

Evaluation of NMFS Toolbox Assessment Models on Simulated Groundfish Data Sets

by Liz Brooks, Chris Legault, Paul Nitschke, Loretta O'Brien, Kathy Sosebee, Paul Rago, and Al Seaver

A Working Paper in Support of Term of Reference 2

This information is distributed solely for the purpose of pre-dissemination peer review. It has not been formally disseminated by NOAA. It does not represent any final agency determination or policy.

GARM 2008 Methods Meeting
Woods Hole, MA
25-29 February 2008

Table of Contents	
Executive Summary	3
Introduction.....	4
PopSim.....	4
Cases Evaluated	5
Georges Bank Yellowtail Flounder	6
Georges Bank Cod	6
White Hake	6
Model Implementation.....	7
Results.....	8
Georges Bank Yellowtail Flounder (GBYT).....	8
Georges Bank Cod (GBCOD)	9
White Hake (WHAKE).....	10
Conclusions.....	10

Executive Summary

A simulation study was performed to evaluate the performance of five NOAA Fisheries Toolbox assessment models (AIM, ASPIC, SCALE, VPA, and ASAP). Data sets corresponding to three representative groundfish stocks (Georges Bank yellowtail flounder, Georges Bank cod, and white hake) were simulated with PopSim, a simulation program in the Toolbox. For each simulated stock, a base case data set was produced as well as three data sets with a known error. There were 12 data sets in total (three stocks with four data sets each) and for each data set, 100 random realizations were generated with PopSim. Each model performed an “assessment” on the simulated datasets, and the results were compared with the “true” value (i.e., the known parameter values used to generate the data sets). Results for each model in each of the 12 cases were summarized with respect to bias and precision (CV). The base case served as a benchmark to determine how well each model could replicate the truth, and as a point of comparison for model performance on the data sets with known error. In general, no model was a clear winner in all cases. Data sets that reflected errors associated with sampling (aging error or number of length samples) were best handled by models that either did not use age (AIM) or models that incorporate error into catches (ASAP). The VPA, because it matches catch exactly, suffered the most bias and had the poorest precision in these cases. However, when the source of error introduced a “break” in the time series (as in all of the yellowtail flounder cases), none of the model configurations was robust to the effect. The “east coast” approach of tuning to age-specific survey indices appears to be robust to the shape of the selectivity function. In the case of misspecification of the fleet selectivity (assuming logistic when it is dome), both forward and backward projecting models were impacted, but the effect was only apparent at the oldest ages (as would be expected). ASPIC failed in all simulated data sets, but this was due to the nature of the simulated data (all of which were one-way trips), and not to deficiencies in the model.

Introduction

This exercise was conducted to examine general properties of a range of stock assessment models available in the NOAA Fisheries Toolbox (NFT) under known conditions similar to those seen in New England groundfish. It was not designed to pick one model as the “winner” and require its use in all stock assessments. The population simulator (called PopSim in the NFT) was specifically designed to work seamlessly with the NFT stock assessment programs and allow comparison of identical datasets among models. Random noise is added to known conditions to create multiple realizations from a given scenario. Each random realization contains the same information for each of the models and therefore comparisons can be made directly among models using fewer than typical number of realizations.

As in all simulation work, there are many limitations to the conclusions that can be drawn from this study. Due to the automatic nature of the simulations and stock assessment, there is limited ability to “tune” each stock assessment model to a particular dataset. Each model is therefore set to “typical” conditions, and the result from each data set is “final” in the sense that the analyst does not have the ability to examine diagnostics and re-configure or re-run the model. In order to ensure that all datasets contained the same random deviates, a consistent time period was required for all models in each scenario. This often does not occur in real assessments when some models are able to use information from time periods without age composition data, for example. Additionally, the PopSim model is limited in the types of uncertainty that can occur in the datasets. Specifically, there are no spatial components, sex-specific differences, density-dependent effects, nor calculated management interventions. However, even with these limitations, PopSim provided an excellent framework for this study and some general conclusions were reached.

PopSim

PopSim is both an age and length based simulator. The user defines:

- dimensions of the scenario in terms of number of years, ages, and lengths
- initial stock abundance at age
- recruitment values or a stock recruitment relationship
- annual length based fishery selectivities and fishing intensities (separable F)
- biological characteristics
 - natural mortality
 - growth, length-weight, and maturity equations
 - variability in growth
- market categories for sampling length distributions
- annual sampling intensities by market
- ageing precision matrix
- survey characteristics (common to all stock assessment models)
- index characteristics (specific to stock assessment models).

Standard fishery science equations are used throughout the model to generate population abundance at age, catch, and survey values for the true conditions. These true values are used as the basis for comparison with stock assessment model estimates from 100 random realizations of the true data. There is the possibility of a small amount of bias being introduced between the true data and any assessment model—even if the dynamic equations were identical—because only 100 data sets were examined. Nevertheless, this was a feasible amount of data to examine for this meeting, and examination of most of the base case models showed fairly low bias.

There are two input standard deviations which determine the uncertainty in growth. The first gives the spread of lengths about the initial population numbers at age as well as the spread of all future recruitments (age 1 in PopSim). The second standard deviation is used to create a growth transition matrix for each age. The expected growth from one age to the next is based on calculated lengths at age from the von Bertalanffy equation and then a normal distribution about this expectation is created from the second growth standard deviation, with the limitation that fish are not allowed to decrease in size. The normal distributions from all possible starting lengths at age are summed to create the growth transfer matrix for that age. By changing the two growth standard deviations, variability in growth can be made to follow a number of patterns at age.

Variability in the realizations provided to the stock assessment models is incorporated in PopSim in a number of ways. The annual total catch in weight values and recruitments are lognormally distributed. Noise is added to each age of each survey and applied to all lengths within the age according to a lognormal distribution. The indices used by a specific stock assessment model are formed by summing the surveys with noise over length, age, or both. For example, if a survey has ten ages with 20% CV at each age, a total biomass index would be formed by summing all ages and lengths for a year times the weight at length. Lengths are sampled from the catch according to market category. These length distributions are used with either the true age-length key, or by randomly sampling ages from a fraction of the sampled lengths to create a user age-length key. The sampled fish are expanded to the total catch based on the weight of the sample and the total weight of the catch. Additionally, an age imprecision matrix can be used to add uncertainty or bias to the ageing process.

Cases Evaluated

Three stocks were chosen to form the basis for scenario testing: Georges Bank yellowtail flounder (GBYT), Georges Bank cod (GBCOD), and White Hake (WHAKE). Base scenarios for each of the stocks were created to mimic, but not exactly replicate, features of the stock assessment. The basic dimensions of the problem were set according to the last assessment for each stock (Table 1). Initial population abundance at age, recruitment series, and annual fishing intensities were taken directly from the previous assessment, resulting in all three cases having high fishing mortality throughout the time series. Each stock had a base case scenario and three test scenarios that were selected to examine

specific features of that stock assessment. The base case for each stock introduced no bias or model misspecification, rather it served as a point of reference to evaluate a given model's ability to reproduce the true values.

