors) for XSAM in Figure 5. The value 328 thousand tons (corresponding to agreed TAC) was used and an uncertainty in the prediction corresponding to a relative standard error of 10% (which is 5% higher than suggested by the empirical data) was assumed. The effect is to greatly reduce the variability in the estimate of F for 2015 and lower the point estimate of fishing mortality for 2015 compared to not utilizing this information. The effect on SSB or previous estimates on F is marginal. This finding is not sensitive to choice of observation model or which data are included. Catch predictions in current year are used in the rest of the analysis shown in this document and values according to Table 1 are applied for retrospective runs of the model.

The estimates of SSB and average F (ages 5-14) are shown in Figure 6 for 3 versions of observation models: applying equal weights obtained by $\Sigma_\mathrm{t}^\mathrm{c} = \sigma_c^2\mathbf{I}$ and $\Sigma_\mathrm{t}^I = \sigma_I^2\mathbf{I}$ where σ_c^2 and σ_I^2 are estimated by the model (observation model 0), applying empirical values of standard errors but with no correlations according to $\Sigma_\mathrm{t}^\mathrm{c} = h_c\widehat{\boldsymbol{\sigma}}_{ct}\widehat{\boldsymbol{\sigma}}_{ct}^t\mathbf{I}$ and $\Sigma_\mathrm{t}^I = h_I\widehat{\boldsymbol{\sigma}}_{It}\widehat{\boldsymbol{\sigma}}_{It}^t\mathbf{I}$ where the scaling factors h_c and h_I are estimated (observation model 1), and finally introducing the empirical correlations for the observations (observation model 2). First we notice that the estimates in general are similar to WGWIDE in 2015 but somewhat lower for SSB and higher for F. As the results from WGWIDE do not include measures of uncertainty it is difficult to assess whether the estimates are significantly different when they fall outside the confidence limits of XSAM, although they are not likely to be. Accounting for the error in the data reduces the SSB estimates (and increases F) somewhat in the period 2009-2015, and accounting for the correlation structure reduces the SSB estimates even further. This is caused by the down weighing of the survey data in the years 2008-2011 where the survey results were relatively more imprecise at the same time as the correlation in the estimates is high (which effectively reduces the effective sample size in the survey) such that the high survey abundances in that time correspond to a “year-effect” caused by sampling. This effect is also seen from the residuals for the survey (Figure 7): As the survey weight is decreasing, we see a shift from negative to positive residuals around 2005 which can imply a change in survey catchability (confounded with unobserved mortality), and the year effects (series of residuals with the same sign within a year) covering age ranges corresponding to the correlation length (Figure A2.4) is apparent. It is important to note that this change is within the uncertainty of the survey, although the systematic pattern raises some concerns. In addition, the systematic effect is magnified on aggregate measures as SSB as it appears much less of a problem considering time series of numbers at age (not shown).

Figure 6 also include the SAM estimates for comparison. To achieve this we use an analogue setup if the model (e.g. same setting for catchabilities). For the SAM I found that using correlated F's gave a better fit to the data (see Nilesen and Berg 2014 for details) with rather high value of the correlation. In effect this brings the models close to a separability assumption for fishing mortality. For the first version of the observation model, SAM estimates a relatively large process error, and causes a rather big difference between XSAM and SAM. If forcing the process error to be very small SAM produces similar results as XSAM. Interestingly, when introducing reasonable error structures the SAM estimates becomes more similar to the XSAM estimates as the estimated process error is reduced. In the latter case the SAM results are practically indistinguishable from the XSAM estimates when forcing the process error to be very small.

Also the retrospective error depends on both which error structure is (Figure 8). There appear to be a systematic downscaling of SSB (and upscaling of F) in the assessment years 2009-2011, when weighting data equally, reduced when accounting for standard errors to be visually absent when also accounting for the correlation structure.

Including more data: fleet 1, 2 and 3

The catchability in surveys are set constant from ages 11 for fleet 5, 8 for fleet 1, 8 for fleet 2 and 7 for fleet 3, respectively. Otherwise the settings are as before unless otherwise stated.

Including fleet 1 into the analysis has the effect of increasing the estimate of SSB from ~2006 and onwards, whilst the additional inclusion of fleet 2 and 3 has marginal effect on the estimates (Figure 9). The peak in 2009 is estimated lower than by WGWIDE in 2015, whereas the current estimate match very well with WGWIDE. As for the single fleet run with fleet 5 only, the effect of using estimates of sampling error is to reduce the SSB estimates and increase F, and the inclusion of correlation to reduce SSB and increase F even further. Again, notice that none of the estimates are significantly different. The estimates of fishing mortality are imprecise due to the noisy catch data, but fairly similar to WGWIDE, except for the 2015 estimate. The increased estimates of stock size resulted in a decrease in the time trend for the residuals for fleet 5 compared to the single fleet run (Figure A5.1 when adding only fleet 1 and Figure A5.2 when adding fleets 1, 2 and 3), although the main pattern concerning the effect of observation error remains. The residuals for the other fleets (Figure A5.2) also display patterns both in time and to a certain degree cohorts (fleet 2). The retrospective error is shown in Figure 10. The variability in the different fits is generally larger when including more data, than using only abundance indices from fleet 5. Despite the larger variability, the retrospective error is decreased the more information about observation error is used, and whether the errors really are systematic can be discussed. The differences are all within the confidence intervals for the respective assessment years (not included in the plot).

Also for the full data set using covariance matrices for all data, the data support a time varying selectivity as the AIC is reduced from ~2180 to ~2169. This can also be seen by the residuals for catch at age where some systematic residuals are somewhat reduced, but not removed (Figure A5.3). However, estimates of SSB and F changes very little although the estimates of age specific fishing mortalities are improved.

To test the effect of fleet 5 on the estimates, the model was fitted by excluding data from fleet 5. The effect is to increase the SSB and lower the F, but at the same time the uncertainty is drastically increased (Figure A5.4). The residuals of catch at age did not change much, and neither did the residuals for fleet 2.

Increasing the plus-group by setting A+=12 resulted a higher point estimate of SSB and a lower F in the most recent years (Figure 11), but as predicted from the simulation study (Figure A4.2) the variability increased somewhat, and the estimates are not significantly different from setting A+=15. The effects of the different observation models are the same as before. The increased variability is more clearly visible by the retrospective analysis (Figure 12), and the errors appear visually as random over time. The residuals have decreased and some of the systematic behavior appears to have been reduced (Figure A5.6, although difficult to read from the figure).

Discussion and conclusions

It is shown that the data available for assessing NSS herring is rather variable with complex error structures. The level of precision in catch data, which informs fishing mortality, provides a warning sign that it will be difficult to estimate precise levels of fishing mortality, which is confirmed by estimates of F.

By simulations it is shown that it is necessary to appropriately account for the error in the data, which effectively means to provide appropriate weights to data points, to avoid biases in estimates of stock parameters such as SSB and F to obtain useable measures of uncertainties.

The model presented here is not able to estimate the complex forms of observation error without detailed information from the various survey sampling programs used to establish data (catch at age and abundance indices at age) used by fish stock assessment models. It is difficult to see how any model could estimate the error structures without proper analysis of the survey sample data itself since the sources of variability in the model (sampling variability and process variability) most likely are confounded. Uncertainties in data do have an impact on the estimates used for management and emphasize the need for appropriate use of available data to obtain a realistic estimate of levels as well as the uncertainty.

XSAM produces similar values (level and trends) as reported by WGWIDE. Utilizing error structures appear to reduce uncertainty in estimates as predicted by simulations. Although the estimates changes by how we inform the model concerning observation error structures, the estimates are not significantly different. Based on comparisons with SAM and trials with ADAPT type of models, there are strong indications that the data and how they are weighed within a model is more determinant for the estimates than the precise model itself.

When specifying the error structures of the data, we only had data available for parts of the data sources and we predicted structures based on averages of correlations structures and average effective sample sizes for parts where details were missing. The results are sensitive to assumptions made on particularly the correlation structure as it appears to have a big impact on the effective sample sizes (positive correlation effectively reduces effective sample size). It is therefore important to have as good knowledge as possible about these structures, since variability from averages is likely to affect the estimates of stock parameters. This means that the survey sample data (from catch and acoustic- or trawl surveys) should be properly analyzed as far as possible, and that the assumptions concerning the predictions applied here should be carefully evaluated. For example: If careful analysis of historical data not is achievable by methods such as ECA and StoX, the estimates of error structures can be improved by the model presented in Appendix A3, and further informed by a qualitative evaluation of the respective survey sampling designs together with knowledge about effective sample sizes.

In the analysis shown here, all data ages 3-15 that is available have been used applying the philosophy that the observation error would provide each data point with the appropriate weight. This may in some instances be too simplistic, e.g. if the survey in question does not cover the target population. For example, a survey on the spawning grounds may not give a correct representation of the abundance of the juvenile herring in the stock, but may estimate the number of juveniles in the surveyed area precisely, and thus be in conflict with the assumption of the model for catchability. The consequence is likely to be that the uncertainty is transferred to the catchability for the juveniles the specific survey, but the effect on estimates of stock parameters has not been examined for such situations. In general it is not advisable to include erroneous data in model fitting, and the weighing-by-error procedure rely on having performed a careful validation of the data before including them in the analysis as proposed by Salthaug and Johnsen (2014). A general evaluation of a survey for should consider coverage of the survey (e.g. to obtain fishery independent abundance indices) as an important quality indicator is to what extent the survey has covered the entire target population. If some surveys only partly cover the target population the risk of introducing bias into the input data for stock assessment increases and subsequently affects estimates of both catchability- and variance-estimates. This has not been explicitly considered in this document, but is an important candidate to consider for further interpretation of results.

I notice that the estimates of standard errors agrees well with the conclusions made by Salthaug and Johnsen (2014) for determining which data to include in the assessment since the data they propose to exclude often are the same data that are given little weight due to large errors. These data often correspond to data for young and old fish or for weak year-classes and therefore also qualitatively match the weighting procedure that traditionally has been applied by WGWIDE.

The effect of the positive correlations in estimates was to reduce the effective sample size and the effect on the estimates was apparent. This also implies that replacing unknown correlation structures with average structures requires justification. For Fleet 5 and 1; the correlation structure appeared fairly constant for all the analyzed years. However, I had only one years of estimate of the correlations available for catch at age (which showed ~0 correlation), and made the assumption that the correlation was 0 throughout the time series. This assumption needs to be evaluated as the common finding for other sample data from other fisheries suggest that positive correlation is the rule rather than the exception.

The positive correlations of the errors have another important implication, namely that the error, which appears as residuals after the data are fitted to the model, may have the same sign over age ranges corresponding to the correlation length. Therefore, it should not be a surprise that residuals appear as year effects and must be taken into account when using traditional residual plots as diagnostics. It does of course not exclude the possibility of a real year effect in the traditional sense, but it decreases the likelihood.

All fits to the model indicate that the model for fishing mortality (both with and without time varying selectivity) may be poorly specified as the residuals for catch at age appear to follow some cohorts. It is therefore natural to consider including cohort effects in the model for fishing mortality. To test the consequence and potential bias of the possible misspecification we fitted an ADAPT model (Quinn and Deriso p 352) to the time series 1996-2015 including catch at age and abundance indices at age from fleet 5 (which both are complete in this time window). This model includes an observation model and therefore treats observations similar to XSAM, but estimates initial values and fishing mortality for all years and ages such that it does not depend on the structure imposed by the time series models in XSAM. The lack of structure in the model results in imprecise estimates, particularly of F, but estimates of SSB and average fishing mortalities are very similar to the estimates given by XSAM (not shown). This implies that the potential misspecification of the model for fishing mortality is of minor importance for estimating key parameters for NSS herring. Also recall that the precision in catch at age alone suggest that estimates of F will be variable even if the model for F could be improved.

There are some signs of a possible change in catchability or mortality for Fleet 5 (the only data source except catches that is available each year since 1996). However, the uncertainty in the abundance indices (including the reduction of effective sample size due to positively correlated estimates) will inevitably produce rather large uncertainty concerning the actual size of the spawning stock, particularly at the peak in 2009. Therefore, based on these data it will be difficult to conclude on the real cause of change. Based on the assumption of a constant catchability, total mortality could be attempted estimated using the approaches in Aanes et al. 2007 and Bjørkvoll et al. 2012. However, this approach will produce high uncertainties of estimates of absolute levels due to difficulties in estimating the level of natural mortality (and thus the scaling of the entire population).

It can be questioned whether the retrospective results indicate that there are some systematic trends in error of point estimates of SSB and F as time progresses. Particularly the assessment year 2009 estimated a higher level of spawning stock and lower F than later assessments do. Do notice that these changes are smaller if the sampling variance is accounted for and particularly if the covariance structure id fully accounted for, the retrospective differences becomes very small. In the first two cases one might argue that the change is systematic from 2009-2011 as the SSB is gradually decreased while F is increased, but the changes after 2011 does not appear to be systematic. More important is it to notice that the changes in estimates are well within the confidence bounds produced by the model. This means that even for the 2009 assessment, the uncertainty is big and contain information that the point estimate could be well above the actual level at that time.

The results in this document have been restricted to analyses of data ages 3-15. Preliminary analysis show that data for younger ages could be included the same way as proposed here, and that the effect on estimates of SSB and F considered here will be marginal.

It has not been tested to include biomass indices in this framework. It is technically straight forward, but should be discussed conceptually as it is not straight forward how to combine data measured at different scales (numbers and biomasses) in the same likelihood.

This framework can accommodate predictions needed for management advice. This can be done in several ways, but one way is simply to run the model forward for a given TAC and to utilize the time series structure that offers prediction of future selectivity. This can in practice be obtained by the same method as proposed for estimation of fishing mortality in the current year (Appendix A1), where the fitted model can be run forward for a specified TAC (without informing the likelihood). Additional assumptions on weights at age in stock and catch are required, but otherwise the model will do the prediction while maintaining the errors in the prediction such that predictions of SSB and F is given with appropriate measures precision. Generally, prediction of the recruitment is also needed for the prediction of the total population. These predictions are generally very imprecise (e.g. Subbey et al. 2014). However, the short term prediction of SSB and F (ages 5 and above) that is used for management advice will not be affected by the imprecise predictions of recruits.

References

Aanes, S., Engen, S, Sæther, B.-E. and Aanes, R. 2007. Estimation of fish stock dynamical parameters from catch-at-age data and indices of abundance: can natural and fishing mortality be separated? Canadian journal of fisheries and aquatic sciences, 64: 1130-1142.

Aanes, S., and Pennington, M. 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES J. Mar. Sci. 60: 297–303.

Aanes, S., and Vølstad J.H. 2015. Efficient statistical estimators and sampling strategies for estimating the age composition of fish. Canadian journal of fisheries and aquatic sciences 72 (6): 938-953.

Bjørkvoll, E., Grøtan, V., Aanes S., Sæther, B.-E., Engen, S. and Aanes, R. 2012. Stochastic Population Dynamics and Life-History Variation in Marine Fish Species. The American Naturalist, Vol. 180, No. 3 (September 2012): 372-387.

F. Fenton. 1960. The sum of log-normal probability distributions in scatter transmission systems. IRE Transactions on Communication Systems, 8:57–67.

Gudmundsson, G., 1994. Time series analysis of catch-at-age observations. Jour-nal of the Royal Statistical Society: Series C (Applied Statistics) 43, 117–126.

Gudmundsson, G., Gunnlaugsson, T., 2012. Selection and estimation of sequentialcatch-at-age models. Canadian Journal of Fisheries and Aquatic Sciences 69,1760.

Hirst, D., Storvik, G., Rognebakke, H., Aldrin, M., Aanes, S., and Volstad, J.H. 2012. A Bayesian modelling framework for the estimation of catch-at-age of commercially harvested fish species. Can. J. Fish. Aquat. Sci. 69(12): 2064– 2076.

Hrafnkelsson, B., and Stefansson, G. 2004. A model for categorical length data from groundfish surveys. Can. J. Fish. Aquat. Sci. 61(7): 1135–1142.

ICES. 2015. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 25 August - 31 August 2015, Pasaia, Spain. ICES CM 2015/ACOM:15. 588 pp.

ICES. 2015b. Report of the Working Group of International Pelagic Surveys (WGIPS), 19-23 January 2015, ICES HQ, Copenhagen, Denmark. ICES CM 2015/SSGIEOM:05. 279 pp.

Kristensen, K. 2014. TMB: General random effect model builder tool inspired by ADMB. R package version 1.1.

Lehtonen, R., and Pahkinen, E. 2004. Practical methods for design and analysis of complex surveys. John Wiley & Sons.

Nelson, G.A. 2014. Cluster Sampling: a pervasive, yet little recognized survey design in fisheries research. Trans. Am. Fish. Soc. 143: 926–938.

Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158: 96–101.

Salthaug, A. and Johnsen, E. 2014. Validation of Norwegian spring-spawning herring surveys. Working document in the Report of the working group on widely distributed stocks (WGWIDE). ICES CM 2014 / ACOM:15.

Salthaug, A. and Aanes, S. 2015. Estimating the Norwegian catch at age of blue whiting, mackerel, North Sea herring and Norwegian spring-spawning herring with the ECA model. Working document in the Report of the working group on widely distributed stocks (WGWIDE). ICES CM 2015 / ACOM:15.

Sarndal, C.E., Swensson, B., and Wretman, J. 1992. Model assisted survey sampling. Springler-Verlag, New York.

Skaug, H., Fournier, D., 2006. Automatic approximation of the marginal likelihoodin non-Gaussian hierarchical models. Computational Statistics & Data Analysis51, 699–709.

Stenevik, E., Vølstad, J.H, Høines, Å., Aanes, S., Óskarsson, G., Jacobsen, J., and Tangen, Ø. 2015. Precision in estimates of density and biomass of Norwegian spring spawning herring based on acoustic surveys. Marine Biology Research 11 (5): 441-461.

Subbey, S., Devine, J.A., Schaarschmidt, U., and Nash, R.D. 2014. Modelling and forecasting stock–recruitment: current and future perspectives. ICES Journal of Marine Science **71**(8): 2307-2322.

Quinn, T.J., and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.

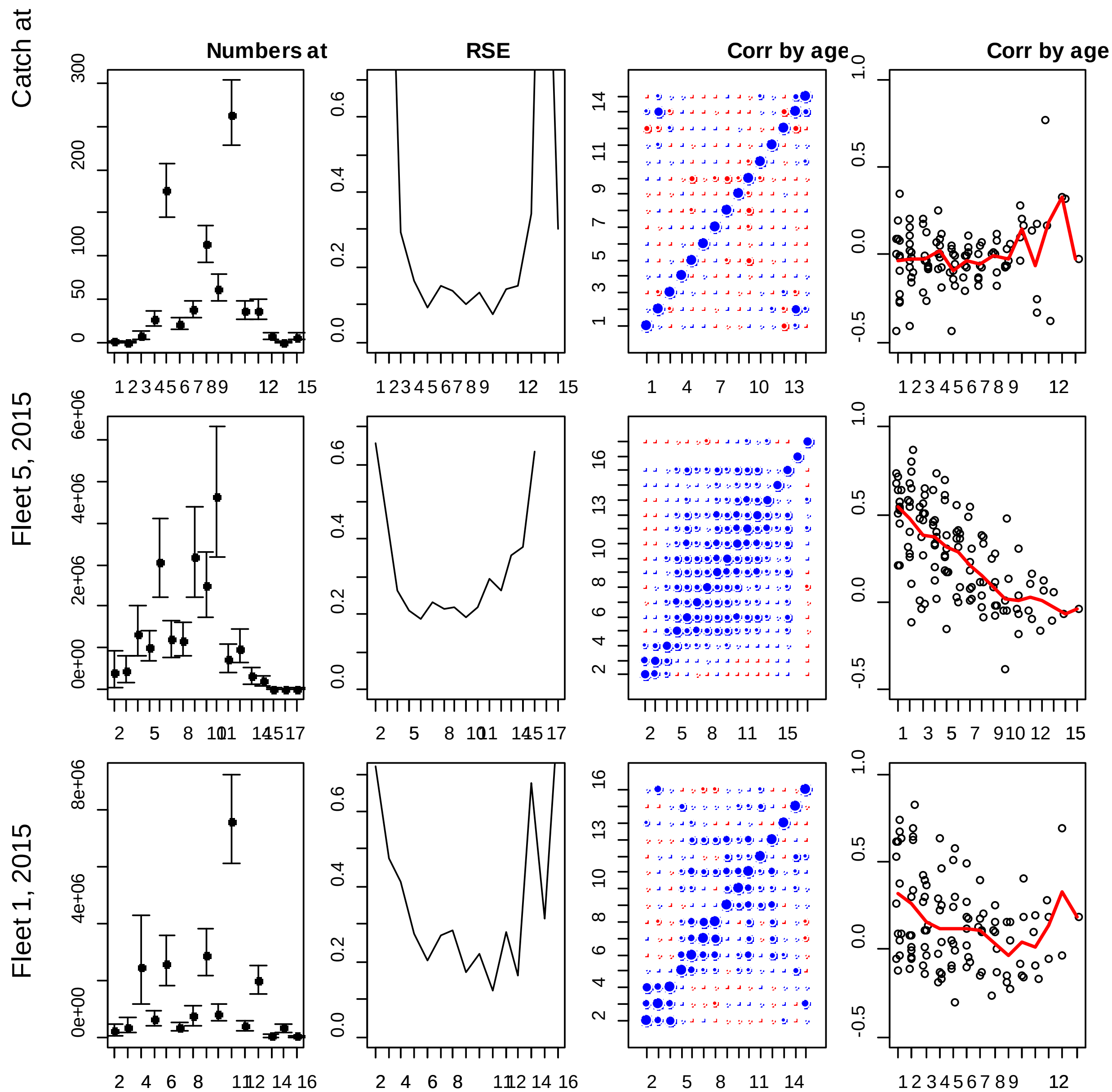


Figure 1. Summary of estimates of Norwegian catch at age in 2008, abundance indices at age from the May survey (Fleet 5) in 2015, and Norwegian acoustic survey on spawning grounds (Fleet 1) in 2015 for NSS herring. Numbers at age (1. column) with 95% confidence intervals, relative standard error at age (2. column), correlation of abundance estimates by age (3. column), and correlation by distance in age (4. column). Estimates of catch at age are based on ECA, while estimates of abundance at age are based on StoX.

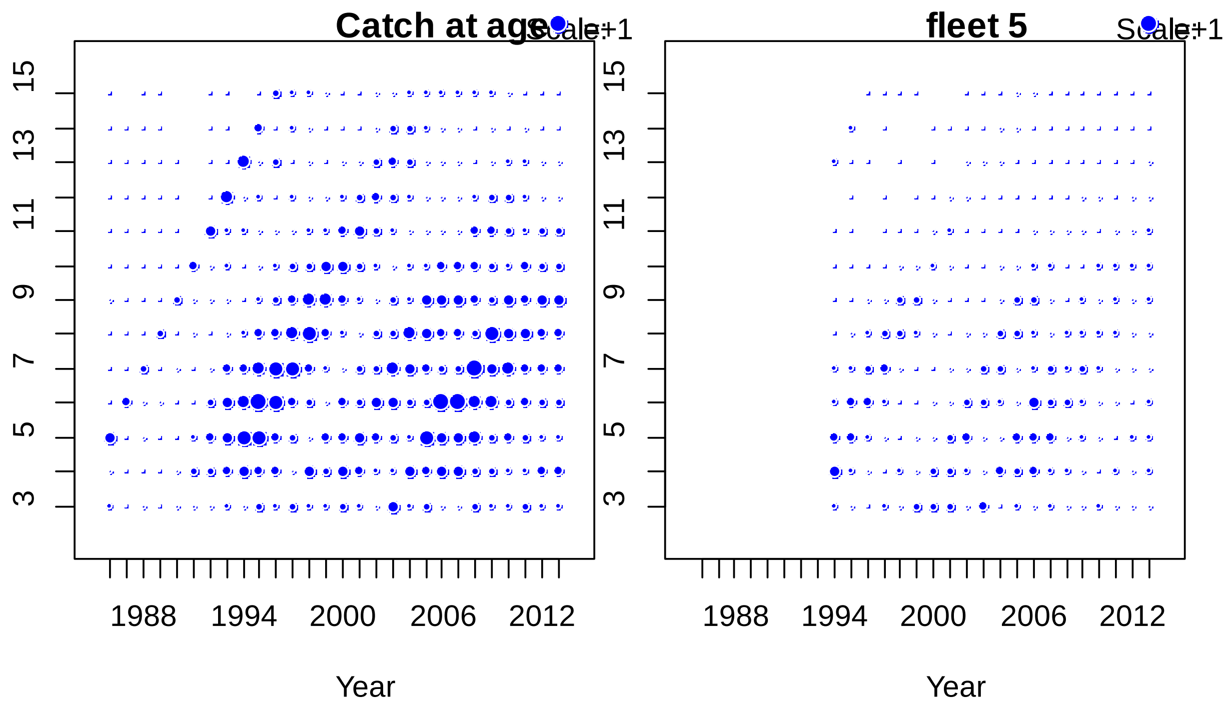


Figure 2. The weights by age and year for estimates of catch at age and abundance index at age from fleet 5. The weights are the inverse of the sampling variances of the estimates.

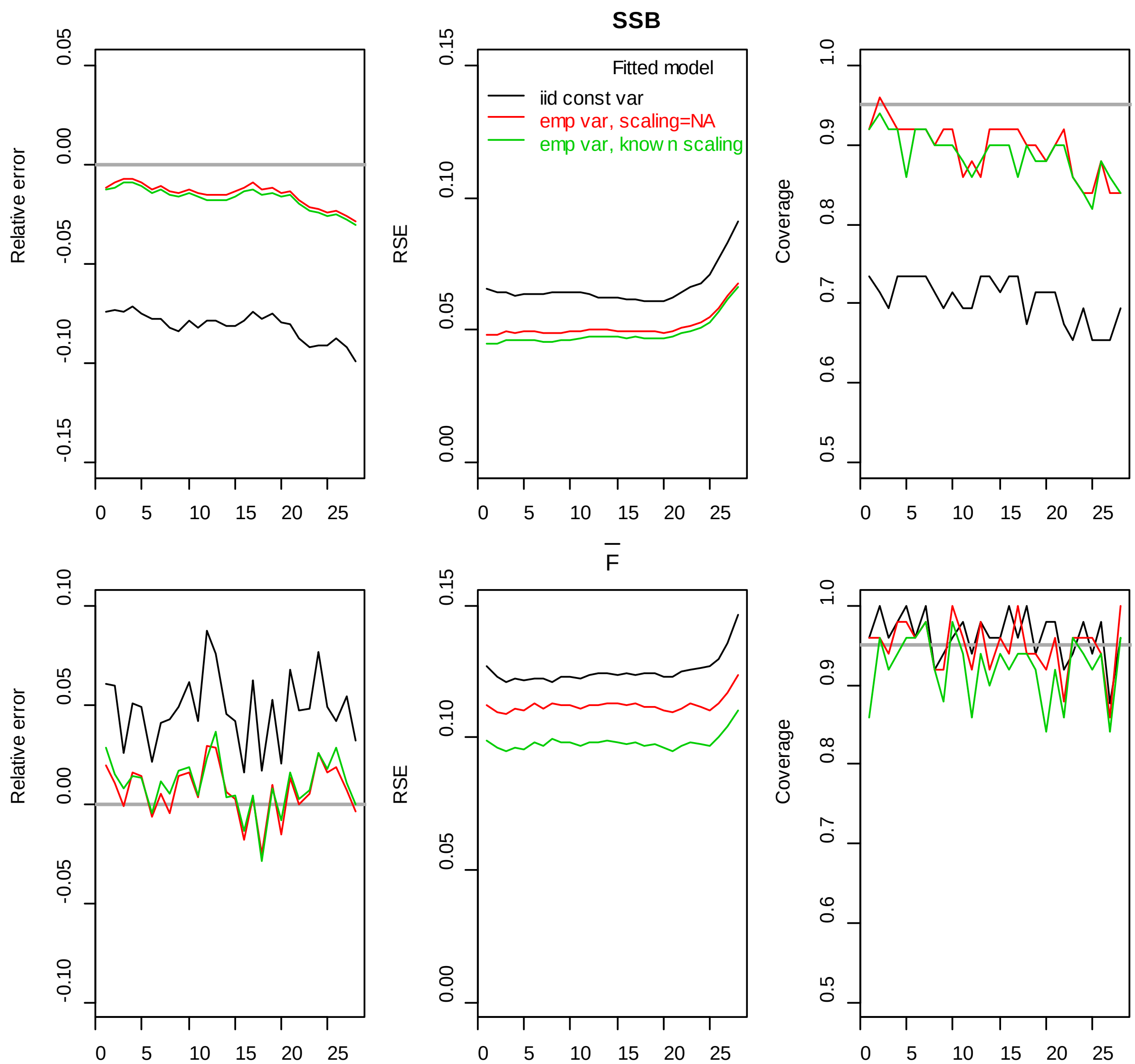


Figure 3. Testing the effect of using correct variance structure on SSB (top row) and average fishing mortality, ages 5-14 (bottom row): The data comes from a distribution with variances that depend on the mean such that the relative standard errors depend on the mean according to empirical data for NSS herring. The correlations of observations are set to 0. The model is fitted by assuming the errors are independent and identically distributed at the log scale (black lines), informing the covariance matrix of the observation error with unequal variances from data but with a unknown common scaling factor for each dataset (catch and index) (red lines), and finally informing the covariance matrix of the observation error with unequal variances from data with known common scaling factor for each dataset (catch and index) (green lines). The statistics are mean relative error (left column), relative standard error (middle column) and coverage (right column) based on 50 replicates of population trajectories with corresponding observations. The nominal level of coverage is set to 95% as indicated by the gray line.

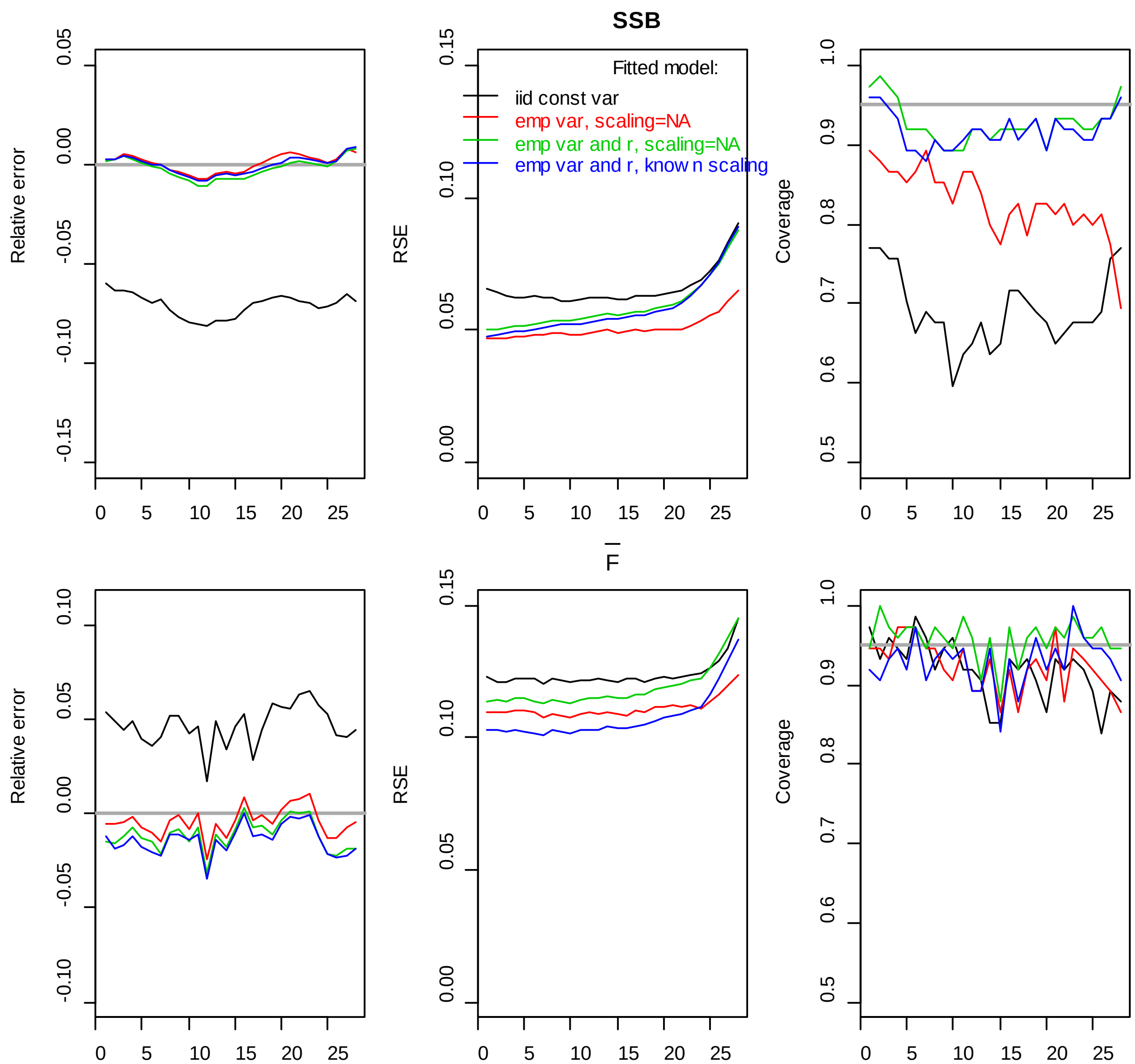


Figure 4. Testing the effect of using correct variance structure on SSB (top row) and average fishing mortality, ages 5-14 (bottom row): The data comes from a distribution with variances that depend on the mean such that the relative standard errors depend on the mean according to empirical data for NSS herring. The correlations of observations are equal to empirical data. The model is fitted by assuming the errors are independent and identically distributed at the log scale (black lines), informing the covariance matrix of the observation error with unequal variances from data but with a unknown common scaling factor for each dataset (catch and index) (red lines), informing the covariance matrix of the observation error with unequal variances and correlation structure from data but with a unknown common scaling factor for each dataset (green lines), and finally informing the covariance matrix of the observation error with unequal variances and correlation structure from data with known common scaling factor for each dataset (blue lines). The statistics are mean relative error (left column), relative standard error (middle column) and coverage (right column) based on 50 replicates of population trajectories with corresponding observations. The nominal level of coverage is set to 95% as indicated by the gray line.

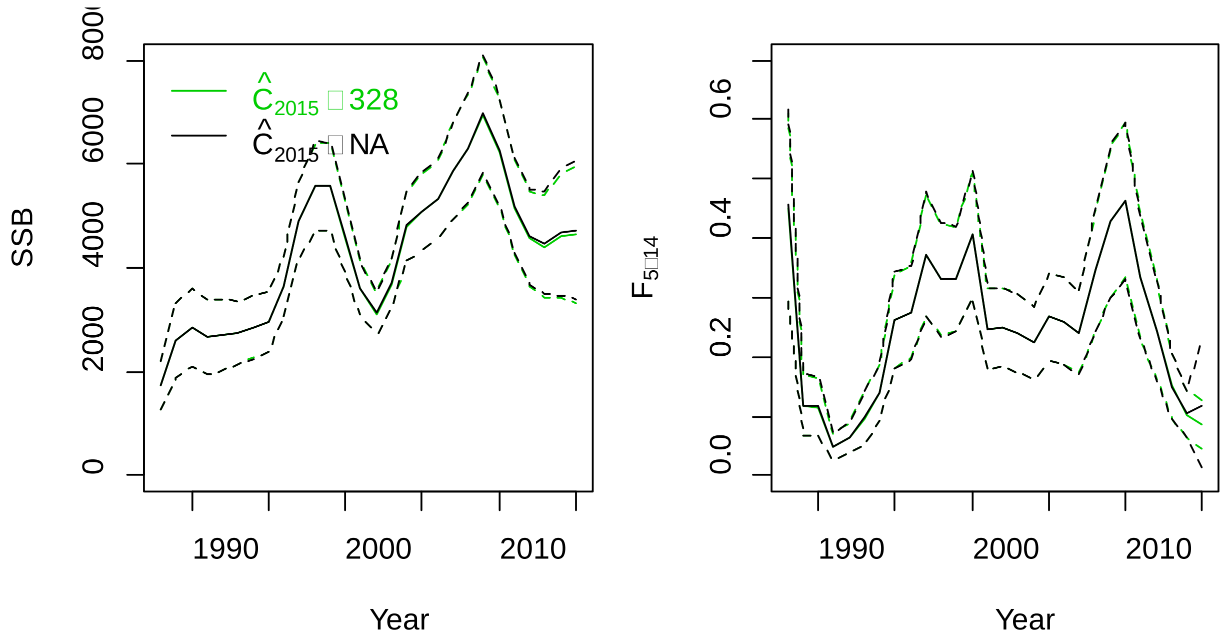


Figure 5. The effect of utilizing prediction of total catch in estimation (green line) compared to not utilizing information on total catch in the current (i.e. 2015) year. The prediction for 2015 was set to 328 thousand tons according to the agreed TAC with an uncertainty corresponding to a relative standard error of 10%.

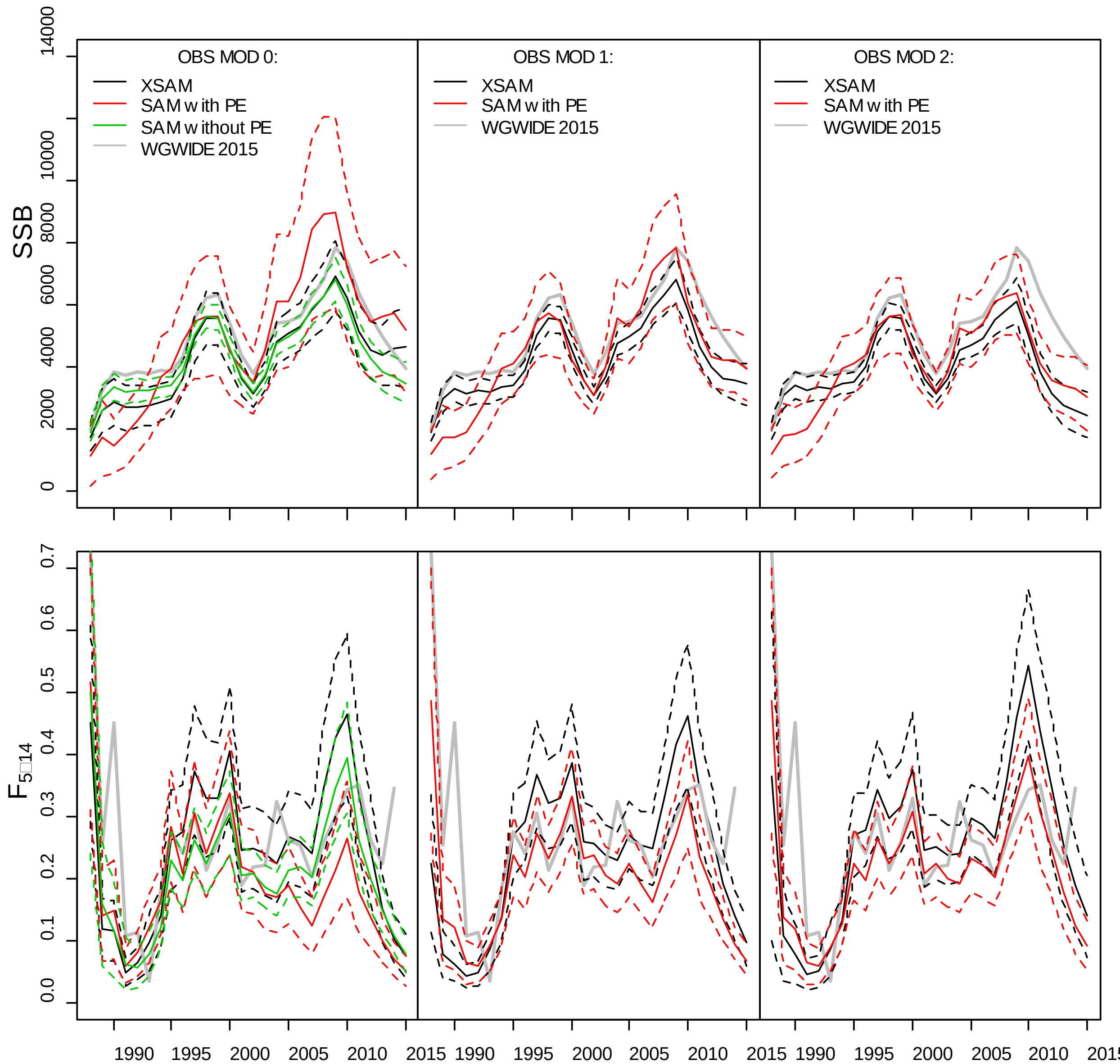


Figure 6. Estimates of spawning stock biomass (top row) and fishing mortality (bottom row) for 3 different formulations of the observation model: assuming iid errors and constant variance across ages and time for each data source (OBS MOD 0), using variance structure from estimates of the data sources (OBS MOD 1), and adding estimated correlation structure for the observations (OBS MOD 2) for the XSAM model (black), the SAM model with process error (red) and the SAM model without process error (green). The estimates from WGWIDE 2015 are included for comparison (gray line). The data used for fitting XSAM and SAM are restricted to catch data and fleet 5 (ages 3-15). For OBS MOD 1 and 2, the results from SAM without process error are nearly identical to XSAM and therefore not shown. Broken lines are approximate 95% confidence intervals.