Georges Bank Yellowtail Flounder

A major feature of the Georges Bank yellowtail flounder assessment is a strong retrospective pattern observed in the Base Case VPA. The three test cases for GBYT were chosen to create retrospective patterns in VPA by changing the catch, natural mortality rate, or survey catchability for years 1995-2006. The change catch scenario was created by changing the sample length-weight equation relative to the population length-weight equation such that the catch at age in numbers were approximately half the true values beginning in 1995. This approach was selected because of concerns regarding changes in average weight at age observed in many New England groundfish stocks. Choosing this approach created a situation where models which did not use age-based catch information did not "see" the problem because the total catch in weight was not changed relative to the truth (beyond the small amount of uncertainty which was applied in all years). The natural mortality change was from 0.2 for years 1973-1994 to 0.5 for years 1995-2006. This survey catchability change scenario tripled each age specific value in years 1995-2006 relative to the base values for years 1973-1994.

Georges Bank Cod

The possibility of domed selectivity in either the fishery or surveys (or both) has been suggested for Georges Bank cod in the past. Doming of selectivity is expected to cause problems if an assumption is made that selectivity is flat-topped, and vice-versa. The scenarios examined for GBCOD were 1) base with flat-topped selectivity for both the fishery and surveys, 2) fishery domed and survey flat-topped, 3) survey domed and fishery flat-topped, and 4) both fishery and survey domed. Some difficulties were encountered when creating these scenarios. The flat-topped selectivity pattern was originally set based on matching actual fishery characteristics, but PopSim did not allow a reasonable dome to be formed due to the double-logistic equation employed not being rescaled to one. The fleet dome scenario therefore had the ascending limb of the selectivity pattern shifted to the left approximately 20 cm. Additionally, it was discovered after running many of the models that the sampling of lengths by market category in all four scenarios was much lower than it should have been. This led to problems with SCALE in particular due to the large number of length bins in each market category.

White Hake

White hake have a history of uncertain ageing and even species identification for smaller animals (they are often confused with red hake at small sizes). Furthermore, the fishery lands white hake headless, meaning the fishery samples cannot be aged but instead survey age-length keys must be applied to commercial landings and discards. This makes the catch at age information much more uncertain for white hake than the other two species. The four scenarios selected for this species were 1) base, 2) additional age

uncertainty due to using sampled age-length keys instead of the true age-length key when creating catch at age (denoted Fraction below), 3) use of an ageing error matrix that contained bias at the youngest ages and increasing noise but not bias at older ages, and 4) a combination of cases 2 and 3. These ageing errors did not impact the ages of survey indices, nor did it impact the “known” growth parameters in the SCALE model.

Model Implementation

All models were initially tuned to the base case data set for each stock, and those modeling decisions were maintained throughout the data sets that introduced misspecification.

AIM tracks the relative fishing mortality rate of the population and computes a measure of the finite rate of increase from the survey indices. As such, its quantities are not strictly comparable to standard outputs of assessment models that generate absolute measures of abundance and fishing mortality rates. To facilitate this comparison with the simulated population, the true value for relative F was computed as the sum of the total catch in weight times a weighted average q and that product was then divided by the sum of the survey indices (in scaled units). The weighted average q was computed as the age specific q times the true initial population abundance divided by the total population abundance. Relative F was estimated as the ratio of total catch to a 3-yr lagged average abundance estimate. The finite rate of increase, or replacement ratio, was computed as the ratio of the survey index in year t to the average index in the preceding 5 years. Note that in the AIM comparison plots, the first five years of replacement ratio and the first three years of relative F are not plotted because those are the intervals used to calculate the first value.

ASPIC initial guesses and parameter bounds were derived from the simulated data. Initial guesses for MSY and K were the maximum catch and ten times the maximum catch, respectively, and bounds were fixed at 100 times larger and smaller than this. Initial guesses for $CPUE$ and index q_s were derived by dividing each index by the catch series. The level of depletion in the first year was initially estimated, but none of the base models had data that were informative for this parameter. Additional runs where the initial depletion was fixed at 1.0 (unexploited) or 0.75 were made. None of these fits was satisfactory, but for the sake of comparison with the other assessment models, results where the initial depletion was fixed at 1.0 are summarized.

SCALE initial guesses (F_{start} , F_s , initial recruitment, q_s , logistic selectivity parameters) were set in the vicinity of the actual estimates of the simulated population. Higher relative weight was given to fitting the catch compared to the survey abundance. The fits to the catch length frequencies had the same weight as the surveys. The overall growth model (mean lengths at age) was taken from the simulated population. The variation around the mean length at age was also approximated from the simulated population. In the fix case runs, we assumed the misspecification was known and in most cases the weights on the components of the objective junction were changed to down weight the

component that produces the misspecification. In the case of a change in M from 0.2 to 0.4 we assumed a constant M of 0.3 over the time series.

VPA was initialized with stock estimates and partial recruitment similar to the assessment, the catch equation was estimated using Pope's approximation, and the plus group was estimated 'backward'. In the terminal year, full F was estimated using the 'classic method', i.e. average F of fully recruited age classes, and F on the oldest true age was estimated from the partial recruitment vector. Prior to the terminal year, F at the oldest true age was estimated using Heinke's method. All simulation data sets were analyzed using these same settings.

ASAP was run in a "typical" first pass setting for all 12 cases (3 stocks by 4 scenarios). Fishery selectivity was estimated as a logistic equation with 2 parameters. A Beverton-Holt stock-recruitment relationship was estimated without priors and recruitment deviations were estimated assuming a 50% CV in each year. Initial stock abundance at age was estimated without constraints. Input effective sample sizes used with the catch at age proportions were set to 200 for GBYT and GBCOD and to 50 for WHAKE based on expected levels of precision. Surveys were entered "east coast" style as age-specific indices, each assumed to have 40% CV. In no case was any "tuning" (meaning changes in input effective sample size or CVs for components of the objective function) conducted for the ASAP runs, as would be done in a real assessment.

Results

Results for all simulations are summarized in a standardized format to facilitate comparisons between models. For each stock, the % bias and the CV of common model output is summarized by overlaying the results from each of the stock-specific simulated data sets. The scale of the y-axis was fixed to a range of +/- 75% for bias plots, and [0,1] for CV plots. The presence of "blank" plots indicates that results for that model were either very biased, very imprecise, or both. Figures for each stock are labeled sequentially from 1-18, with a prefix that corresponds to the stock being considered.

Georges Bank Yellowtail Flounder (GBYT)

Model estimates of an aggregate measure of abundance for the base case data set were generally unbiased for AIM and SCALE, while ASAP had a slightly positive bias (~ 2-20%); the VPA estimates of spawning biomass at age was positively biased, most notably at the youngest age classes (Figures GBYT.1-GBYT.18). The pattern of fishing mortality estimates was complementary to the pattern for abundance. All results for ASPIC were severely biased due to the data following a one-way trip, and the lack of a relationship between annual changes in biomass and annual removals. Compared to the base case, all models failed in a predictable manner when each of the misspecifications was introduced. Increasing natural mortality for the last ten years led all models to underestimate abundance and to overestimate fishing mortality. When catch at age was "underreported" by one half of the true values in the last ten years of data ("Ch.catch" in figure legends), abundance was underestimated for all models but AIM, and fishing

mortality showed a strong spike of positive bias in the year just prior to the reduced catch, followed by low to moderate negative bias. In AIM, estimates of replacement ratio for the underreported catch data set were identical to the base case, because AIM only uses the survey indices to estimate that ratio, and the indices were not affected by a change in the sample length weight equation. Tripling catchability for the last ten years produced positive bias in abundance estimates and negative bias in fishing mortality estimates. In AIM, the positive bias in replacement ratio only persisted for five years, which is the width of the interval used in the calculation.

Precision for the base case run of all models was generally $< 20\%$ CV, and did not vary for the other data sets. The only exception to this result was for the SCALE model, where CVs increased from the base case, with the underreporting data set having the poorest precision.