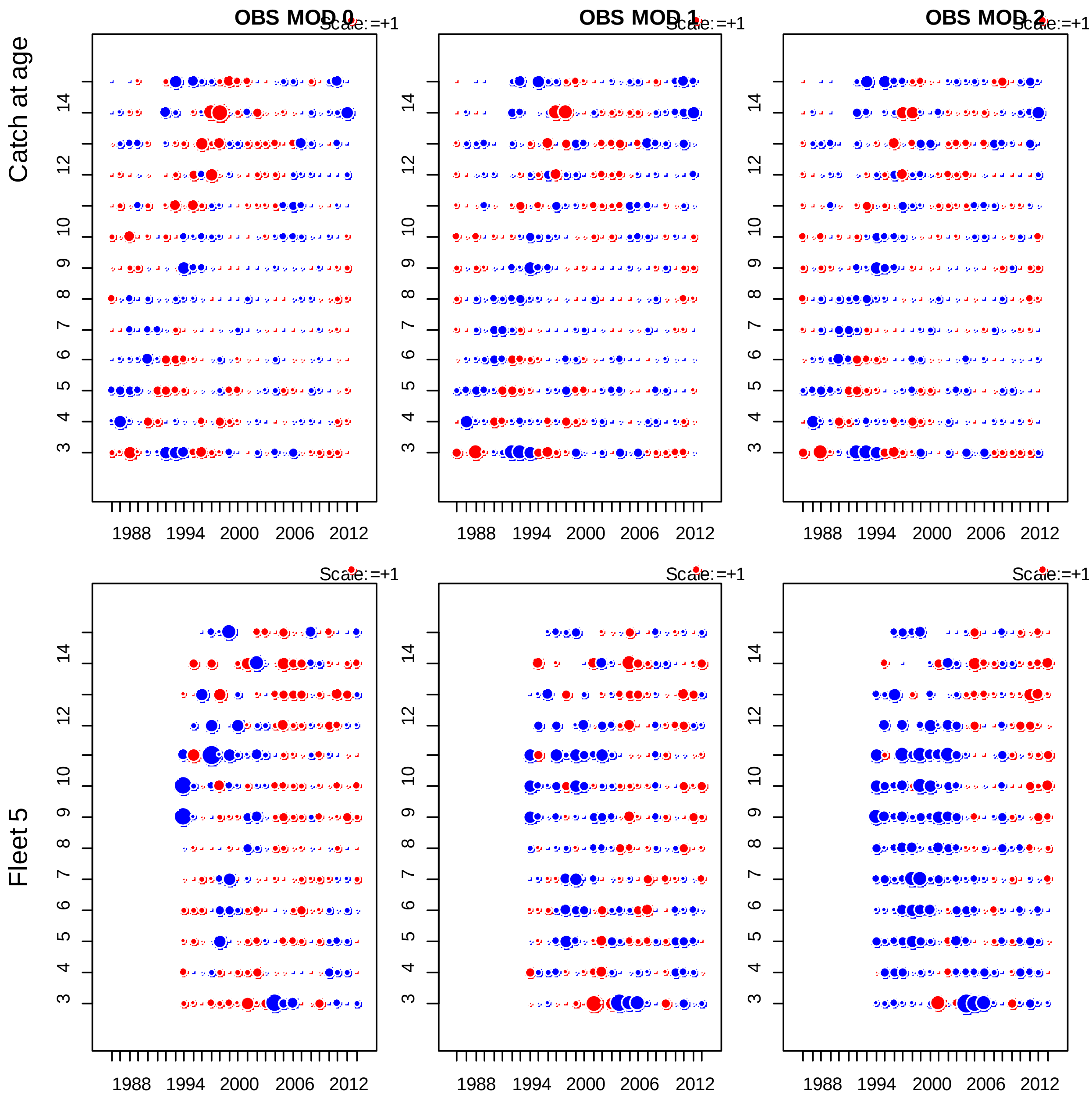


Figure 7. The residuals for catch at age and fleet 5 for the fitted model shown in Figure 6.

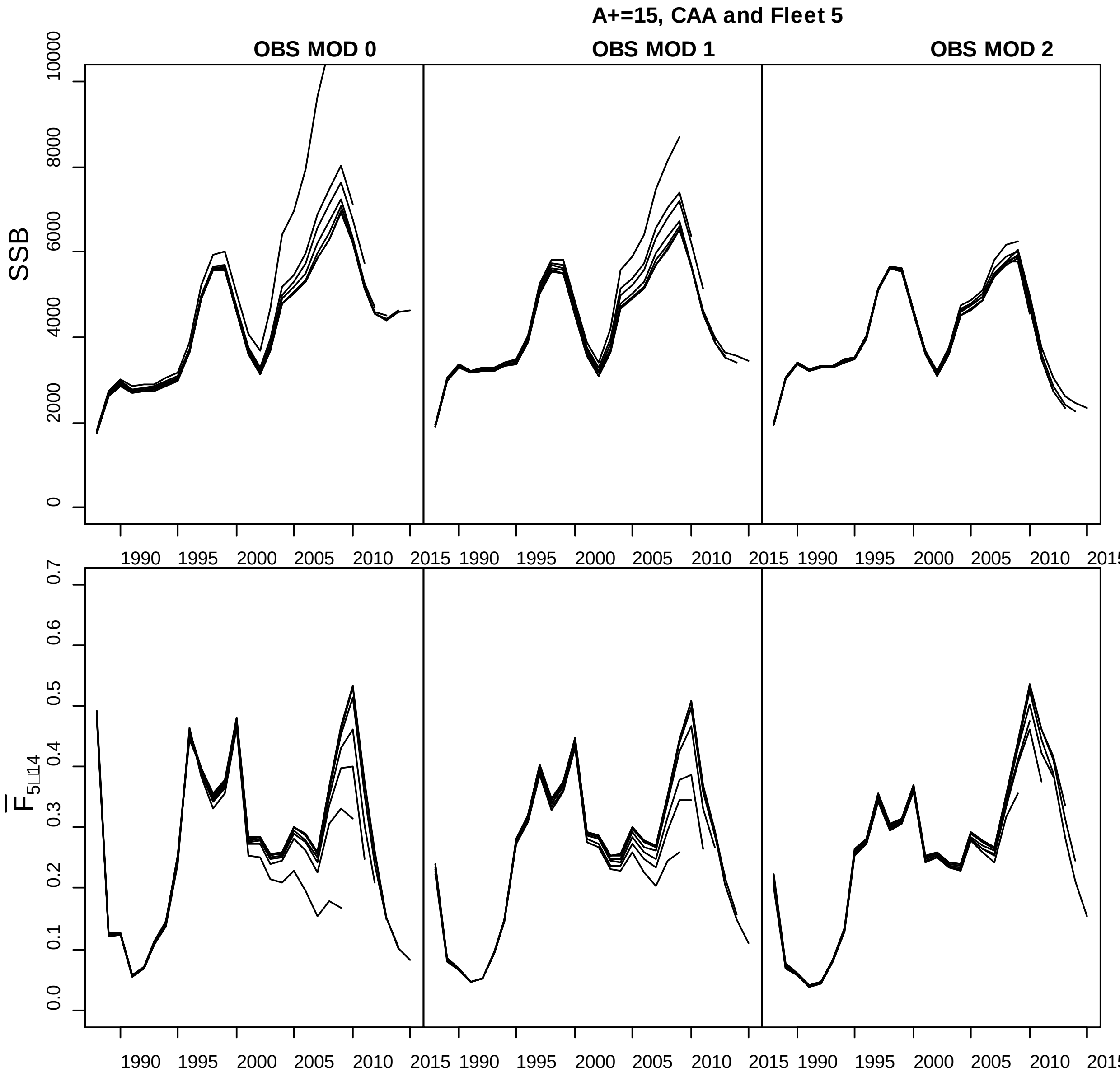


Figure 8. Retrospective results for SSB (top row) and average fishing mortality (bottom row) from XSAM including only catch at age and abundance indices from fleet 5. The retrospective run started in 2009.

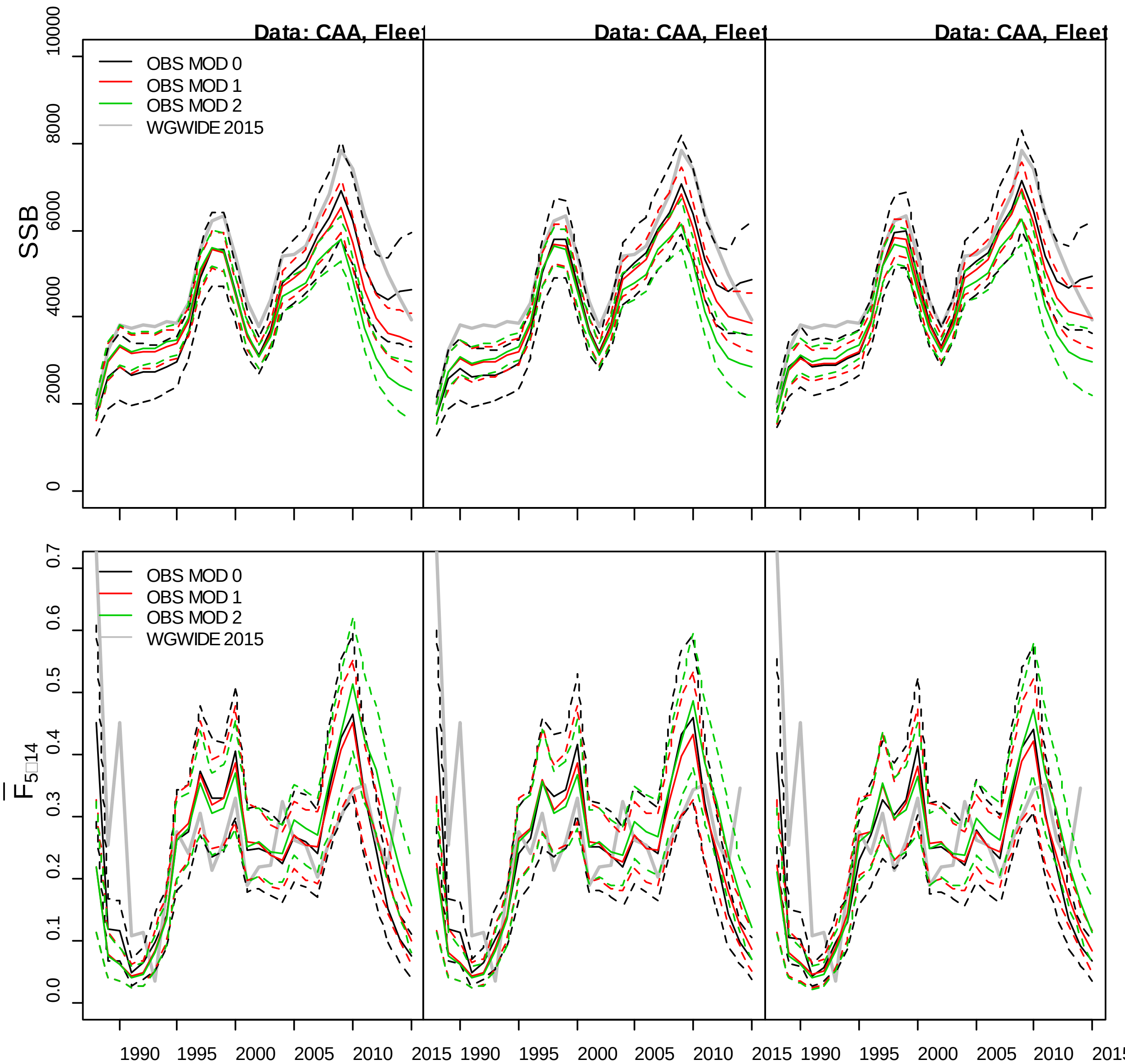


Figure 9. Estimates of SSB (top row) and F (bottom row) for different data inputs: the first column includes catch at age and data from Fleet 5, the second adds fleet 1 and the third adds fleet 2 and 3 to the estimates. The model is fitted with different observation models: iid errors with equal variance (black, “OBS MOD 0”), iid errors but variances according to empirical sampling variances (red, “OBS MOD 1”) and variances and covariances according to sample data (green, “OBS MOD 2”). The estimates from WGWIDE 2015 are included for comparison (gray line). Broken lines are approximate 95% confidence intervals.

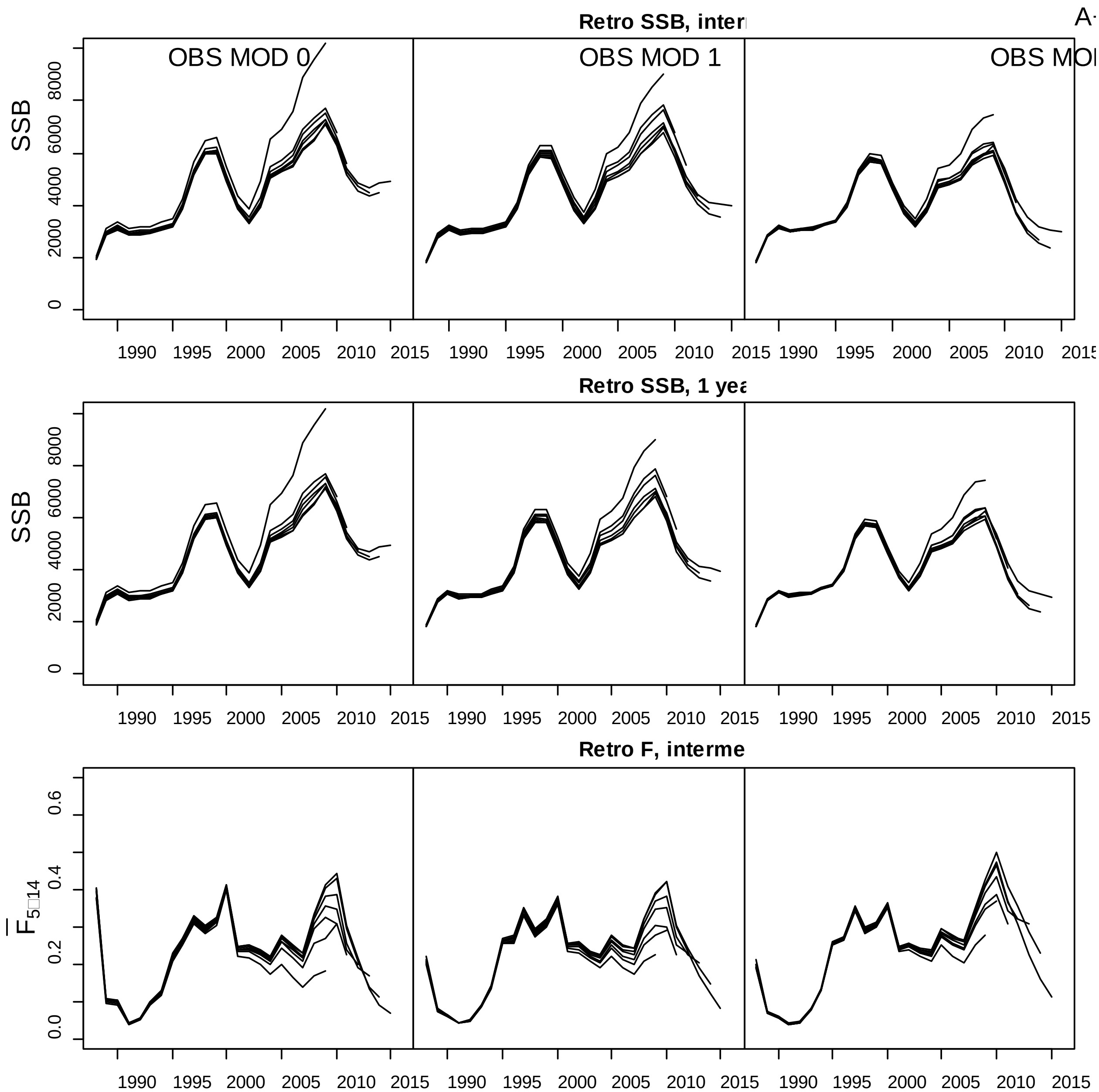


Figure 10. Retrospective runs when including catch at age, fleet1, fleet 2 fleet 3 and fleet 5 and A+=15.

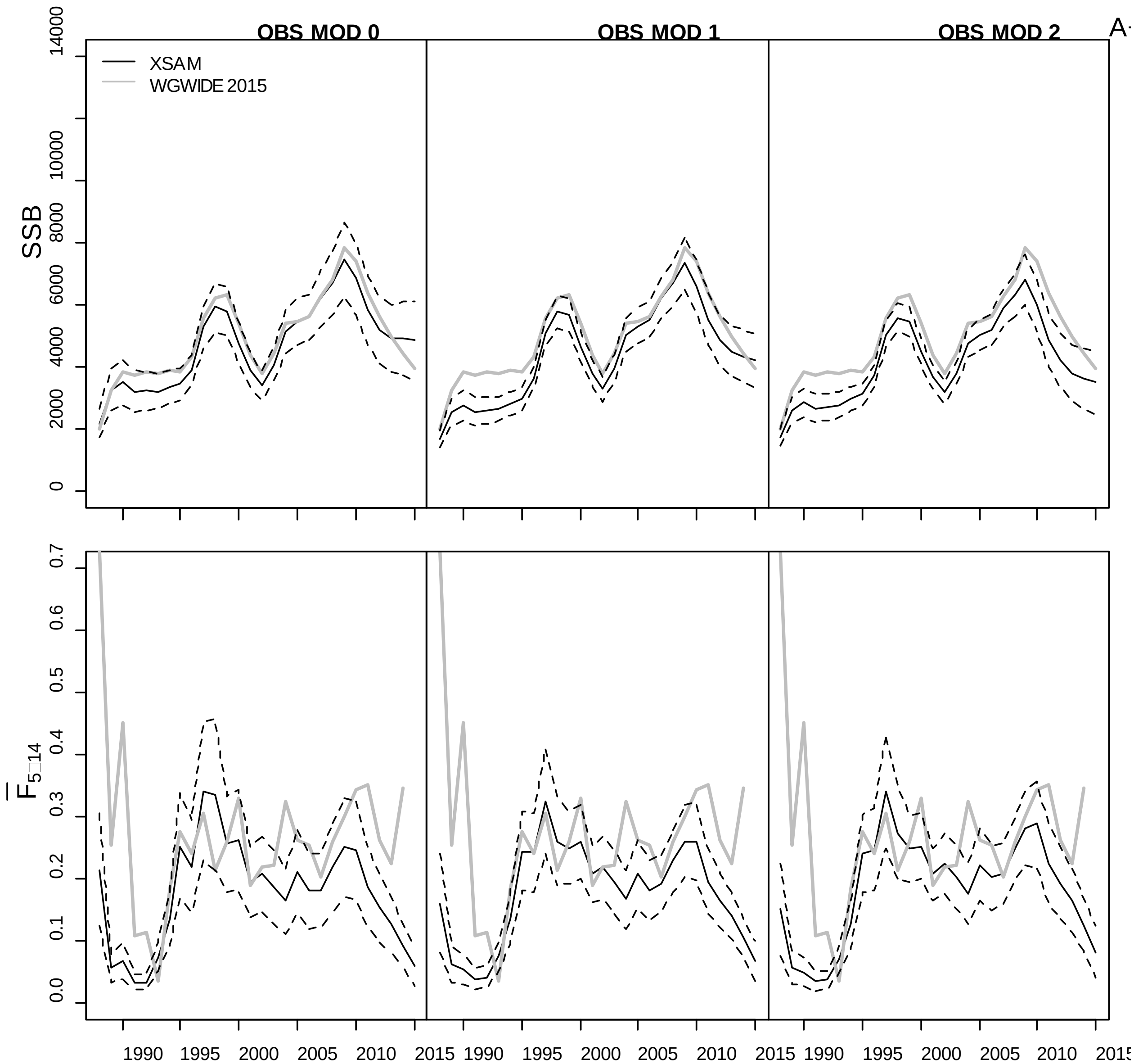


Figure 11. Fit of the model including catch at age, fleet1, fleet 2 fleet 3 and fleet 5 and A+=12.