Georges Bank Cod (GBCOD)

Both the VPA and ASAP models had very low bias in the base case, although age-specific estimates degraded for ages 8-10+ in the VPA, and for the 10+ age class in ASAP. Introducing a dome into the survey selectivity did not cause bias in either VPA or ASAP, because the models are tuned to age-specific indices rather than to a single population index with age composition data. The two remaining data sets, where a domed selectivity was used for the fleet or for both the fleet and the survey, are essentially two different random realizations of just a domed selectivity for the fleet. In the VPA, all age specific measures of abundance were underestimated and the fishing mortality at age was overestimated. In ASAP, total spawning stock biomass was biased low, as was the estimate of average F. The bias in estimated numbers at age was negative for younger ages (1-4) and gradually became positive for older ages (5-10+). The average F corresponds to ages 4-6, so its bias likely reflects the composite effect of opposing biases in those numbers at age. The VPA was very precise for the base case through age 7, with a loss of precision for ages 8-10+. Precision was similar for ages 1-5 in the other datasets, but increased substantially for ages 6-10+. ASAP was extremely precise in the base case, and there was negligible change in precision for the other data sets.

Bias in the SCALE base model ranged from 5-35%. One reason for this is that the level of sampling for catch lengths was very low given the large size range in cod. The high Fs in the simulated population and a logistic selectivity curve shifted to larger fish in the base case contributed to estimation problems for SCALE. When a domed selectivity was implemented for the fleet, the length at 50% selectivity was shifted to a smaller size compared to the base case. The shift in selectivity had the unintended effect of providing more length samples, which improved the model estimates. To evaluate SCALE properly would require that the length sampling be increased, and that the length at 50% selectivity be the same for both the logistic and the ascending limb of the domed selectivity curves.

All results for ASPIC were severely biased due to the data following a one-way trip, and the lack of a relationship between annual changes in biomass and annual removals. AIM had an unexplainable pattern in the bias for replacement ratio, due to the low sampling leading to division by zero issues in this model that were only discovered after all other models had been run. The AIM results for GBCOD should be disregarded.

White Hake (WHAKE)

All models except ASPIC had negligible bias in the base case for most model output (Figures WHAKE.1-18). The SCALE model had difficulty estimating the initial fishing mortality, which carried through to the bias in the number at age. The VPA model only showed large bias in the number at age 1 in the last model year and in the fishing mortality at age one in all years. ASAP accurately estimated numbers at ages 1-7, but the estimates were strongly biased for ages 8 and 9+. Introducing aging error had no effect on AIM because it doesn't use age information. Likewise, using only a fraction (40%) of the available length samples to create the age-length key had an almost imperceptible effect on the bias and precision of AIM's estimates. The SCALE results for the data set with aging error were identical to the base case, because SCALE uses the growth curve rather than an age length key; therefore, there was no misspecification in SCALE for this data set. In reality, if the growth curve had been estimated from samples with biased age errors, one would expect that the growth curve would also be biased. Results from SCALE were insensitive to the use of an age-length key from sampled data. This is not unexpected given that no age data were passed to SCALE. Ageing error in the VPA caused serious bias in the abundance estimates at all ages, but only seriously impacted the estimated F at ages 1-3. Use of an age-length key from sampled data led to very poor accuracy in F at most ages, while abundance estimates were only moderately biased compared to the base case. Results for the data set combining both sources of age imprecision performed very similarly to the reduced length samples. Most estimates from ASAP were robust to all of the misspecification data sets, with only minor differences in bias compared to the base case.

SCALE had very good precision for total biomass and the F multiplier, but estimated numbers at age were very imprecise. The CV was consistent across all data sets. The VPA was very precise in all estimates except for the oldest age classes (8, 9+). ASAP estimates were very precise for all but the 9+ age class, and the CVs were very consistent among data sets.

All results for ASPIC were severely biased due to the data following a one-way trip, and the lack of a relationship between annual changes in biomass and annual removals.

Conclusions

Only simulations for the VPA and ASAP models worked for all stocks. Results of AIM and SCALE were questionable for the Georges Bank cod example due to the nature of the simulated data. ASPIC performed poorly in all simulations, but this result was

expected because of the violation of fundamental production model assumptions for even the base case. All of the stocks had “one-way trips” for both the CPUE and the population abundance index. Furthermore, all of the simulated datasets used a series of recruitments created independently of the spawning stock biomass, which violates the assumed relationship between the annual change in biomass and the annual removals by the fishery. Data of this sort make it impossible for a production model to estimate the initial stock depletion, hence the absolute scale of abundance was extremely biased and imprecise. Although results were summarized by fixing the initial biomass at the carrying capacity, alternate levels of initial depletion were examined and did not improve model performance.

For the Georges Bank yellowtail flounder cases, the retrospective inducing data sets caused retrospective patterns for all four of the converged models (AIM, SCALE, ASAP, and VPA). Bias was of a similar magnitude for all models, and there was therefore no clear winner. Also, the pattern in bias was similar, which precluded drawing inferences from the bias pattern given the model structure. This implies that if a retrospective pattern is observed in an actual stock assessment, it would also be expected to cause problems for alternative stock assessment models in their base configuration.

For the Georges Bank cod cases, only VPA and ASAP runs could be compared. Their performance was very similar with respect to estimates of spawning biomass and fishing mortality. The only apparent difference was the bias at the older ages, where ASAP was more accurate (Figures GBCOD.13 versus GBCOD.17). ASAP was also far more precise in estimating numbers at age (Figures GBCOD.14 versus GBCOD.18).

For the white hake cases, AIM was unbiased and unaffected by the simulated data sets with error, because the errors were age based. SCALE results were invariant to the errors in data sets because it used the correct growth curve, and because the length samples, though reduced, were still unbiased. Between VPA and ASAP, the pattern in bias was very similar, but the magnitude was far less in ASAP. This is likely due to ASAP’s ability to incorporate uncertainty in catch at age.

In general, data sets that reflected errors associated with sampling (aging error or number of length samples) were best handled by models that either did not use age (AIM) or models that incorporate error into catches (ASAP). The VPA, because it matches catch exactly, suffered the most bias and had the poorest precision in these cases. However, when the source of error introduced a “break” in the time series (as in all of the yellowtail flounder cases), none of the model configurations was robust to the effect. The “east coast” approach of tuning to age-specific survey indices appears to be robust to the shape of the selectivity function. In the case of misspecification of the fleet selectivity (assuming logistic when it is dome), both forward and backward projecting models were impacted, but the effect was only apparent at the oldest ages (as would be expected). However, in ASAP, the bias was not serious when examining the total SSB or the average F (Figure GBCOD.15). The low bias in total SSB is likely due to the high level of fishing mortality, which leaves very few fish at the oldest ages and effectively minimizes the consequences of “missing” those few old fish.

Table 1. Summary of basic dimensions along with biological and fishery characteristics of the three stocks used in PopSim.

Parameter	GBYTF	GBCod	WhiteHake
number of years	34	27	12
number of ages	15	25	30
plus group age	6	10	9
number of lengths	70	170	200
number market categories	2	3	2
number surveys	2	3	3
CV of surveys	0.5, 0.4	0.2, 0.3, 0.25	0.2-0.5 by age
CV for recruitment	0.0001	0.0001	0.0001
CV on catch in weight	0.01	0.05	0.05
Growth stdevs	4, 1.5	5, 2	5, 1
von B Linf	50	148.2	189
von B K	0.33	0.1198	0.081
von B t0	-0.75	-0.6173	0.062
50% Mature	30	42	35
50% Selectivity Flat-topped	29	64	39
50% Selectivity Domed	N/A	44, 99	N/A
M	0.2	0.2	0.2

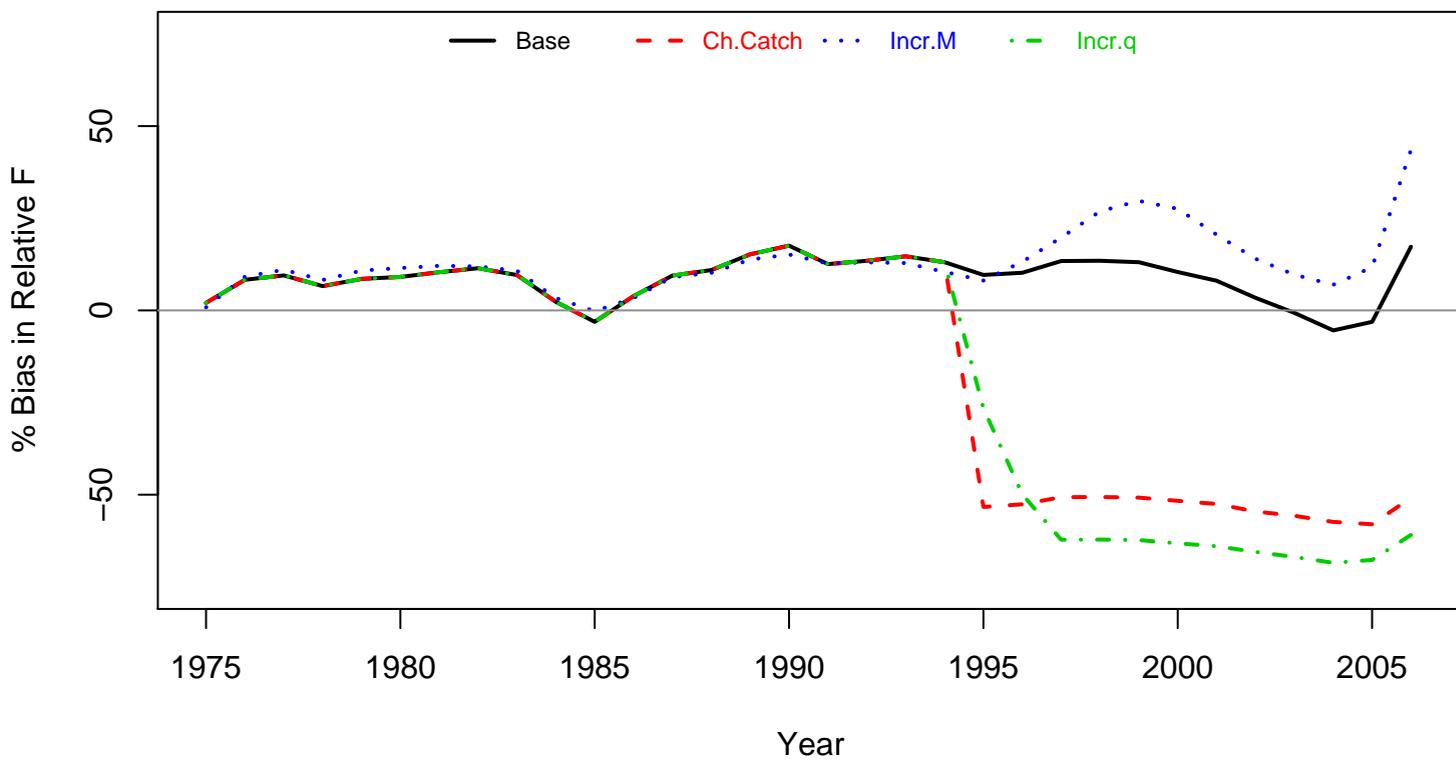
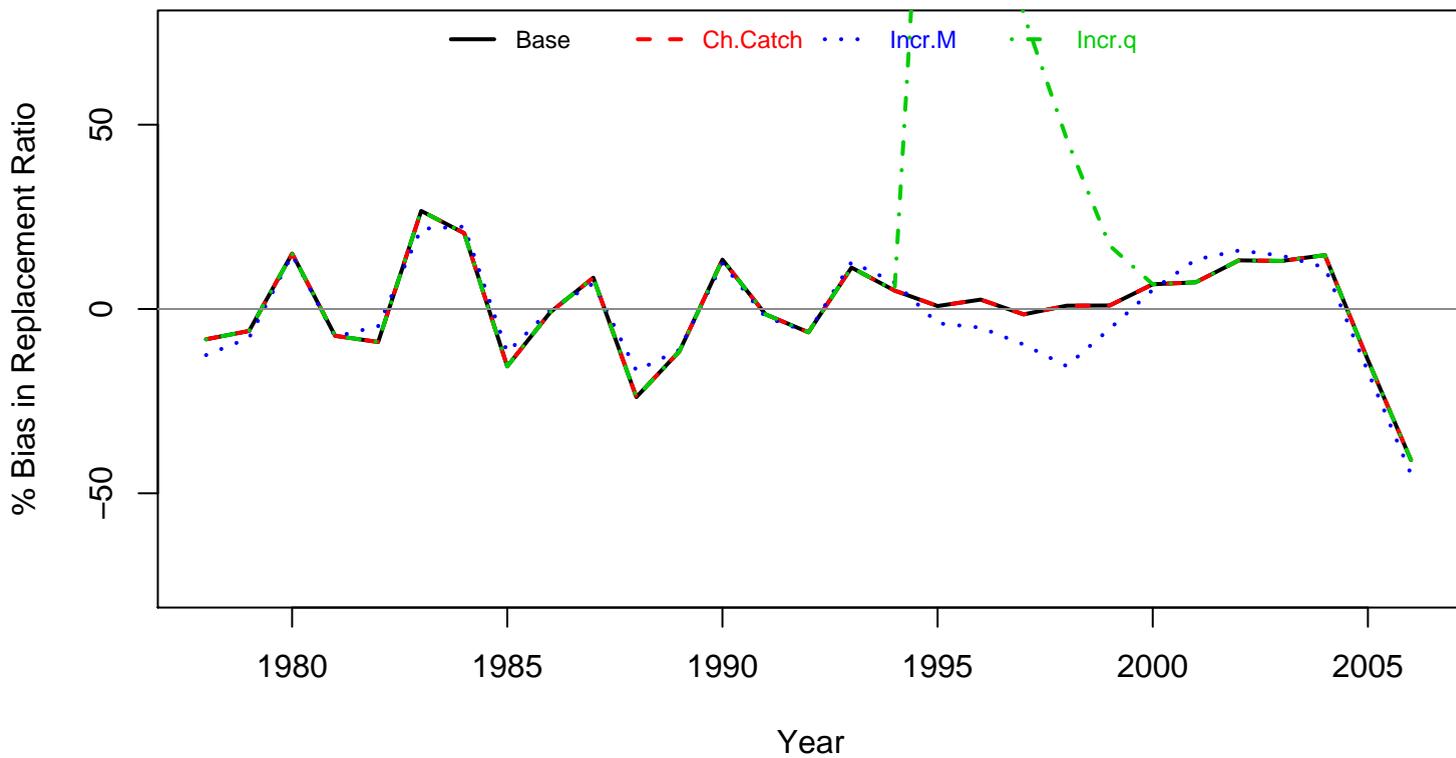