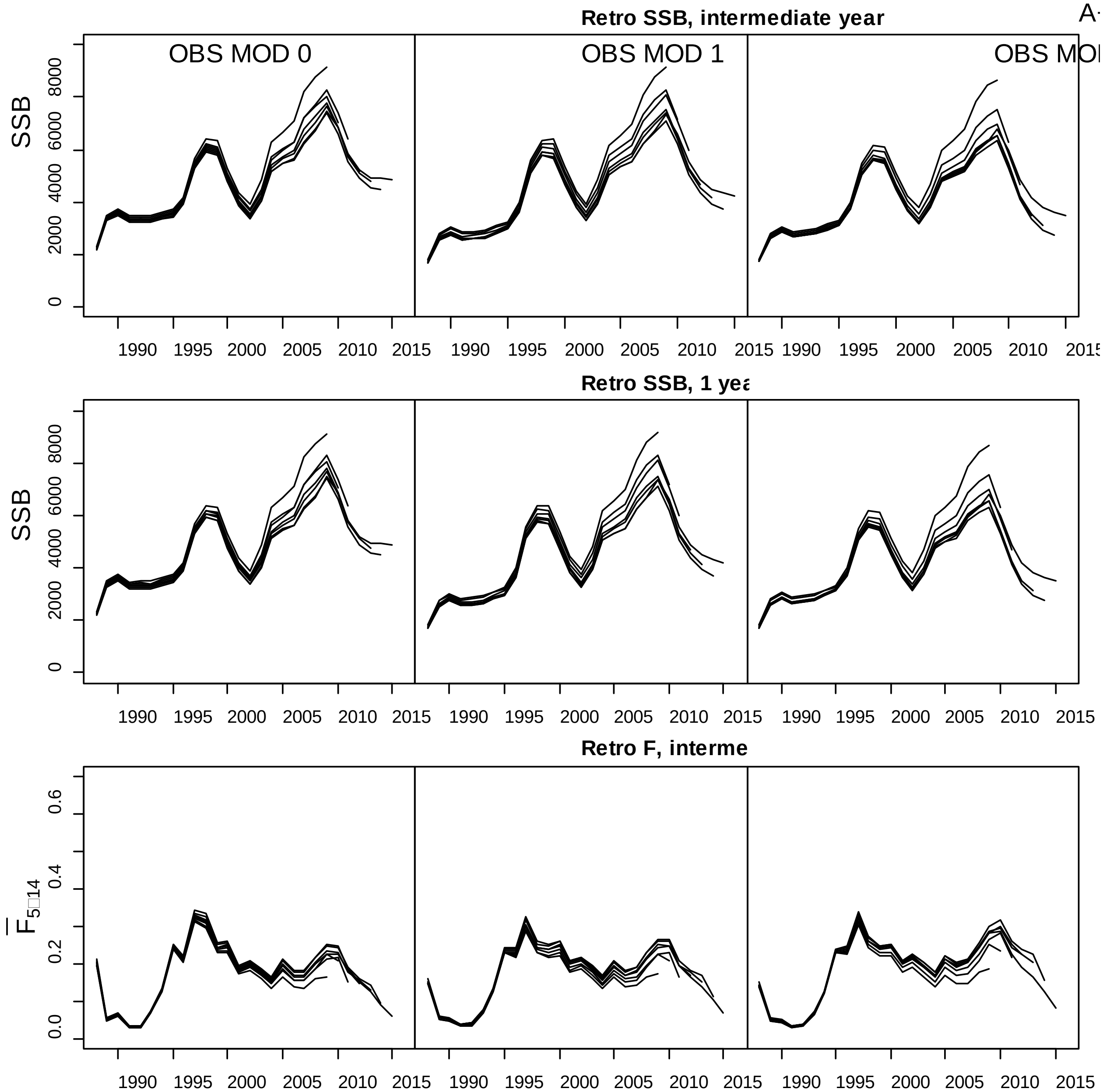


Figure 12. Retrospective runs when including catch at age, fleet1, fleet 2 fleet 3 and fleet 5 and A+=12.

Appendix A1

Prediciton of catch at age based on prediction of total catch weight

First we outline the distributional properties of the total catch weight by establishing the moments. Then we derive the approximate distribution of the mean total catch weight for the model prediction. Finally we link this with the prediction of the total catch weight available as external data.

Mean and variance of total catch weight

The total catch W_t is given by $W_t = \sum_{a=1}^{A} w_{a,t} C_{a,t}$, where $w_{a,t}$ is the mean catch weight at age. Now note that $\widehat{W}_t = \sum_{a=1}^{A} w_{a,t} \hat{C}_{a,t}$ with variance $Var(\widehat{W}_t) = \sum_{a=1}^{A} w_{a,t}^2 Var(\hat{C}_{a,t}) + \sum_{i \neq j} w_{i,t} w_{j,t} Cov(\hat{C}_{i,t}, \hat{C}_{j,t})$ provided $w_{a,t}$ is known. Note that this can only be 0 (total catch weight is known without error) if $\sum_{i \neq j} w_{i,t} w_{j,t} Cov(\hat{C}_{i,t}, \hat{C}_{j,t}) = -\sum_{a=1}^{A} w_{a,t}^2 Var(\hat{C}_{a,t})$, i.e. must be sufficient negative correlation among the catch numbers. In reality the $w_{a,t}$'s are also random variables which are correlated with itself and the $\hat{C}_{a,t}$'s which complicates the exact calculation of variance and is ignored here.

Approximate distribution of model predicted total catch weight

Then we derive the approximate distribution of the prediction of total catch weight. Since the mortality $F_{a,t}$ is modelled as a time series model a prediction of future fishing mortalities $F_{a,T}$ based on $F_{a,T-1}$ is defined and since $N_{a,t+1} = N_{a,t} e^{F_{a,t}+M_{a,t}}$ a prediction of the catch at age is available using the catch equation. This means that the distribution of predicted observations of catch at age, $\hat{\mathbf{C}}_{t+1}$, is available from the observation model

$$log(\hat{\mathbf{C}}_{t+1}) \sim MVN(log(\mathbf{C}_{t+1}), \mathbf{\Sigma}_{t+1}^{c})$$

conditional on the predicted parameters ($F_{a,t+1} = \hat{F}_{a,t+1}$, $M_{a,t+1} = \widehat{M}_{a,t+1}$ and $N_{a,t+1} = \widehat{N}_{a,t+1}$) with variance of the predicted observation $\mathbf{\Sigma}_{t+1}^{c}$ which correspond to the sampling error in $t+1$.

Provided $w_{a,t+1}$ are known we have that sum of the total of the observations

$$\widehat{W}_{t+1} = \sum_{a=1}^{A} w_{a,t+1} \hat{C}_{a,t+1}$$

which has mean

$$E(\widehat{W}_{t+1}) = W_{t+1} = E\left(\sum_{a=1}^{A} w_{a,t+1} \hat{C}_{t+1}\right) = w_{a,t+1} \sum_{a=1}^{A} E(\hat{C}_{a,t+1}) = \sum_{a=1}^{A} w_{a,t+1} C_{a,t+1}$$

provided the estimator for $\hat{C}_{a,t+1}$ is unbiased, and variance

$$Var(\widehat{W}_{t+1}) = \sum_{a=1}^{A} w_{a,t}^2 Var(\hat{C}_{a,t+1}) + \sum_{i \neq j} w_{i,t+1} w_{j,t+1} Cov(\hat{C}_{i,t+1}, \hat{C}_{j,t+1})$$

The total catch weight $\widehat{W}_{t+1}$ is a sum of log normal distributed variables, which is a distribution with no analytical solution. However, a reasonable approximation of the sum of log-normals is that the sum itself is lognormal with moments matching the mean and the variance (Fenton, 1960). Note that $Var(\hat{C}_{a,t+1})$ and $Cov(\hat{C}_{i,t+1}, \hat{C}_{j,t+1})$ can be found by established conversions from normal to log-normal distributions. In this way we assume that

$log(\widehat{W}_{t+1}) \sim N(\mu_y^W, \sigma_{Wt}^2)$, where $\mu_t^W = log\left(E(\widehat{W}_{t+1})\right) - 0.5\sigma_{Wt}^2$ and $\sigma_{Wt}^2 = log\left(\frac{Var(\widehat{W}_{t+1})}{[E(\widehat{W}_{t+1})]^2} + 1\right)$.

Note that the reported total catches correspond to $\{\widehat{W}_t\}_{t=1,\ldots,T-1}$, and the models ability to estimate the actual total catch is represented $Var(\widehat{W}_t)$.

Approximate distribution of the predicted total catch weight

For all years where total catch weight is available we do have data on $\widehat{W}_t$ (reported catches). For the prediction we do not have a direct observation of $\widehat{W}_{t+1}$, but instead we have a prediction of $\widehat{W}_{t+1}$, say $\widehat{\widehat{W}}_{t+1}$. We use the reported total annual catches $\{\widehat{W}_t\}_{t=1,\ldots,T}$ and WGWIDE prediction of total annual catches in the assessment year $\{\widehat{\widehat{W}}_t\}_{t=1,\ldots,T}$ to construct a model for the prediction of catches. We find that the model $\left(\widehat{\widehat{W}}_y\right) = log(\widehat{W}_y) + \varepsilon_y^p$, where $\varepsilon_y^p \sim N(0, \sigma_p^2)$ fits the data well (Table 1, $\sigma_p^2 = 0.003$ with $R^2 > 0.999$).

And implies the observation model for the total catch

$$log\left(\widehat{\widehat{W}}_t\right) = log(\widehat{W}_t) + \varepsilon_t^P$$

But from the observation model for catch at age we have

$$log(\widehat{W}_t) = \log(W_t) + \varepsilon_t^W$$

Such that

$$log\left(\widehat{\widehat{W}}_t\right) = \log(W_t) + \varepsilon_t^W + \varepsilon_t^P$$

Assuming ε_t^W and ε_t^P is independent we obtain

$$E\left(log\left(\widehat{\widehat{W}}\right)\right) = \log(W)$$

With variance

$$Var\left(log\left(\widehat{\widehat{W}}\right)\right) = \sigma_{Wt}^2 + \sigma_p^2$$

And we see how the prediction error increases the variance in the observation model for the total catch. In this way we can add a component to the likelihood for the predicted catch for the assessment year T where only the predicted total catch is available and not the catch at age in numbers.

In reality the mean catch at age weight $w_{a,t+1}$ and the sampling error $\boldsymbol{\Sigma}_{t+1}^{c}$ is not known in advance. The data do however suggest some stability in weight at age and assuming $w_{a,t+1} \approx w_{a,t}$ appears as a reasonable assumption. Similarly, $\boldsymbol{\Sigma}_{t+1}^{c}$ depend on the sample survey programs effort and designs, but provided the same design and approximately the same sampling effort we assume $\boldsymbol{\Sigma}_{t+1}^{c} \approx \boldsymbol{\Sigma}_{t}^{c}$, i.e. the sampling variability in the current year as about the same as previous year.

Note: this derivation is based conditional on the predicted values of mortalities and abundance (The abundance is informed by the survey for which we do have data in the assessment year). On one hand this means that we have not accounted appropriately for the variance in the predictions since we ignore the prediction variance. On the other hand the result simply expresses that the uncertainty in the models ability to estimate the total reported catch weight will be the "same" as for the other years and determined by the sampling effort (through Σ_{t+1}), but we will have extra variability corresponding to our ability to predict the reported total catch.

Appendix 2

Summary of analysis of sample data for catch at age and abundance at age for NSS herring.

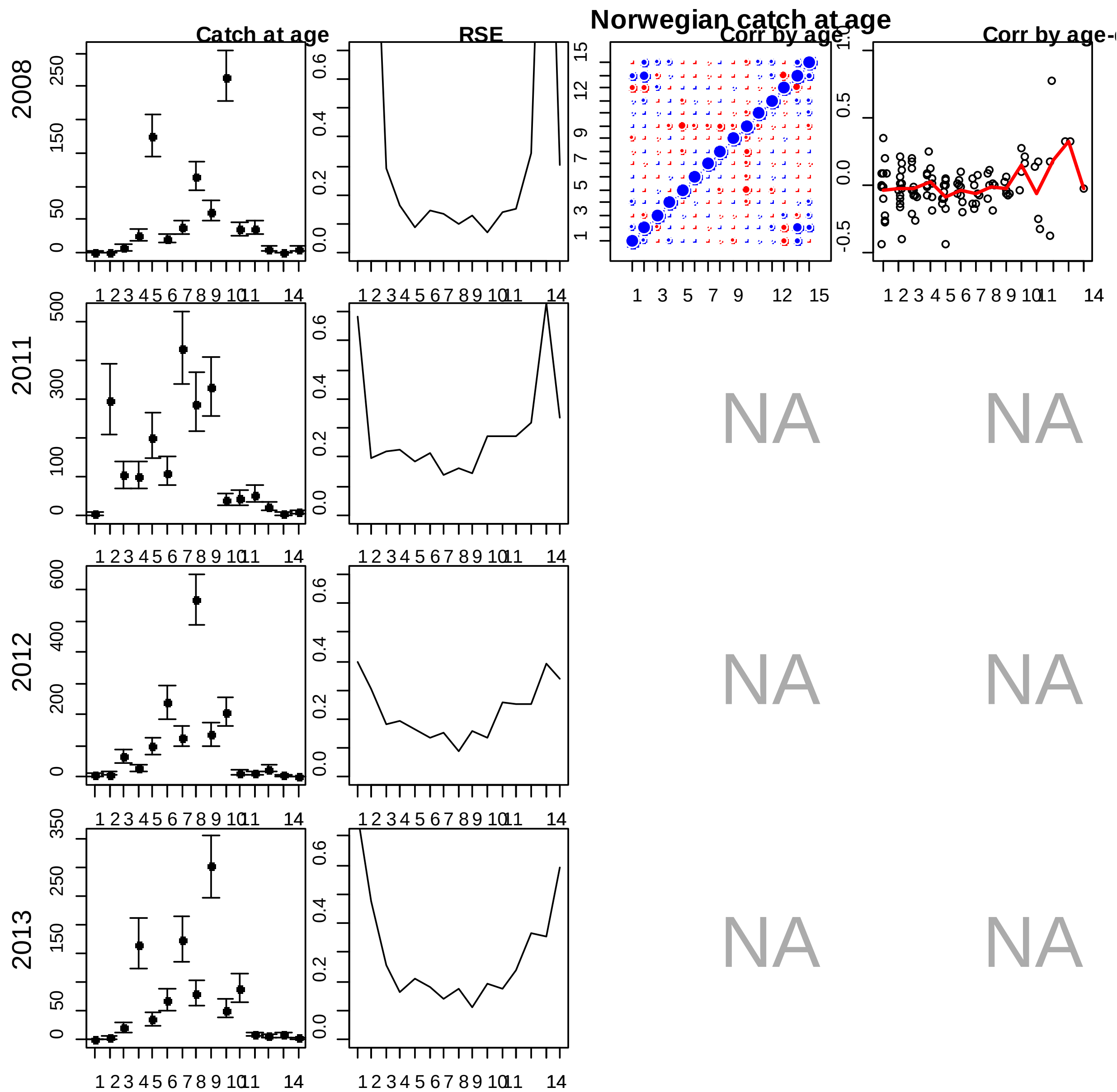


Figure A2.1. Summary of estimated Norwegian catch at age for NSS herring by years (rows) 2009-2015. Abundance at age (1. column) with 95% confidence intervals, relative standard error at age (2. column), correlation of abundance estimates by age (3. column), and correlation by distance in age (4. column). Alle estimates are based on ECA. Data for calculation of correlation structures of the estimates was not available for 2011, 2012 and 2013 at the time of these analysis.

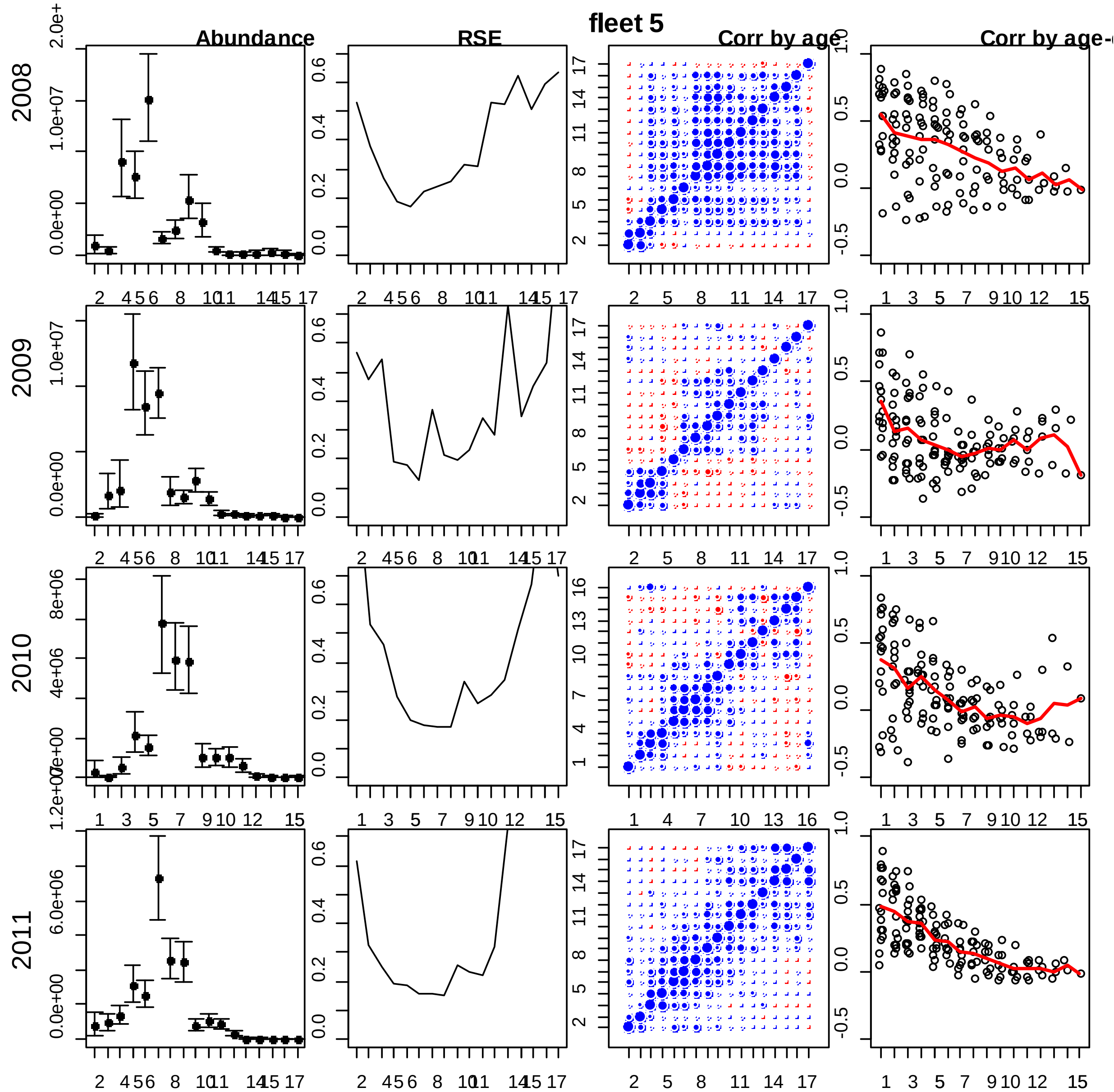

fleet 5
Abundance
RSE
Corr by age
Corr by age-
2008
2009
2010
2011

Figure A2.2.