Figure GBYT.1 Bias in replacement ratio and relative F for all simulated data sets using AIM. ¹³

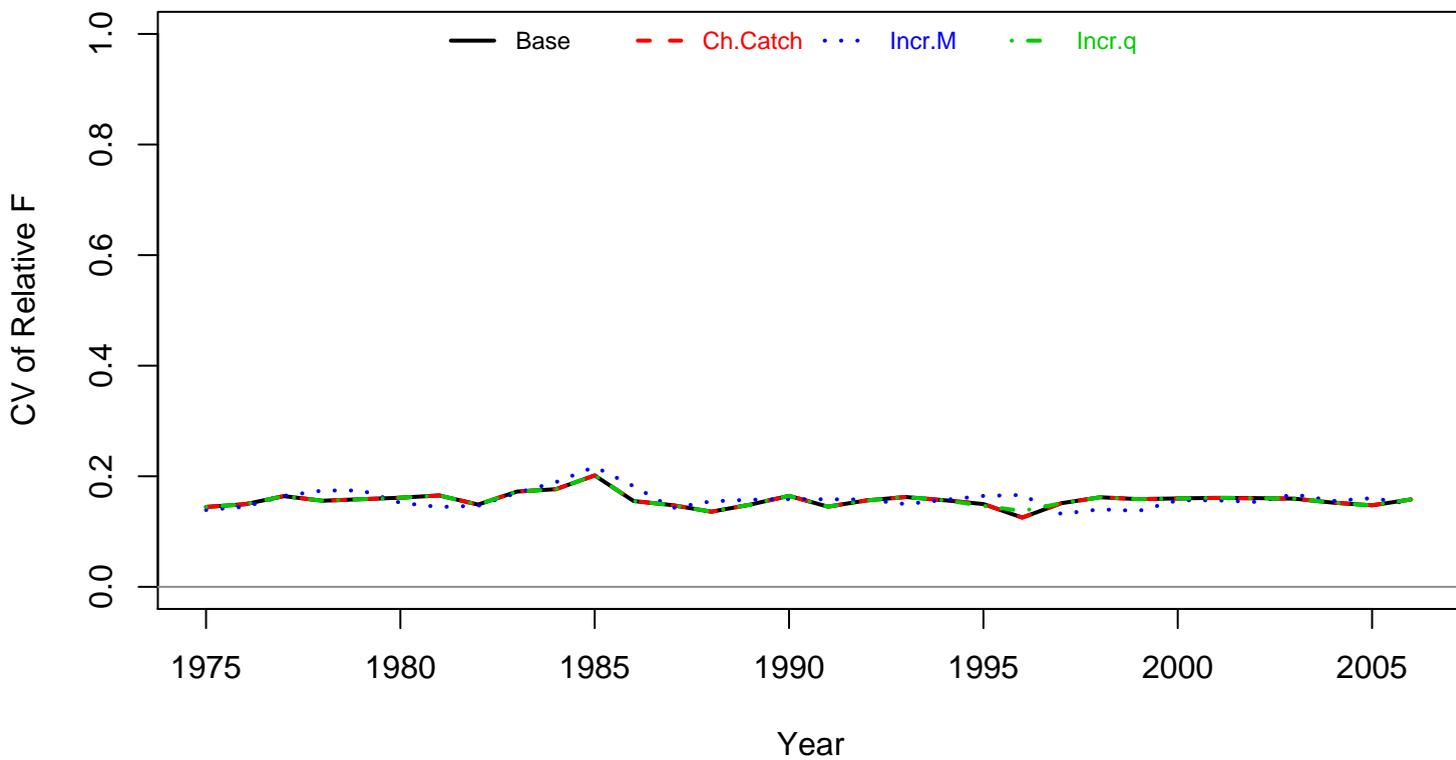
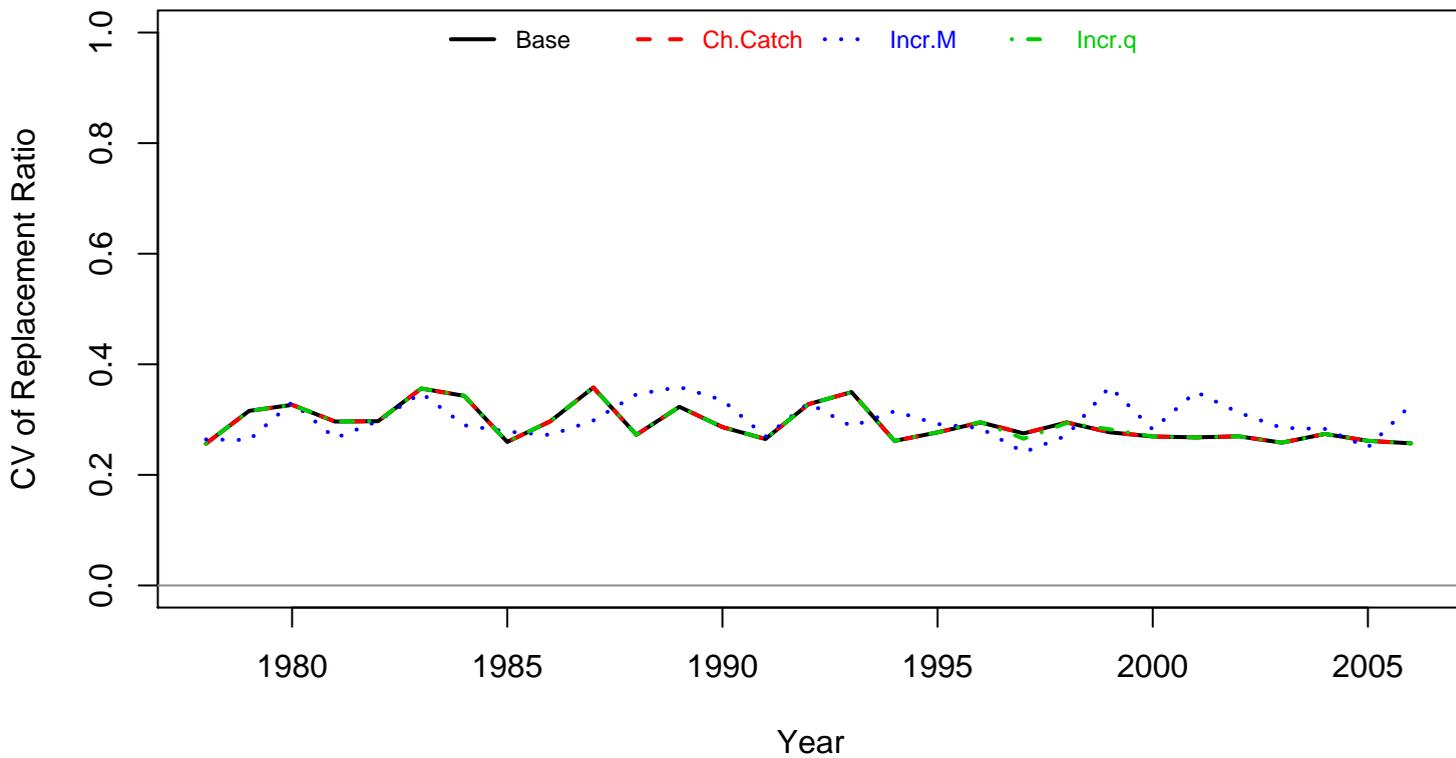