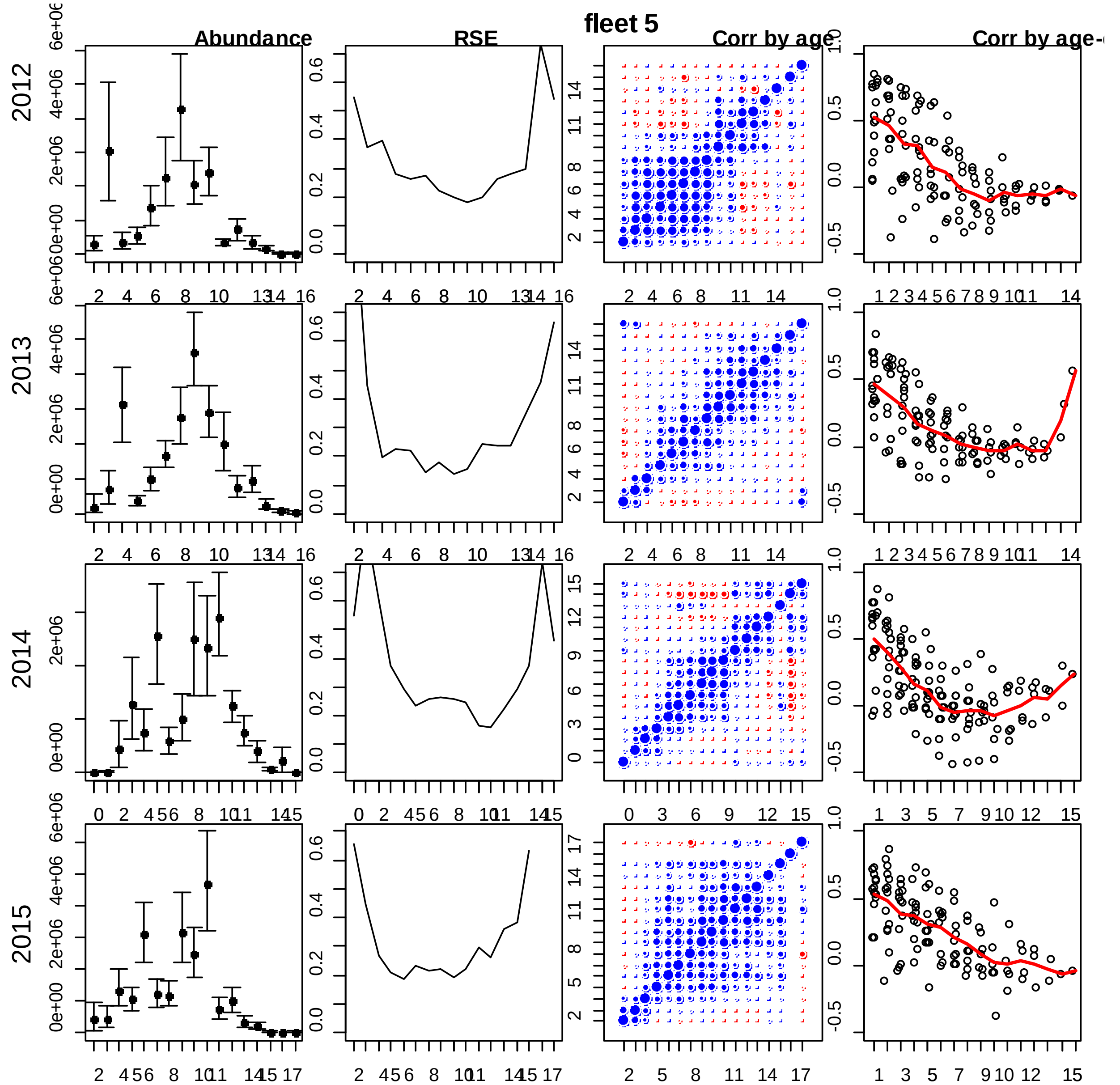


Figure A2.2 cont. Summary of estimated abundances at age for the May Survey (fleet 5) by years (rows) 2009-2015. Abundance at age (1. column) with 95% confidence intervals, relative standard error at age (2. column), correlation of abundance estimates by age (3. column), and correlation by distance in age (4. column). Alle estimates are based on StoX.

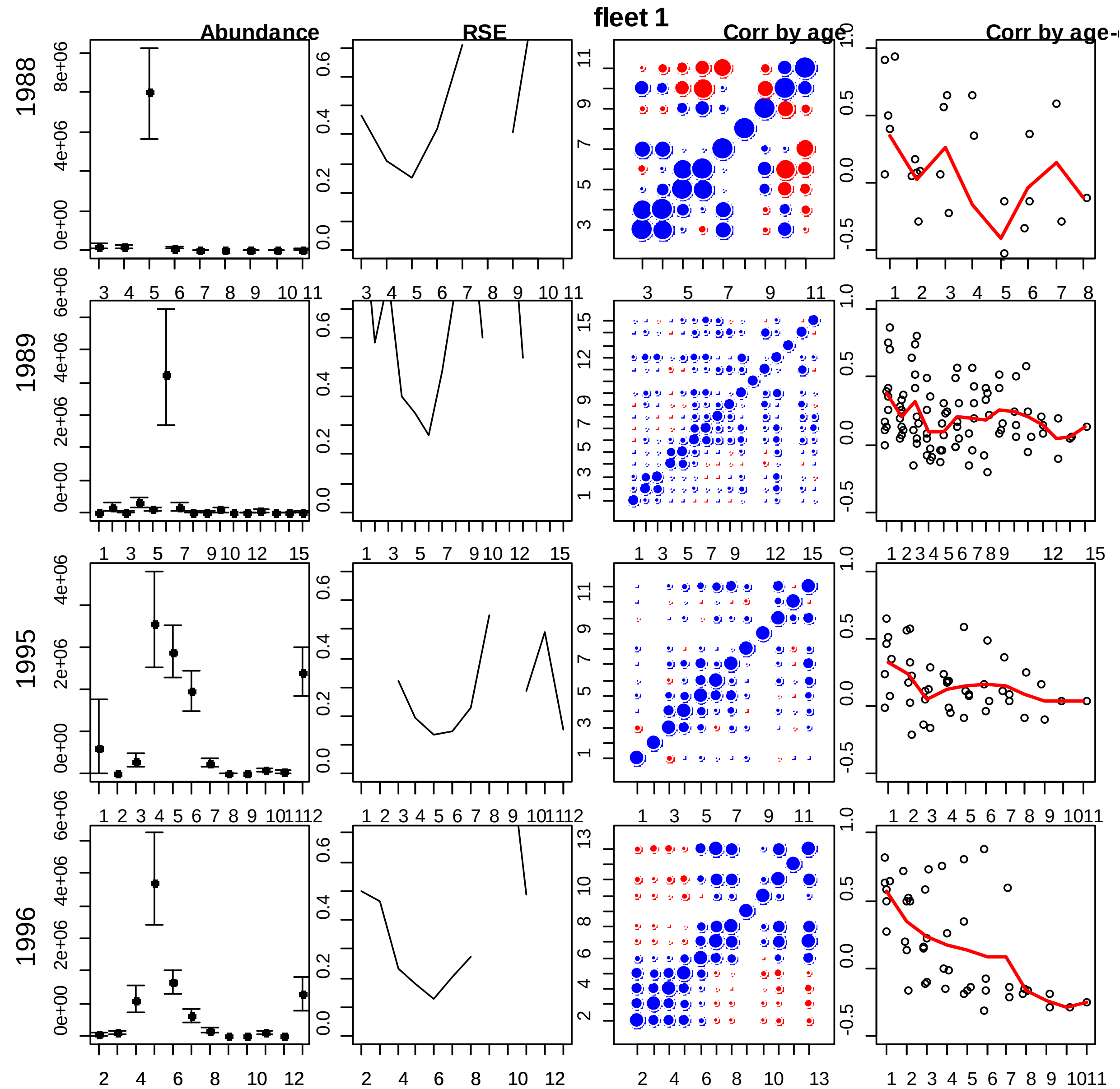

fleet 1
Abundance
RSE
Corr by age
Corr by age-
1988
1989
1995
1996

Figure A2. 3.

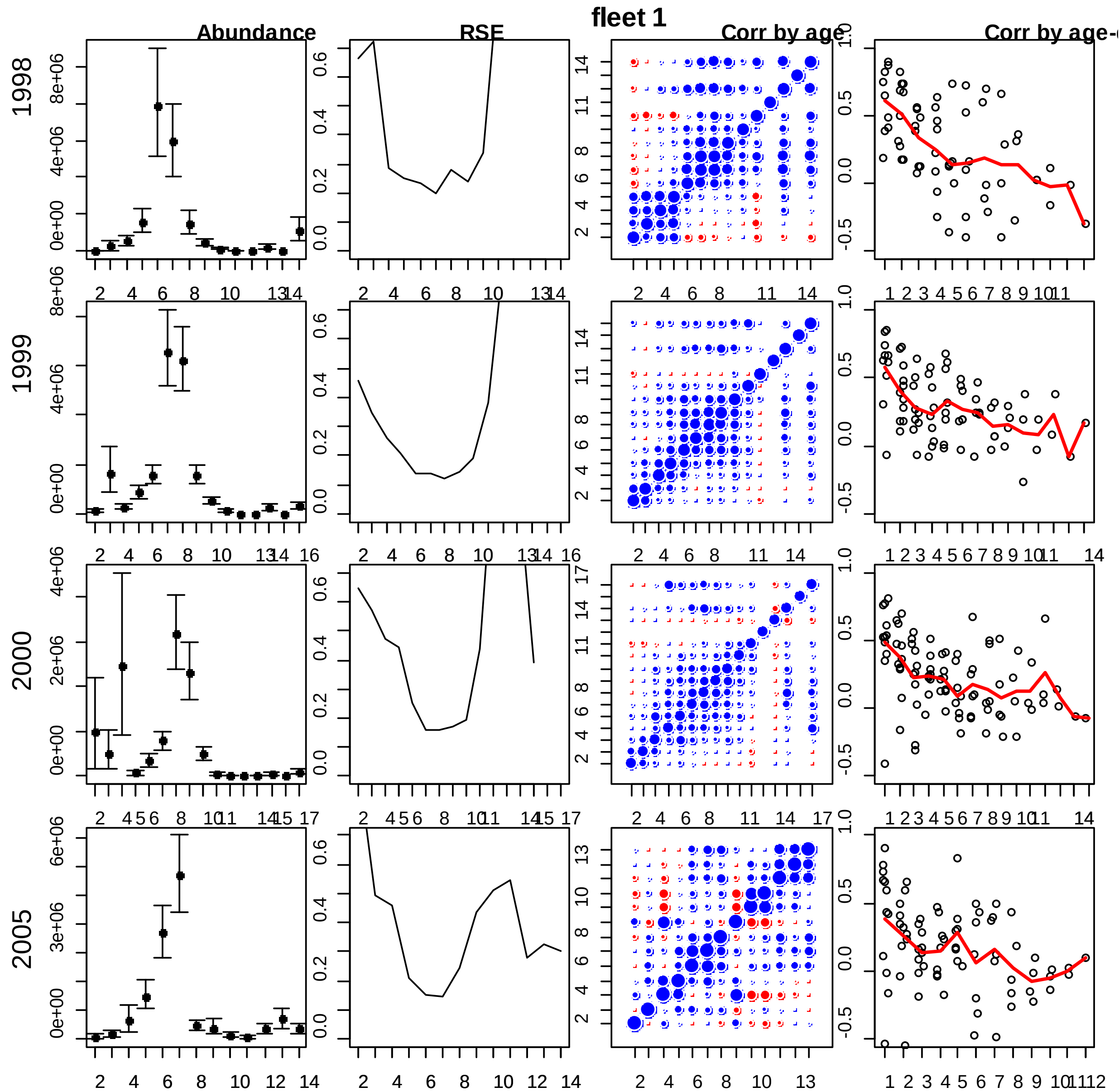


Figure A2.3 cont.

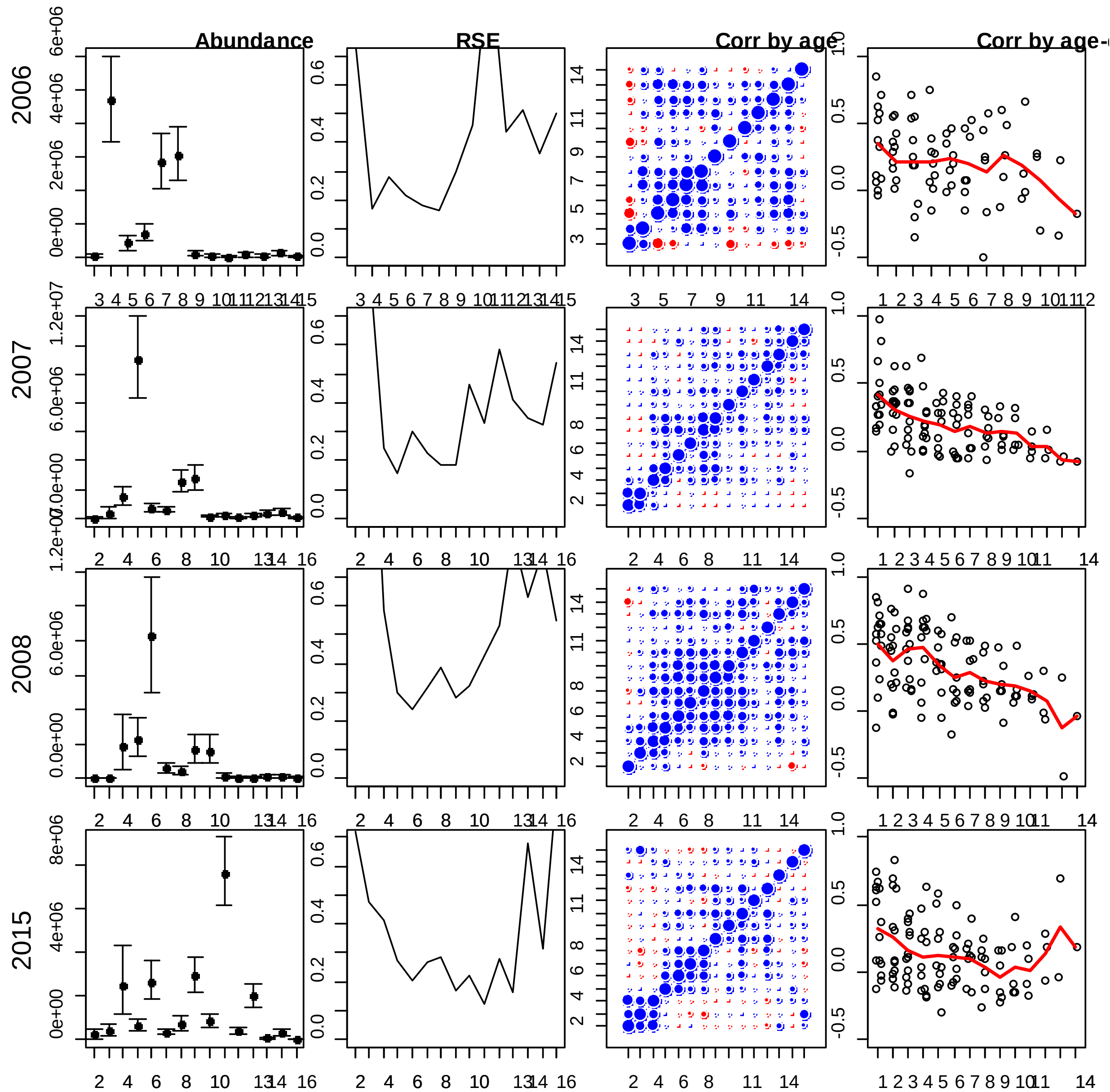


Figure A2.3 cont. Summary of estimated abundances at age for the spawning ground survey (fleet 1) by years (rows) 1988-2015. Abundance at age (1. column) with 95% confidence intervals, relative standard error at age (2. column), correlation of abundance estimates by age (3. column), and correlation by distance in age (4. column). Alle estimates are based on StoX.

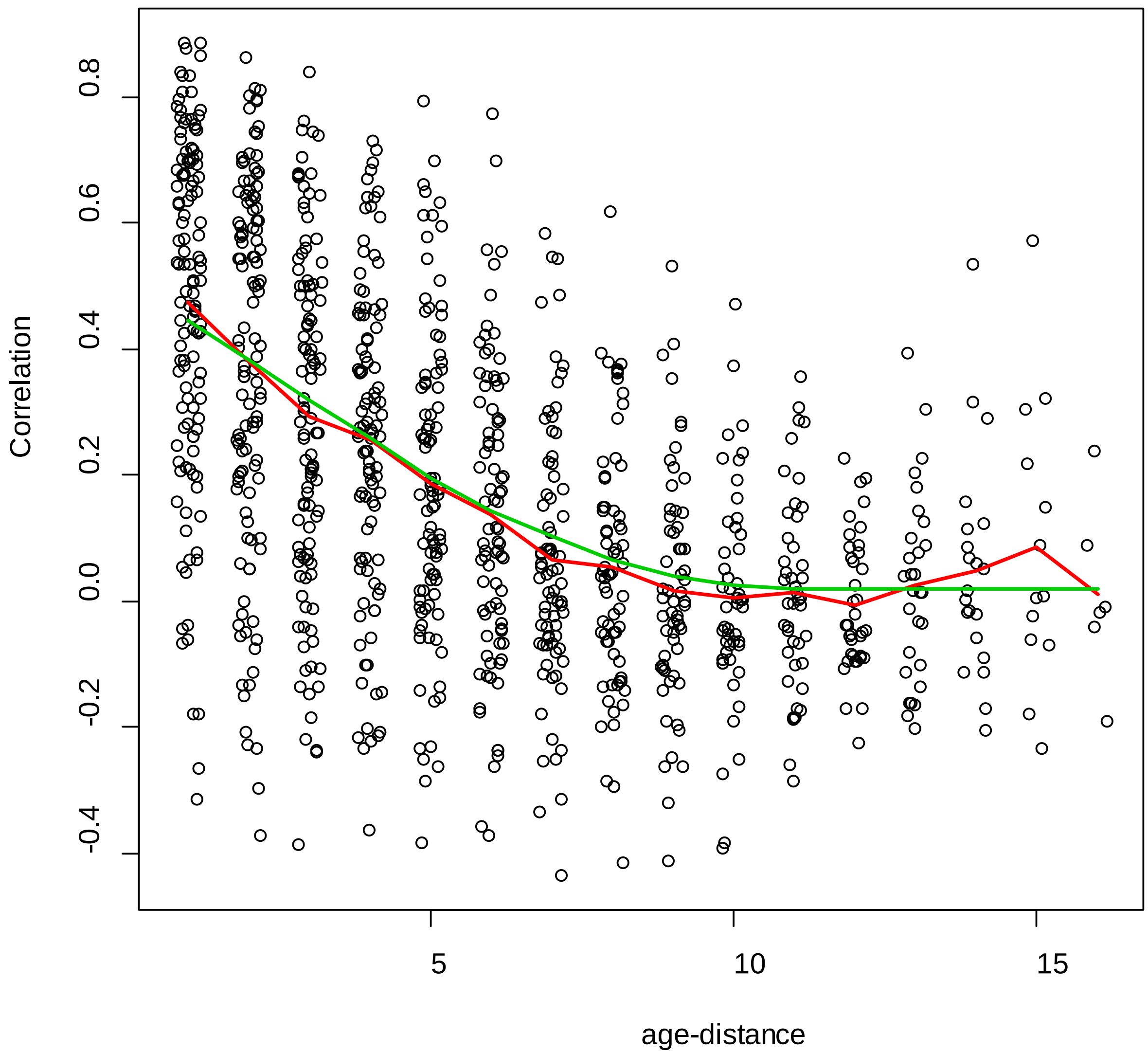


Figure A2.4. Average correlation structure over the years 2008-2015 by age distance of estimated abundance indices for NSS herring in the May survey (fleet 5).