Figure GBYT.2 CV in replacement ratio and relative F for all simulated data sets using AIM. ¹⁴

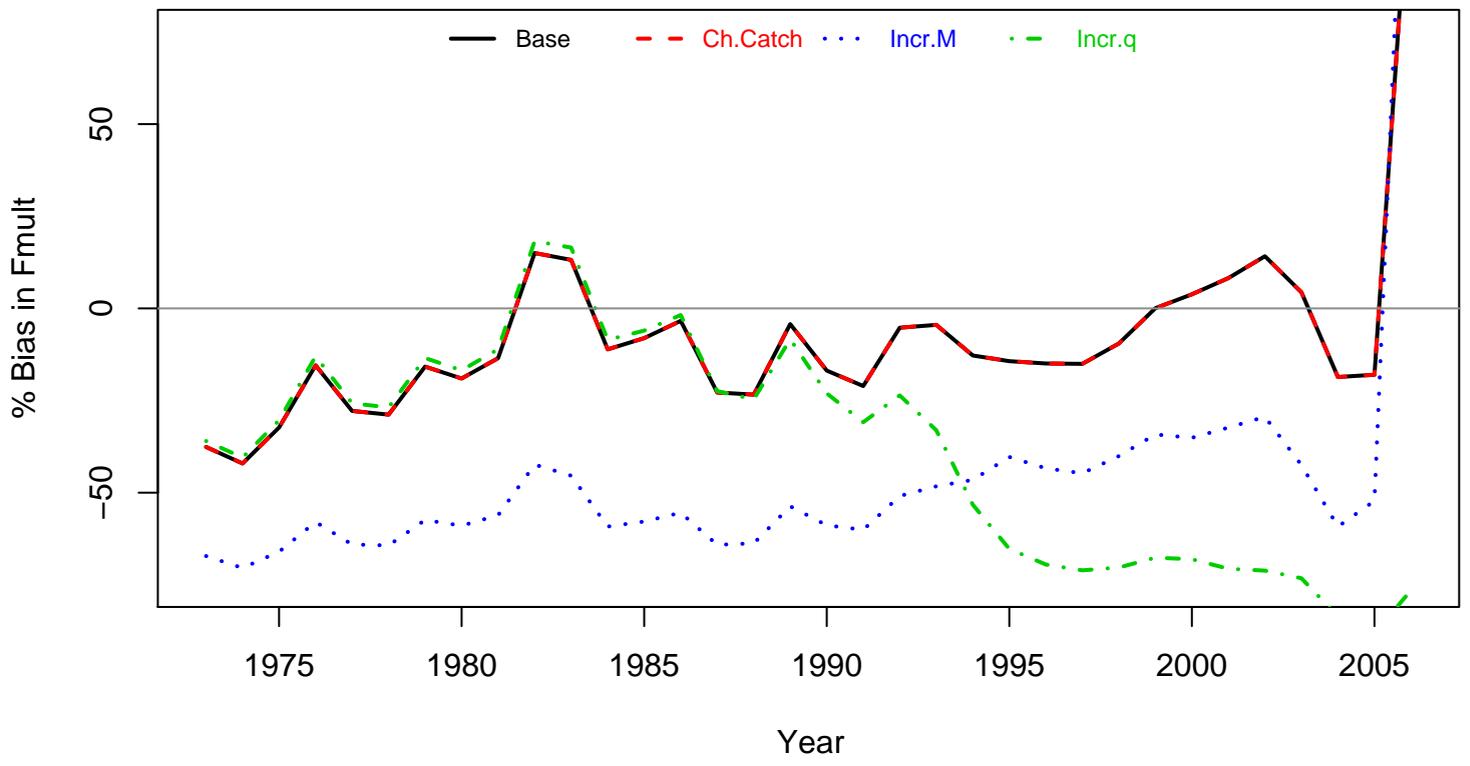
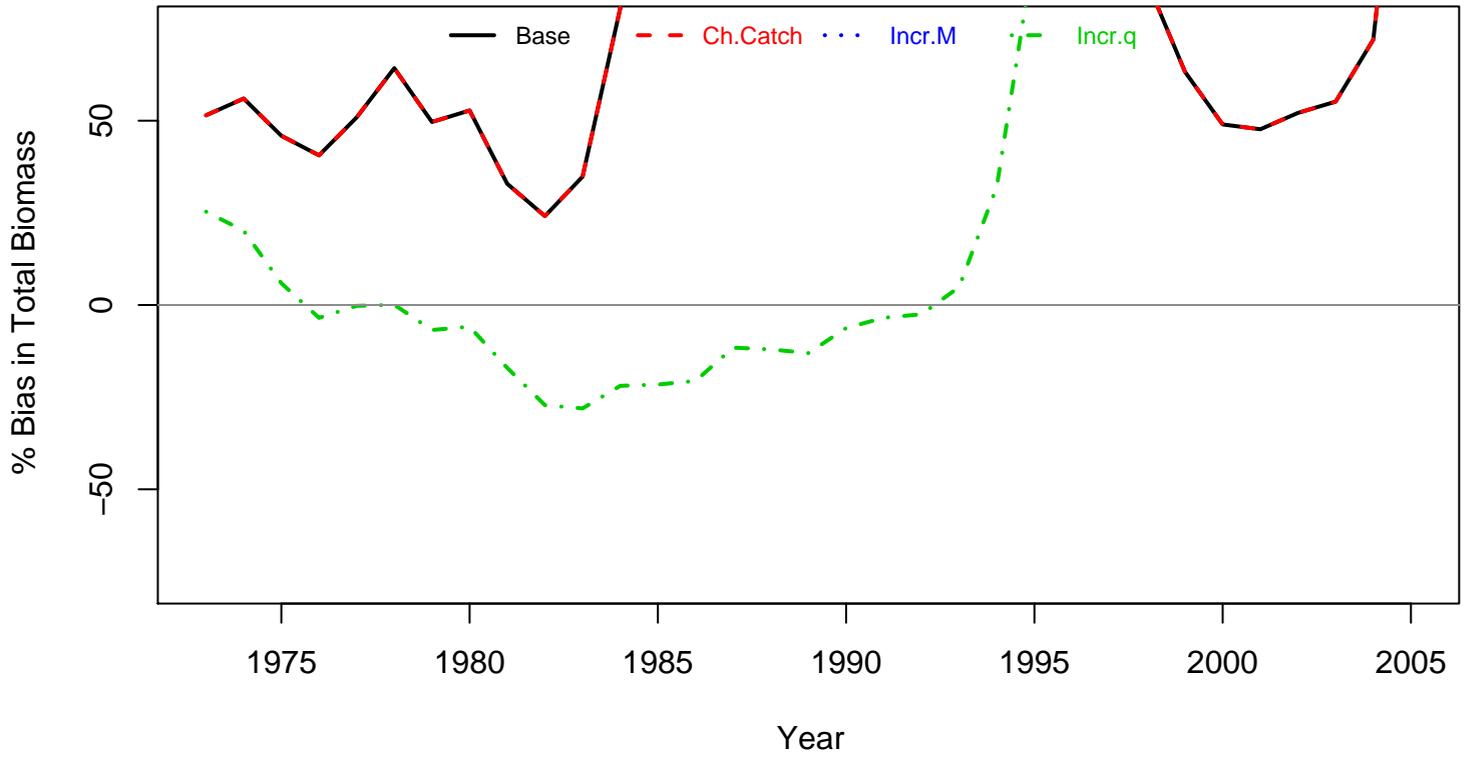


Figure GBYT.3 Bias in total biomass and F for all simulated data sets using ASPIC.

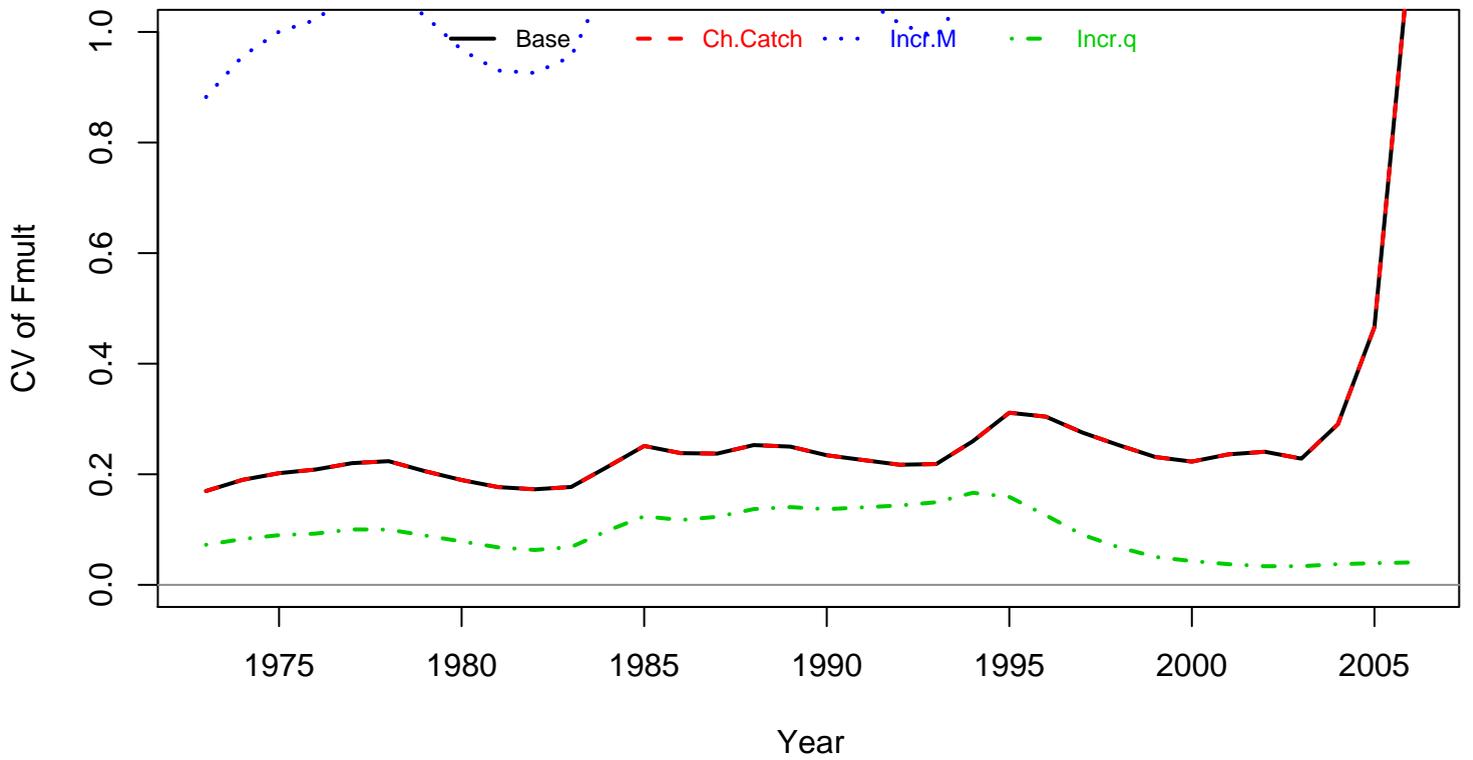
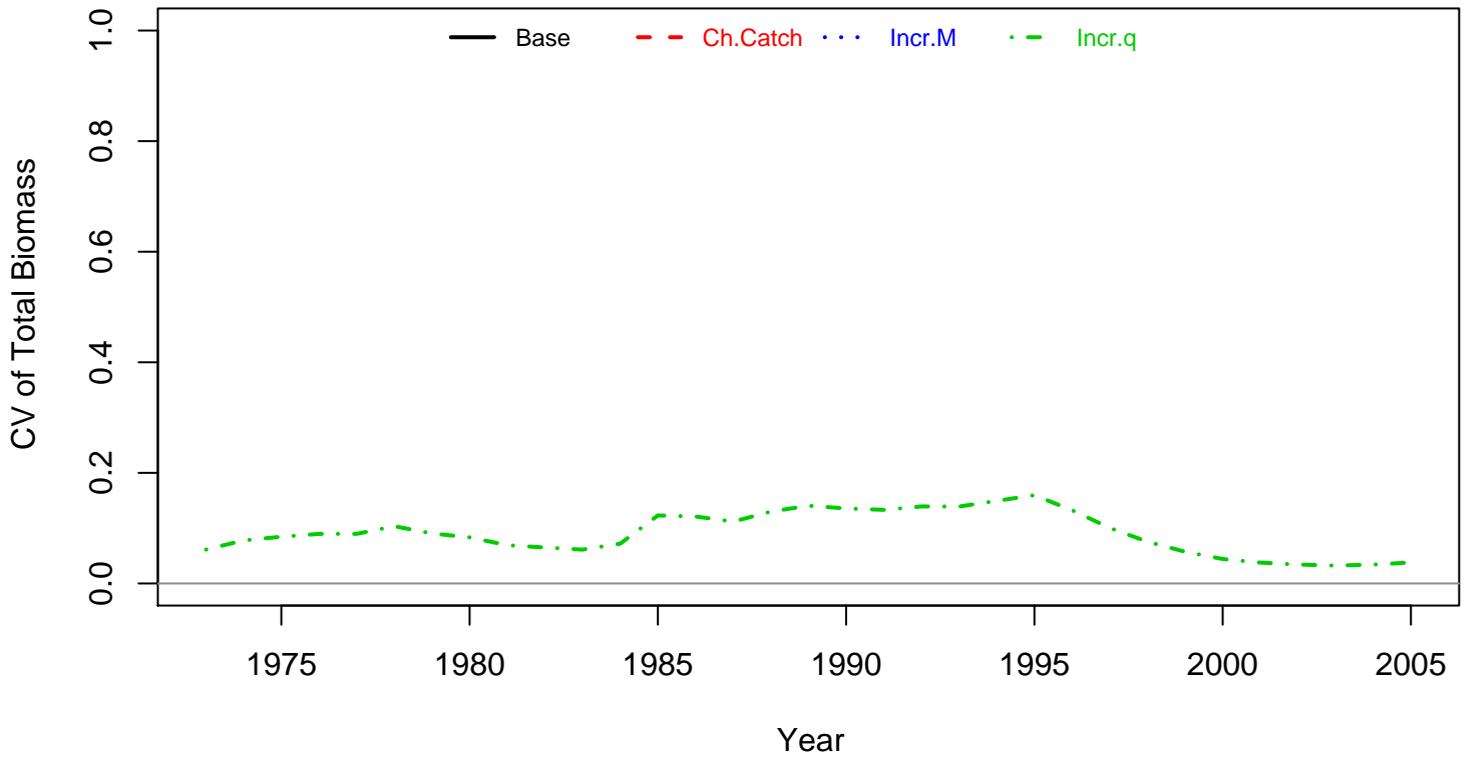


Figure GBYT.4 CV in total biomass and F for all simulated data sets using ASPIC.