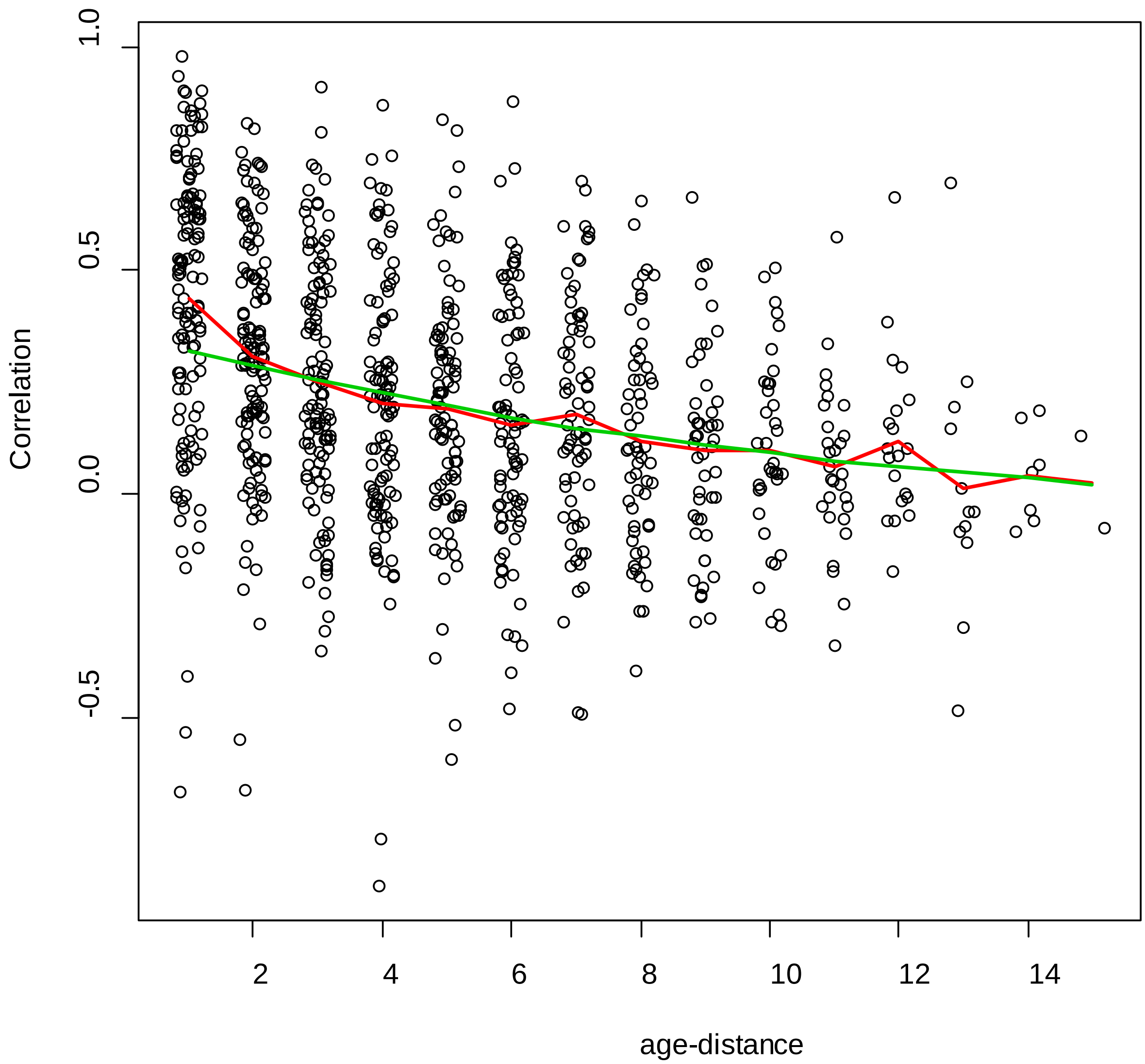


Figure A2.5. Average correlation structure over the years 1988-2015 by age distance of estimated abundance indices for NSS herring in the spawning ground survey (fleet 1).

Appendix A3: Variance structure of observations

The observation variances are available as estimates for the Norwegian catch at age and the survey indices for parts of the time series used for assessing herring. This enables using prior knowledge about these quantities. Estimation of observation error requires raw data (data to the lowest sample level or at least to the primary sampling unit), and unfortunately these have not been available for all years and all data, or have not been analyzed to achieve this. This hampers the evaluation of the data since their quality is a priori unknown. We therefore explore if it is possible to predict the variances of the observations using other and auxiliary information. To achieve this we consider Taylors spatial power law (Taylor 1961). This law postulates that for a population count Y with mean μ the variance of the population count is

$$Var(Y) = \alpha\mu^{\beta}$$

This relationship has proven to hold for a variety of survey data from a number of different populations across taxa (e.g. Taylor and Woiwood 1982) and have sound theoretical basis (e.g. Engen et al. 2008). Note that the formulation implies that for a simple random sample of n, the mean is a unbiased estimator of μ with variance

$$Var(\hat{\mu}) = Var(\bar{Y}) = \frac{\alpha'}{n}\mu^{\beta}$$

which relates the variance-mean relationship to sampling. A simple random sample of population counts is general is impossible for fisheries data such that n refers to the effective sample size $n = n_{eff}$. Empirical evidence implies that $n_{eff} \propto n_{PSU}$ for the survey in question, where n_{PSU} is the number of Primary Sampling Units (PSU) which for survey data typically is number of transects and for catch data is number of catches sampled. This suggest the variance model

$$v = Var(\hat{\mu}) = Var(\bar{Y}) = \frac{\alpha}{n_{PSU}}\mu^{\beta}$$

for population counts, and indeed we find that the model

$$\hat{v}_{a,y} = \frac{\alpha}{n_{PSU,y}}\hat{\mu}_{a,y}{}^{\beta}e^{\varepsilon_{a,y}}$$

where $\varepsilon_{a,y} \sim N(0, \sigma_{\mu}^2)$ fitted to empirical data for $\hat{v}_{a,y}$ and $\hat{\mu}_{a,y}$ for each survey (including the catch data) separately, different surveys rather well. The inclusion of n_{PSU} in the intercept significantly improves the model fits which supports the assumption that $n_{eff} \propto n_{PSU}$. We therefore utilize this relationship to predict the variances for years were only the point estimates are available. A critical assumption for a constant relationship between $n_{eff} \propto n_{PSU}$ over time is that the sampling design has been constant, e.g. same stratification and area coverage, such that variability in the effective sample size is reflected by the variability in the sampling effort in the survey. In cases were even n_{PSU} is unknown, we replace it with average numbers. Note that the parameters in the variance function is scale dependent if $\beta \neq 2$ which becomes relevant if the scale of the observations used to estimate the parameters in the variance function is different to the scale of the observations used when fitting the assessment model to the observations. To translate the meaning of this in terms of weighting in the likelihood, $\beta < 2$ implies that large values is given relatively larger weight than small values, $\beta = 2$ all data is given equal weight independent of size, whereas $\beta > 2$ implies large values is given relatively smaller weight than small values that a value of variance.

Data	Age range	Year range	$\hat{\beta}$	R^2
Catch at age	3-15	2011-2013	1.47 (0.03)	0.97
Fleet 5	3-15	2008-2015	1.47 (0.03)	0.94
Fleet 1	3-15	1988-2015*	1.40 (0.03)	0.93

*The survey have not been run every year and the range correspond to data available in WGWIDE (2015)

Scaling of the variance function

Scaling of the variance function may be necessary for two reasons: 1) Sometimes the time series of data used in the assessment models, $\hat{\mu}'$, are scaled by a constant k compared to the original scale values $\hat{\mu}$, i.e. $\hat{\mu}' = k\hat{\mu}$, necessary to obtain numerical stability when optimizing the likelihood function. 2) In other cases it we may be in the situation that we have knowledge about the relative standard error, but no direct information about the scaling; for example: Based on rigorous analysis of the sample data from a portion of commercial catches, say from one country, we do have good knowledge about the sample distribution, but not for the total. If the sample design and effort is similar for the portion of the catches which lacks the rigorous analysis, it may be good reason to believe that variance function and thus the relative standard error is the same, but the scaling factor is complicated to derive. In this case the we can find the scaling factor based on RSE. In any case this implies that the parameters in the variance relationship must be scaled correctly if it differs from the scale of the data used for inference about the parameters in the function.

In principle k is easily derived directly by matching pairs of the scaled an unscaled observations ($k = \hat{\mu}/\hat{\mu}'$) and then realize that

$$Var(\hat{\mu}') = Var(k\hat{\mu}) = k^2 Var(\hat{\mu}) = k^2 \frac{\alpha}{n_{PSU}} \hat{\mu}^{\beta} = k^2 \frac{\alpha}{n_{PSU}} \left(\frac{\hat{\mu}'}{k}\right)^{\beta} = k^{2-\beta} \frac{\alpha}{n_{PSU}} \hat{\mu}'^{\beta}$$

Alternatively, we could utilize the relative standard error to determine k since

$$RSE(\hat{\mu}) = \frac{\sqrt{Var(\hat{\mu})}}{\hat{\mu}} = \frac{\sqrt{Var(\hat{\mu}')}}{\hat{\mu}'} = RSE(\hat{\mu}') = RSE$$

i.e. it is independent of k, and since

$$Var(\hat{\mu}') = RSE^2 \hat{\mu}'$$

We identify k by solving the above equation substituting for $Var(\hat{\mu}')$ which yields

$$k = \left(\frac{n_{PSU}}{\alpha} RSE^2\right)^{\frac{1}{2-\beta}} \hat{\mu}', \text{ for } \beta \neq 2$$

For $\beta = 2$, the scaling does not effect the variance since it implies that the variance is proportional to the square of the mean independent of the scaling. In practice this means that the constant k is determined by selecting the model predicted RSE for $\hat{\mu}$ with the corresponding value of $\hat{\mu}'$.

In practice, small deviations in methods, including revisions of data over time, or even just rounding errors, causes some random deviations between the two sets. This implies that the actual observation to be used is a random variable conditioned on the observation used for determining the variance function, e.g. $\hat{\mu}' = k\hat{\mu} + \varepsilon$, and consequently we find small deviations in k for different pairs of RSE and $\hat{\mu}'$. This implies that k should be averaged over the matching data sets. This alternative may seem overly complicated and a more transparent way of determining the scaling factor is therefore to use the matching datapoints and estimate k directly based on $\hat{\mu}' = k\hat{\mu} + \varepsilon$, and then apply

$$Var(\hat{\mu}') = k^{2-\beta} \frac{\alpha}{n_{PSU}} \hat{\mu}'^{\beta}$$

to predict variances directly. Note that this implies that $\hat{k} = \bar{k}$, where $k_i = \hat{\mu}_i / \hat{\mu}_i'$, i.e. simply the average of all ratios.

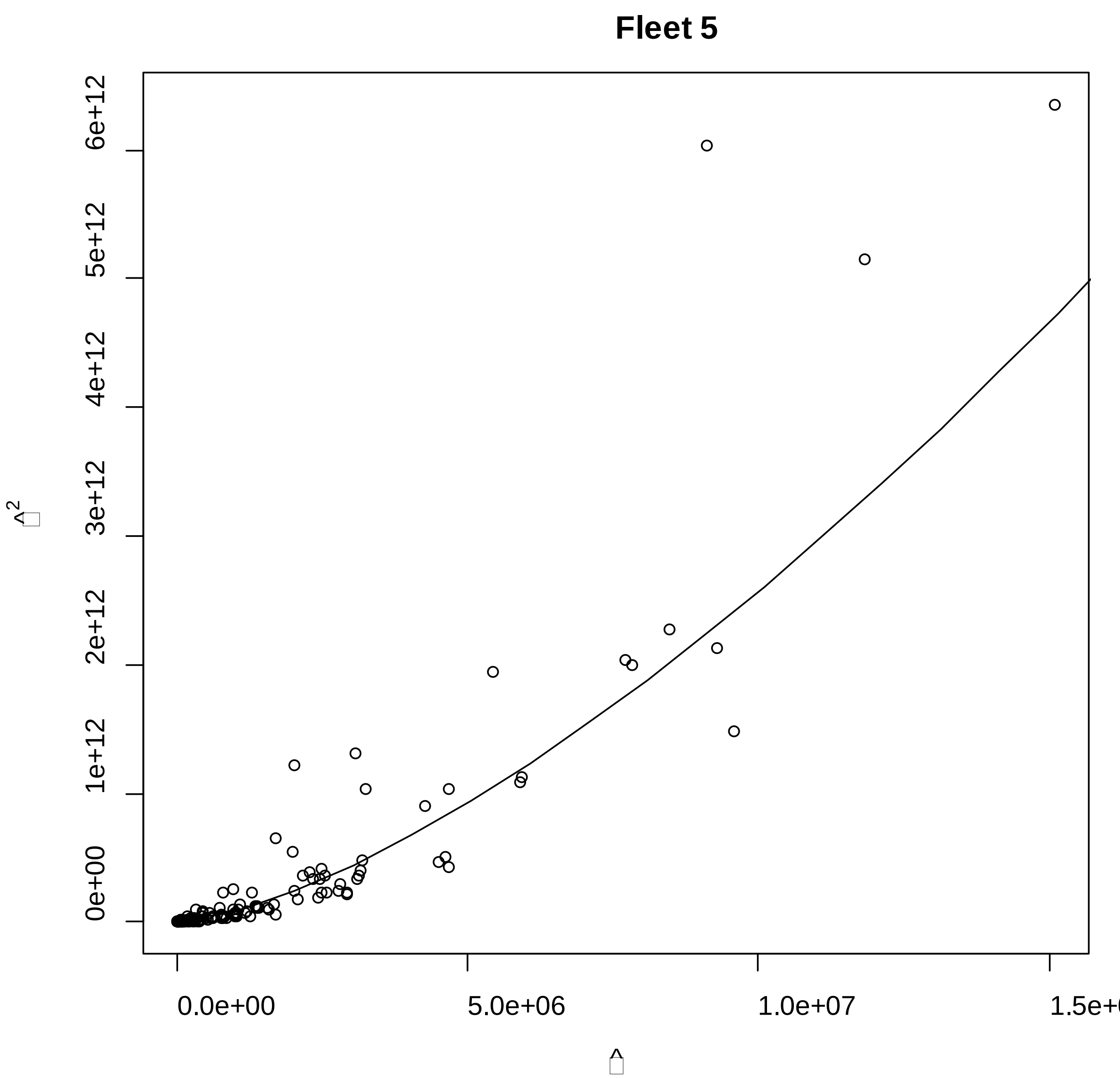


Figure A3.1. The Taylor variance function fitted to data (2008-2015) for the May survey (fleet 5) on sampling variances and estimates of abundance. On the logarithmic scale the R-squared is 0.94.

References

Engen, S., Lande, R. and Sæther, B.-E. 2008. A GENERAL MODEL FOR ANALYZING TAYLOR'S SPATIAL SCALING LAWS. Ecology, 89: 2612–2622. doi:10.1890/07-1529.1

Taylor, L. R. 1961. Aggregation, variance and the mean. Nature 189:732–735.

Taylor LR & Woiwod (1982) Comparative Synoptic Dynamics. I. Relationships Between Inter- and Intra-Specific Spatial and Temporal Variance/Mean Population Parameters. J Animal Ecol 51, 879–906.

Appendix 4
Properties of the model for herring type data

NSS herring is a population with big variability in recruitment forming strong and weak year-classes, this result in two issues that needs to be addressed: First; large year-classes can be traced for a long period of time both in the catches and the surveys indicating that the plus-group should be rather high (can be traced to the age of 16-18). The weaker ones disappears relatively quickly such that no individuals of ages say older than 12 is observed in catches or surveys. This result in the second issue which is that estimates of catch at age and survey indices becomes 0 (concerns ~2% of the catch at age data and 6% of the abundance indices at age) for small yea-classes at old ages. This indicates that the plus-group should be set much lower than for strong year-classes. Values of 0 do obviously contain information that the abundance is low, and the problem of the value 0 is due to modelling the data as log normally distributed variables which implies a log transform of the data (see *Data*). This would not represent any problem if another distributional family (which is defined for the vale 0) for observations had been chosen (e.g. Poisson).

To be able to address 1) the models ability to estimate vital parameters such as SSB, average F and observation errors, 2) how to deal with 0's in data and 3) how to model the plus-group I use simulations. More specifically this means that the population is simulated by the model for the process, a corresponding set of observations are generated by the observation models which then is used to fit the model, and the estimates can be compared with the "true" values. Repeating this procedure a large number of times enables an evaluation of the obtained estimates through statistics such as relative error, relative standard error (measures variability) and coverage (whether the error is in line with the variability estimates). To parameterize the process and observation model, a simplified version of the model is fitted to catch at age data (1988-2014) and abundance indices estimated from the May survey (1996-2015) as reported in WGWIDE (2015). The analysis was restricted to ages 3-14 omitting the data from the plus-group to not interfere parameter estimates with plus-group specific features. This enables evaluation of the model fit to the following situations:

- complete data without any 0's and plus group
- introducing 0's in data by setting the smallest values in data to 0 according to empirical proportions of 0s in the data and evaluating the effect of adding small numbers to the data compared to omitting the data containing 0's
- using the dynamical pool model for the plus-group compared to omitting data for the plus-group and instead model the abundance in the plus-group. This was done for a small (15+) and large (11+) plus-group.

The result from this evaluation provides guidance for later analysis.

Parametrization of simulation model

To parameterize population and observation models that can reproduce key features of dynamics in population and observations of NSS herring I start with a simplified version of the model fitted to real data:

I used the same time range as WGWIDE (2015), i.e. 1988-2015 including estimates of catch at age and abundance indices estimated from the May survey (available from 1996) as reported in WGWIDE (2015). I have restricted the analysis to ages 3-14 omitting the data from the plus-group to not interfere parameter estimates with plus-group specific features. Also all 0's in data was omitted from the likelihood. I set age specific selectivity constant over time such that the model for fishing mortality is the same as in Aanes et al (2007) where the fishing mortality is a separable model with random noise, but the effort modelled as an AR(1) process. The age specific selectivity were set constant for ages above 10, i.e. $a_m = 10$. For the observation models the simplest form was chosen: $\Sigma_{\rm t}^{\rm c} = \sigma_c^2\mathbf{I}$ and $\Sigma_{\rm t}^{I} = \sigma_I^2\mathbf{I}$**,** i.e. one parameter across time and age for catch at age and survey indices respectively.

When simulating the model when the aim is explore the model for plus-group, some care must be taken. Conceptually the plus-group contains all fish above a certain age, which in theory infinite. If the maximum age in the simulation model is set too low, i.e. when there still are some individuals left, it will create to small aggregate catches in the plus-group, such that the initial population size (and thus SSB) will be underestimated since all fish not is accounted for. Therefore, the age in the simulation model must be set so high that there is close to 0 fish left for the maximum age. To achieve this for the simulated population we set the maximum age to 100 but note that few individuals are left after age 18. Then we parameterize a simulation model up to age 100 by setting the selectivity constant for ages above $a_m = 10$, such that $\mu_{a,t}^F = \mu_{a_m,t}^F$ for $11 \leq a \leq 100$. Although the mean fishing mortality within year is constant for ages above a_m there is still variability across ages according to $\sigma_{a,t}^{(1)2} = \sigma^{(1)2}$ as this appears as more realistic than the mortality being equal for all ages. For simplicity, weight at age in stock and catch, as well as proportion mature at age, are set constant over time such that variability in stock numbers and SSB is only due to variability in recruitment and mortality.

Based on the parameter estimates realizations of population are simulated, and 28 years of data (equal to the time span 1988-2015) is extracted, and observations are generated by simulating from the observation models using $\{\hat{q}_a\}_{a=a_{min},\ldots,A}$, $\hat{\sigma}_c^2$ and $\hat{\sigma}_I^2$.

This fit results in very large values of $\hat{\sigma}_c^2$ and $\hat{\sigma}_I^2$ indicating RSEs ~30-50% for the data, i.e. very poor precision in data.

For each replicate the following is done:

1. Use complete observation data ages 3-14 with no plus-group and compare estimates with simulated data ages 3-14
2. Use data ages 3-14 with no plus-group, replace the 2% lowest values of catch at age with 0 and the lowest 6% of the abundance indices with 0 and fit the model by
    a. Omitting the 0's
    b. Add a small number to the 2% 0 catches, but omitting 0s in the survey abundance.
3. Use data ages 3-100 create and create a plus-group 15+ from which observations are generated
    a. Apply a dynamic pool model for 15+
    b. Fit model to data for ages up to 14 with no plus-group but run model to age 18 assuming constant mortalities within year for ages above 14.

4. Use data ages 3-100 and create a plus-group 11+ from which observations are generated
    a. Apply a dynamic pool model for 11+
    b. Fit model to data for ages up to 10 with no plus-group but run model to age 18 assuming constant mortalities within year for ages above 10.

Results

The simulation experiment based on the simplified model shows that for complete data (no 0's) and no plus-group, the estimates of SSB and average F are approximately unbiased and the coverage is close to the true level (Figure A4.1). A small trend in the error appear to underestimate SSB for the most recent part, monte carlo errors (due to too few replicates) makes it difficult to conclude but indicates that SSB may be biased downward with ~2%. When the data for catch at age and abundance indices contains 0's (2% of the observations for catch at age and 6% of the abundance indices), the effect of adding a small number to the catch at age while removing the 0-observations from the survey indices is to introduce a small bias in SSB and a huge bias in estimates of F as the precision lowers for both and the coverage remains approximately the same (Figure A4.1). The effect of adding a small number also to the abundance indices is detrimental and introduces large bias, low precision and very low coverage (not shown). The effect of removing the 0's instead of adding a small number is the bias in estimates of SSB is on the same level compared to the complete data, but changes sign (from slightly negative to slightly positive), the estimates of F remains approximately unbiased, and there is a slight decrease of precision as the coverage remains the same as for the complete data. For the estimates of the observation errors the observation error for catch at age is largely overestimated (by about 40%) and the observation error for abundance indices are also overestimated but to a lesser extent. The other parameters estimates appear reasonable. This is in line with the findings in Aanes et al. (2007) which used the same model for fishing mortality as evaluated here.

The conclusion from this exercise is that a small value should not be added to the data, but instead to remove 0 values, as the small values will drive the likelihood function and introduce large uncertainty and bias into the estimates when a log normal distribution is used for the observations. For the proportion of 0 values represented in these data, omitting the 0 values have a small effect in increasing the uncertainty of the estimates, but otherwise appear to be a satisfactory solution. This will therefore be done for the rest of the analysis shown in this document.

The effect of applying the dynamical pool model for a relatively small plus-group (15+) is to underestimate slightly underestimate SSB as average F (ages 5-14) appear unbiased (Figure A4.2). The precision is relatively high, and the coverage for SSB appears reasonable. Ignoring the data for the plus-group and let the model predict abundances up to age 18 leaves the bias unchanged, but lowers the precision and the coverage. This means that letting the model predict from ages 14 to 18 increases the uncertainty, but not sufficiently to ensure that the coverage is correct. The effect of having a larger plus-group (11+) and modelling it as a dynamical pool yields the same small bias for SSB and average F, but increases the variability with a coverage that that appear appropriate (perhaps slightly lowered, but monte carlo error makes it difficult to conclude). The effect of ignoring data for ages above 11, and apply the model to predict from ages 11-18, results in a big bias for SSB while the bias for F appears low. The precision in these estimates are very low (outside the scales of the plot) due to predicting abundance with uncertainty over 7 ages with no data. The high uncertainty does ensure a reasonable coverage, but with these levels of uncertainty, the estimates may be concluded to be of very little value.

The conclusion from this is that the dynamic pool model works satisfactory, and that increasing the plus-group might result in reduced precision. However, that will depend on the error structure in the available data as large uncertainties for older ages suggest grouping them into a plus-group. Based on this the dynamical pool model will be applied for the rest of the analysis shown in this document.

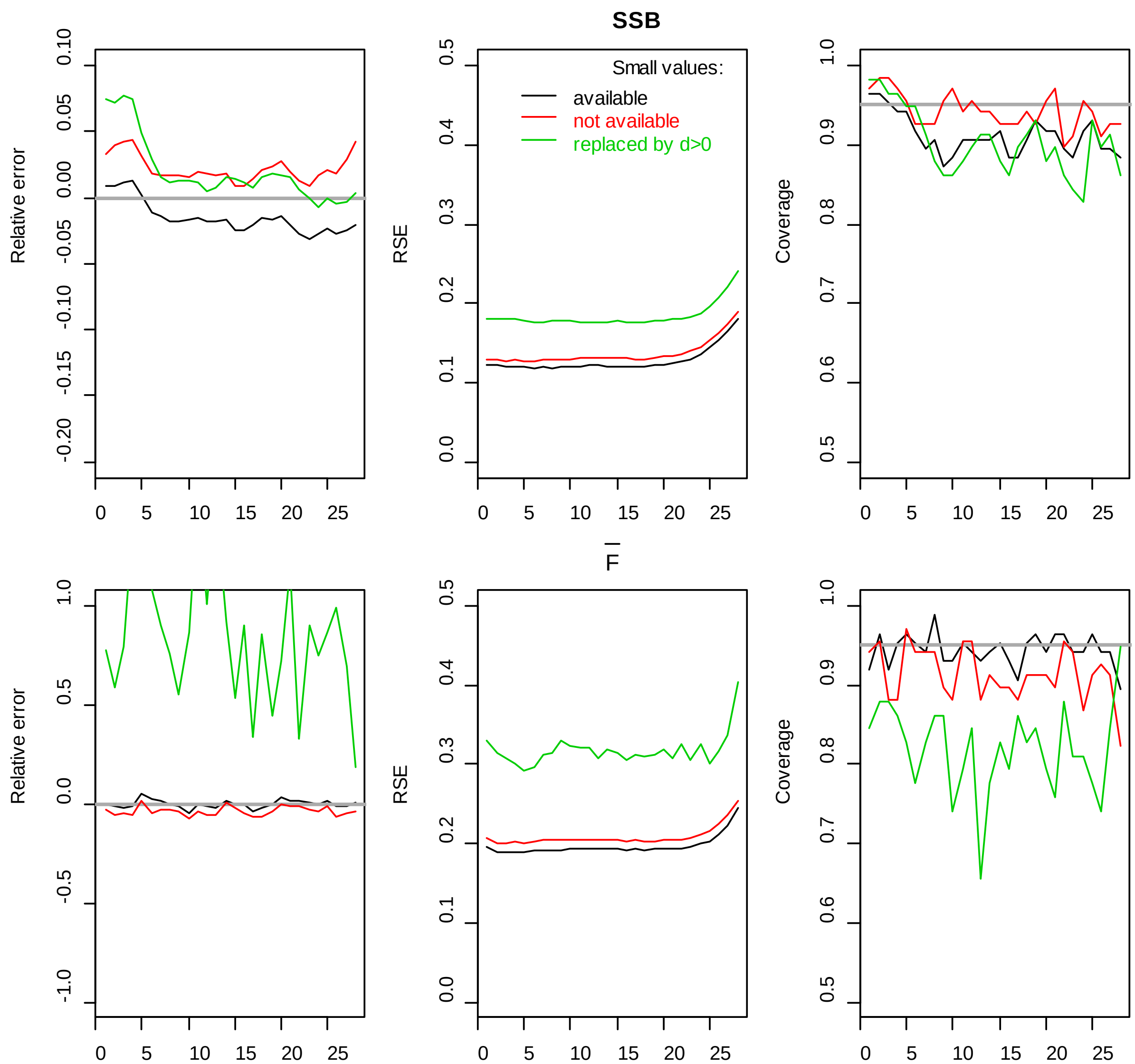


Figure A4.1. Relative bias (left column), relative standard error (middle column) and coverage at 95% level for spawning stock biomass (top row) and average fishing mortality (bottom row) for 3 different ways of dealing with small values: the black shows the means when all data is available, the red when 3% of the smallest catches and 6% of the smallest indices are set to zero and then the zeros are set to omitted from the analysis, the green when 3% of the smallest catches are replaced by the same small constant and 6% of the smallest indices are omitted from the analysis. The statistics are means of 100 replicates of population trajectories and catches with corresponding observations from the observation model.

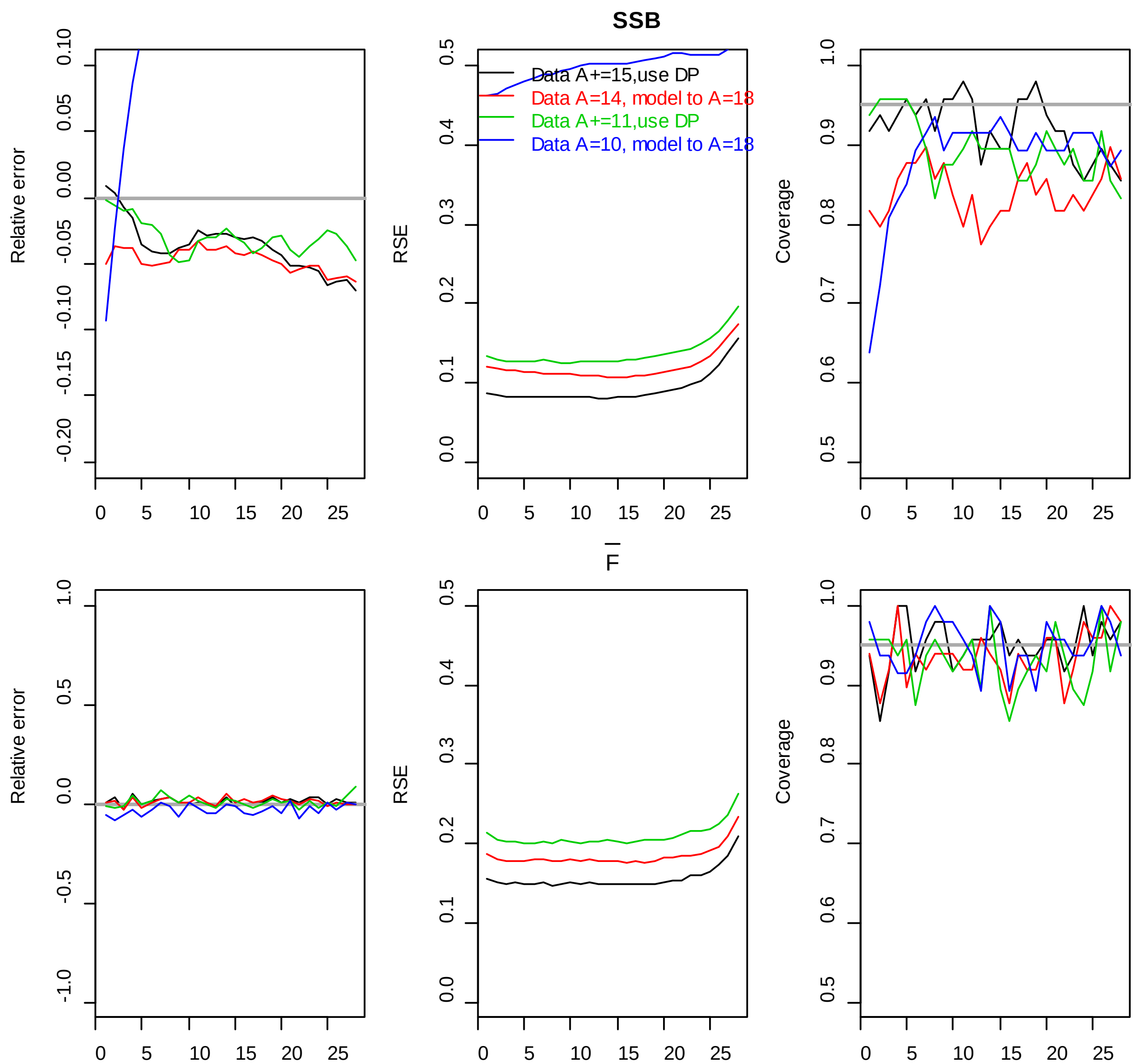


Figure A4.2. Relative bias (left column), relative standard error (middle column) and coverage at 95% level for spawning stock biomass (top row) and average fishing mortality (bottom row) for 2 different plus-groups and two different ways of dealing with the plus-group in the model: First we set the plus-group to age 13 for the data and use the dynamic pool model to model the plus-group (black lines), then the plus-group is removed from the data (i.e. using ages 1-12) but the model is run to age 18 (red lines), the plus-group for the data is reduced to 9, and we apply the dynamic pool model to estimate the total (green lines) and finally the plus-group is removed from the data (i.e. using ages 1-8) but the model is run to age 18 (blue lines). The statistics are means of 100 replicates of population trajectories for a population with maximum age 18 and catches with corresponding observations from the observation model.

Appendix 5

Additional plots

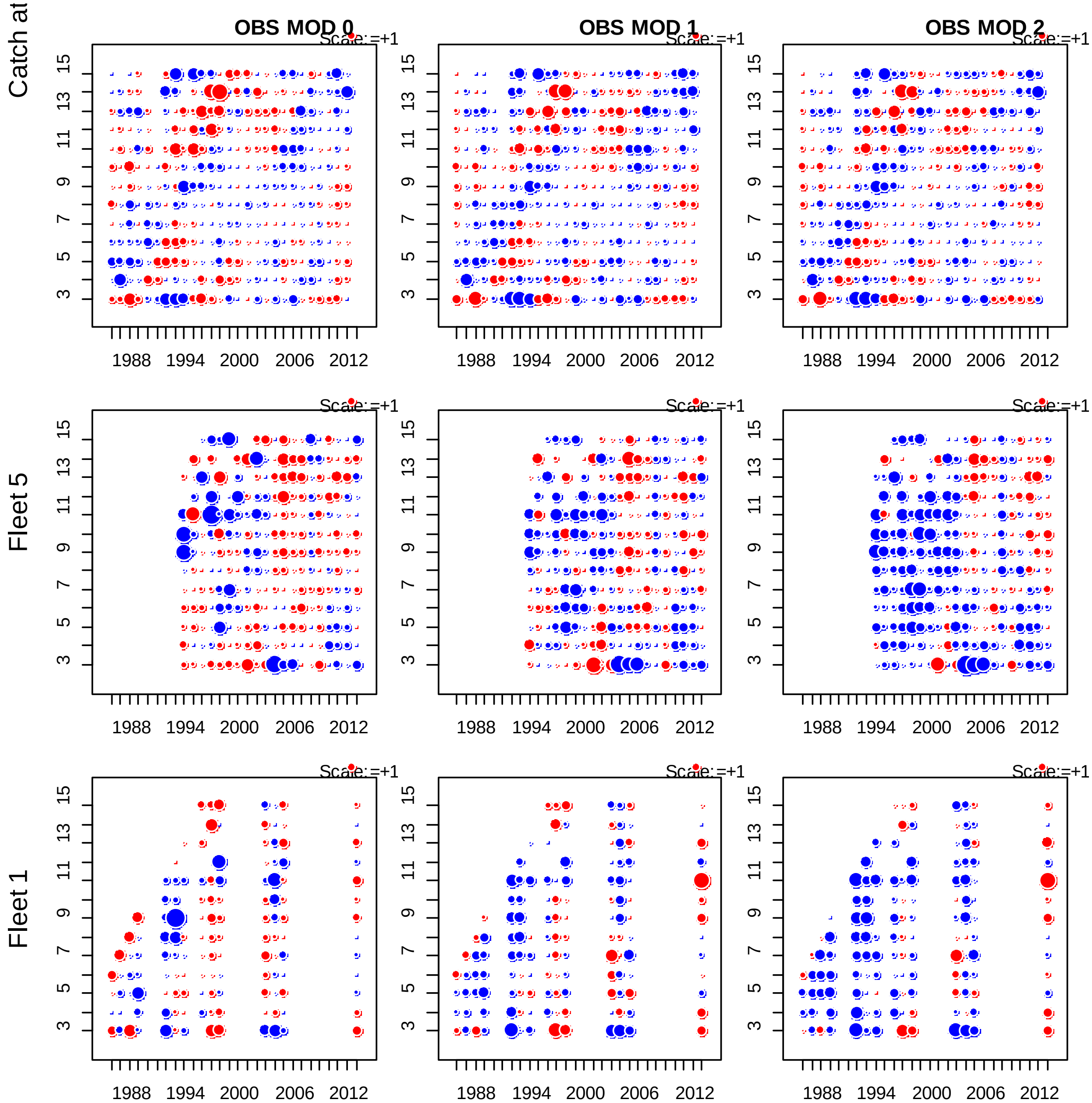
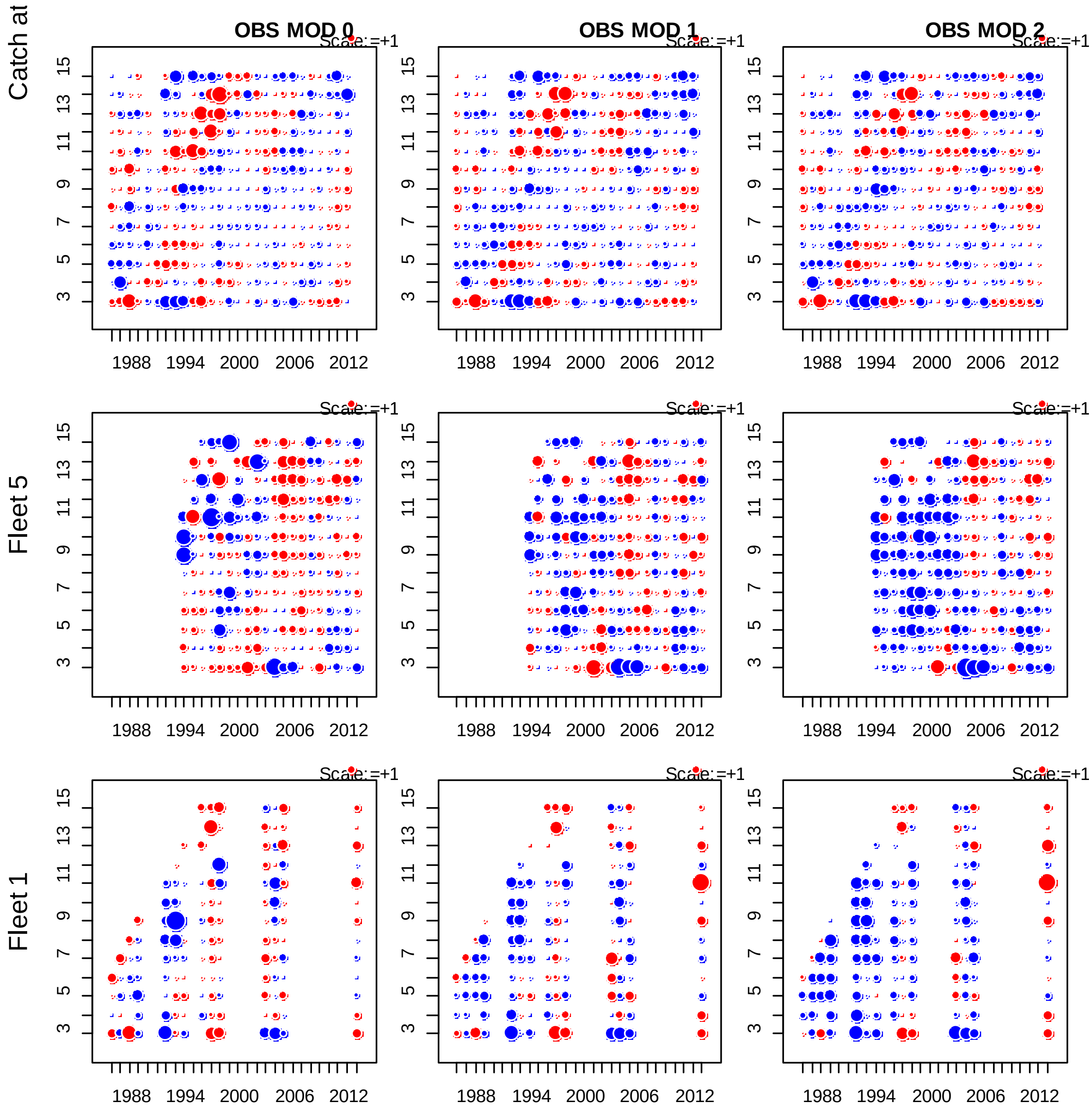


Figure A5.1. Residual when including fleet1 and fleet 5



Catch at
Fleet 5
Fleet 1
OBS MOD 0
OBS MOD 1
OBS MOD 2
Scale:=+1
3
5
7
9
11
13
15
1988
1994
2000
2006
2012

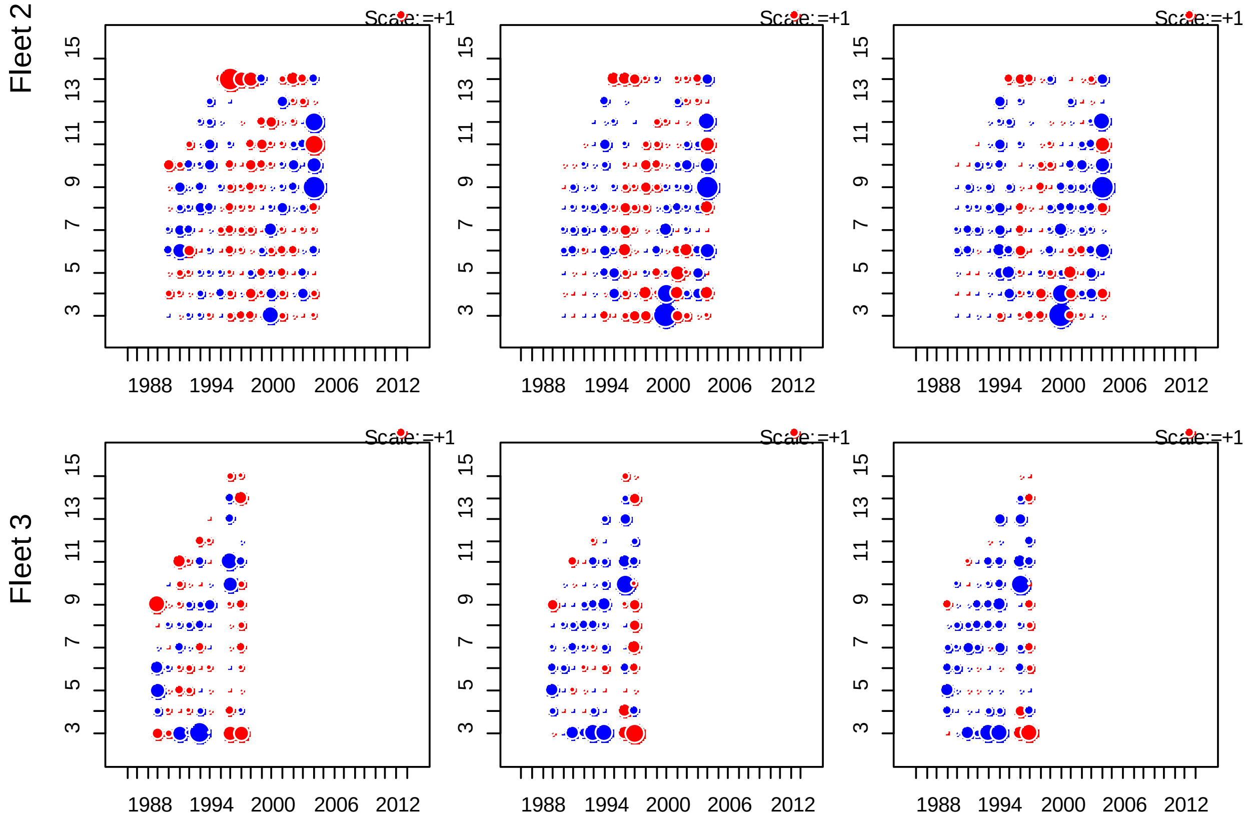


Figure A5.2. Residual when including fleet1, fleet 2 fleet 3 and fleet 5

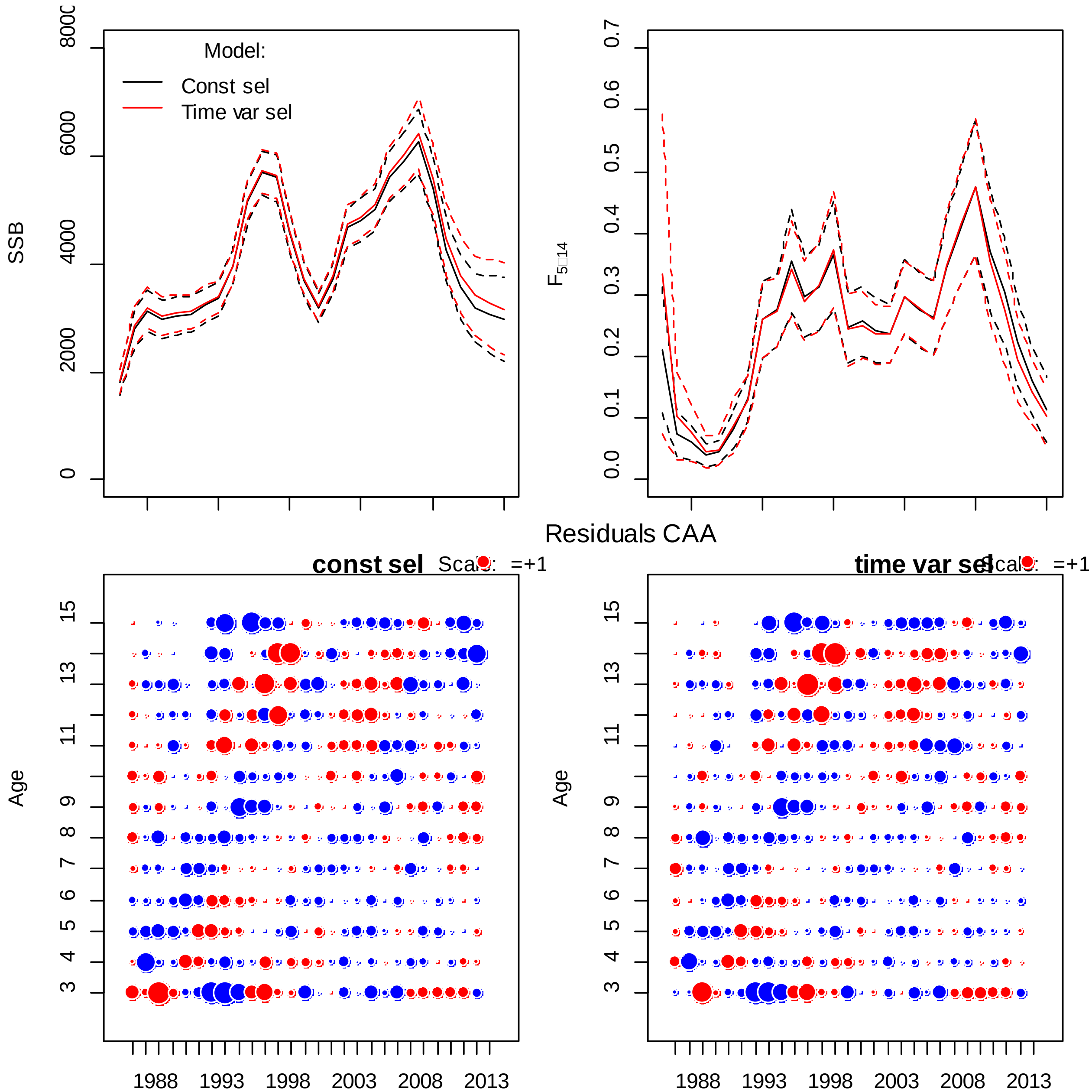


Figure A5.3. Fitting XSAM to catch at age and abundance indices from fleet 5, 1, 2 and 3 with constant selectivity (black lines) and time varying selectivity (red lines). The residuals for catch at age is shown in the bottom row. The observation model 2 was used.

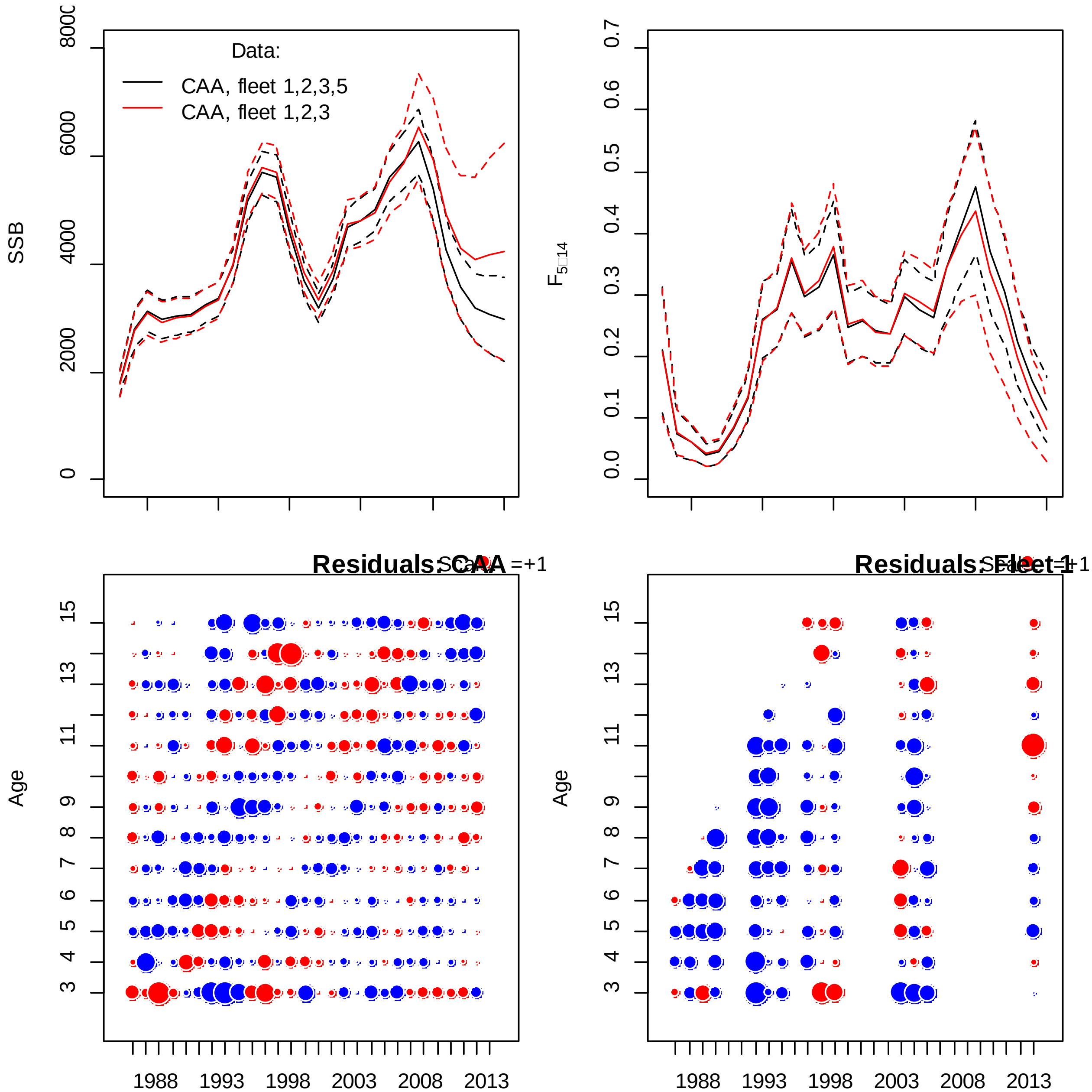
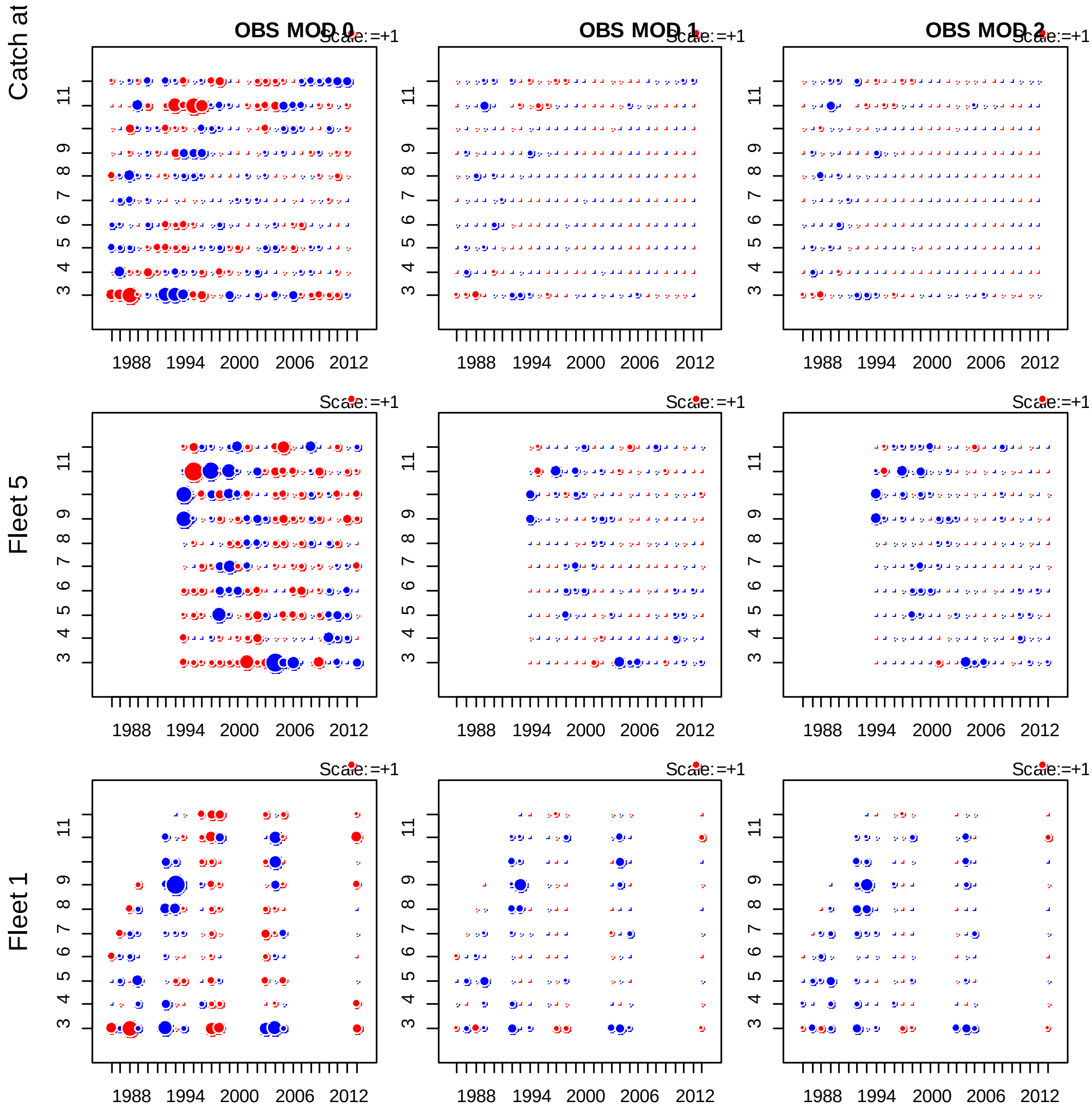


Figure A5.4. Comparing estimates by excluding fleet 5 from the analysis. The residuals for the latter is shown in the bottom row.



Catch at
Fleet 5
Fleet 1
OBS MOD 0
OBS MOD 1
OBS MOD 2
Scale:=+1
1988 1994 2000 2006 2012
3 4 5 6 7 8 9 11

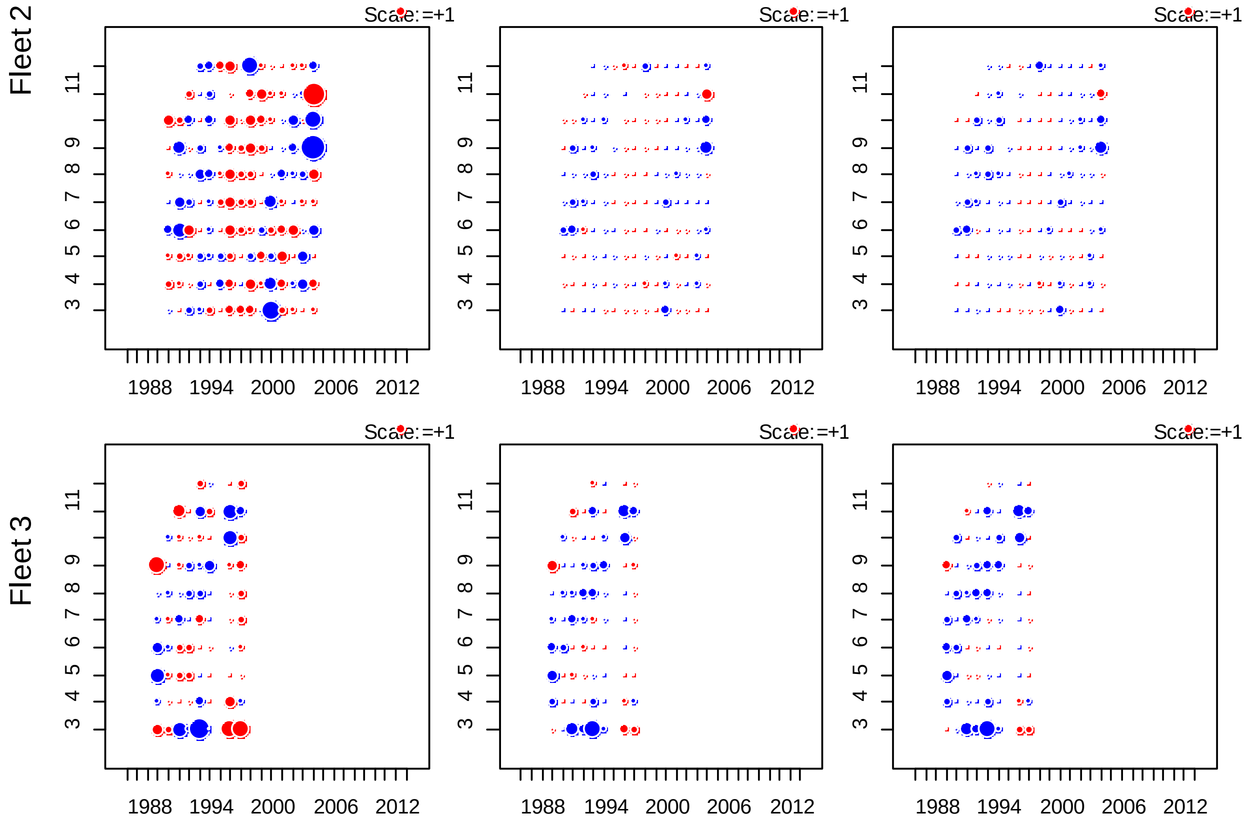


Figure A5.6. Residuals by observation model and data source for A+=12