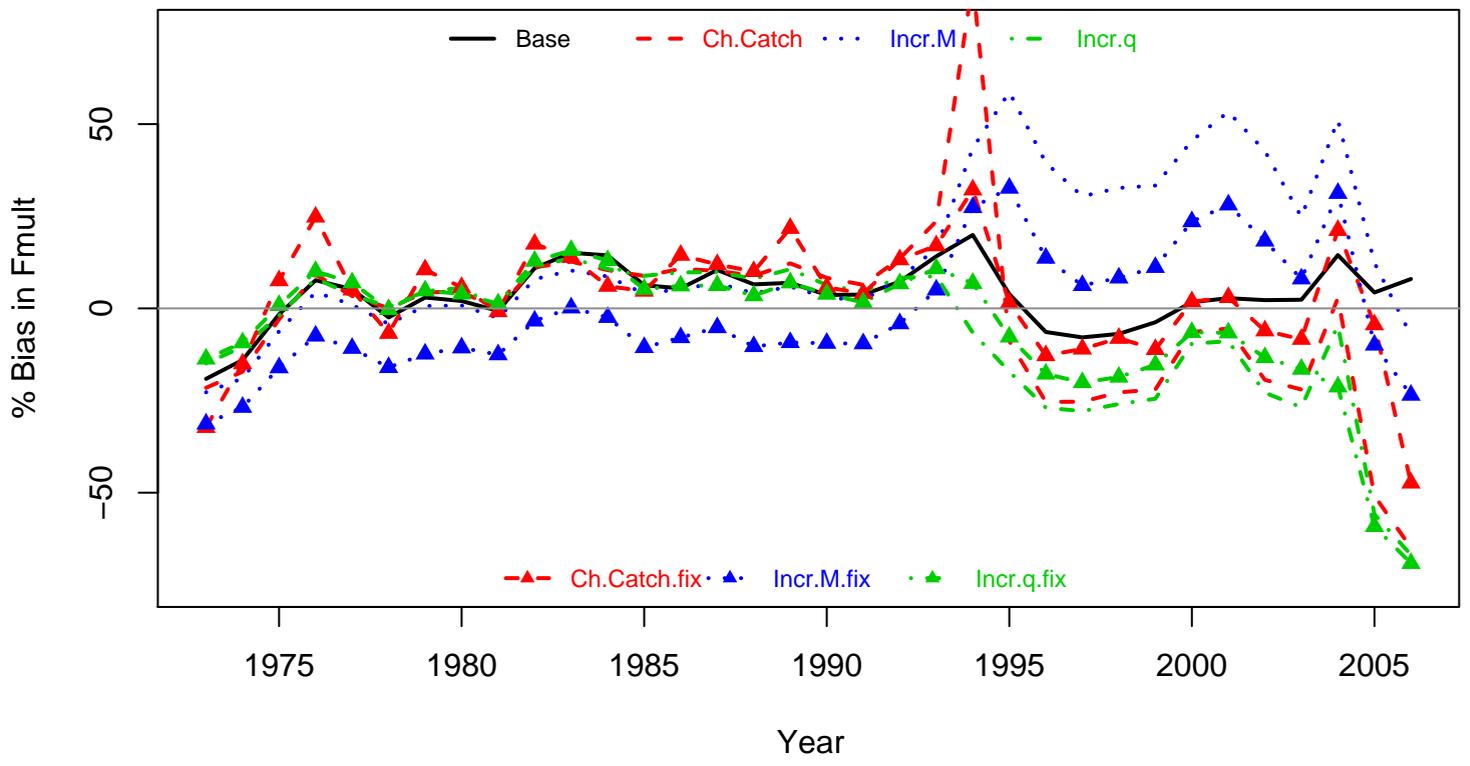
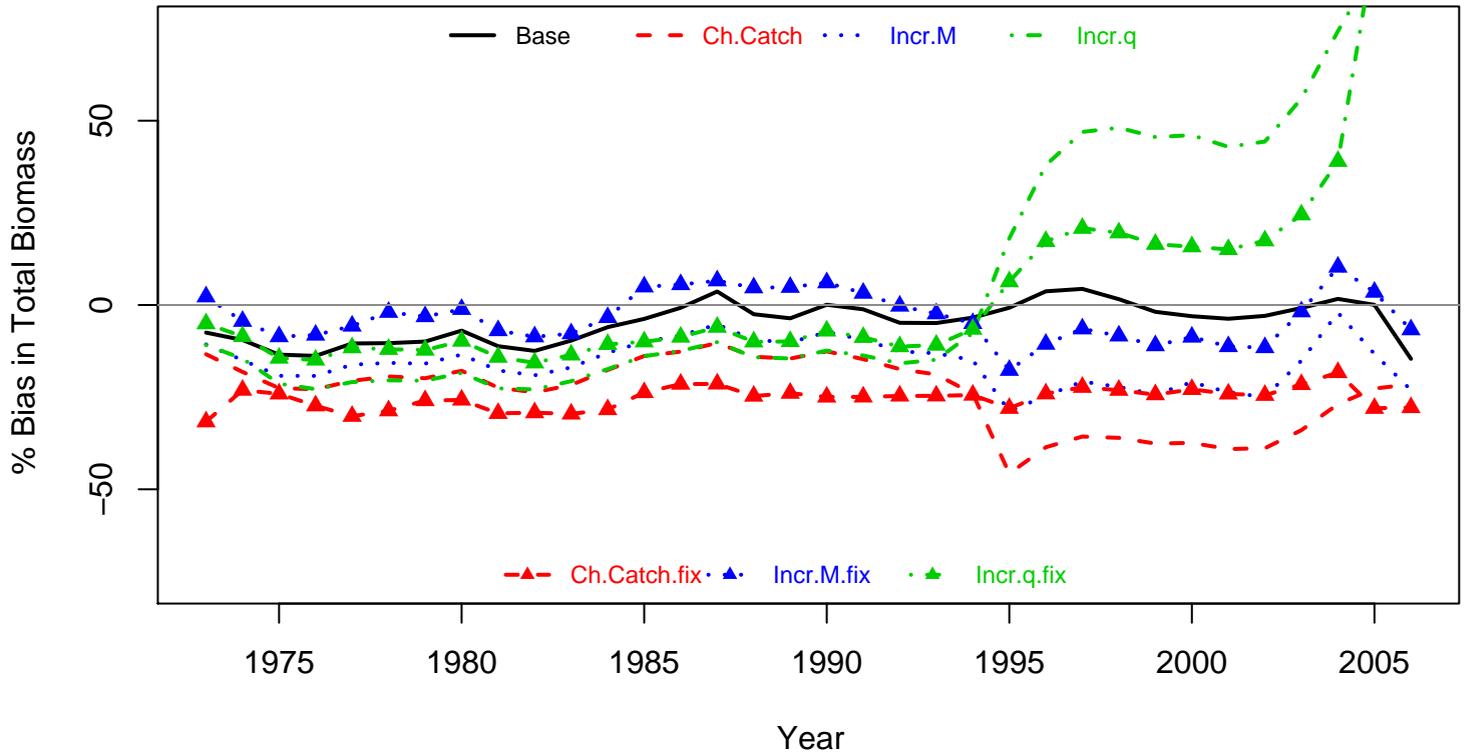


Figure GBYT.5 Bias in total biomass and F multiplier for all simulated data sets using Scale. ¹⁷

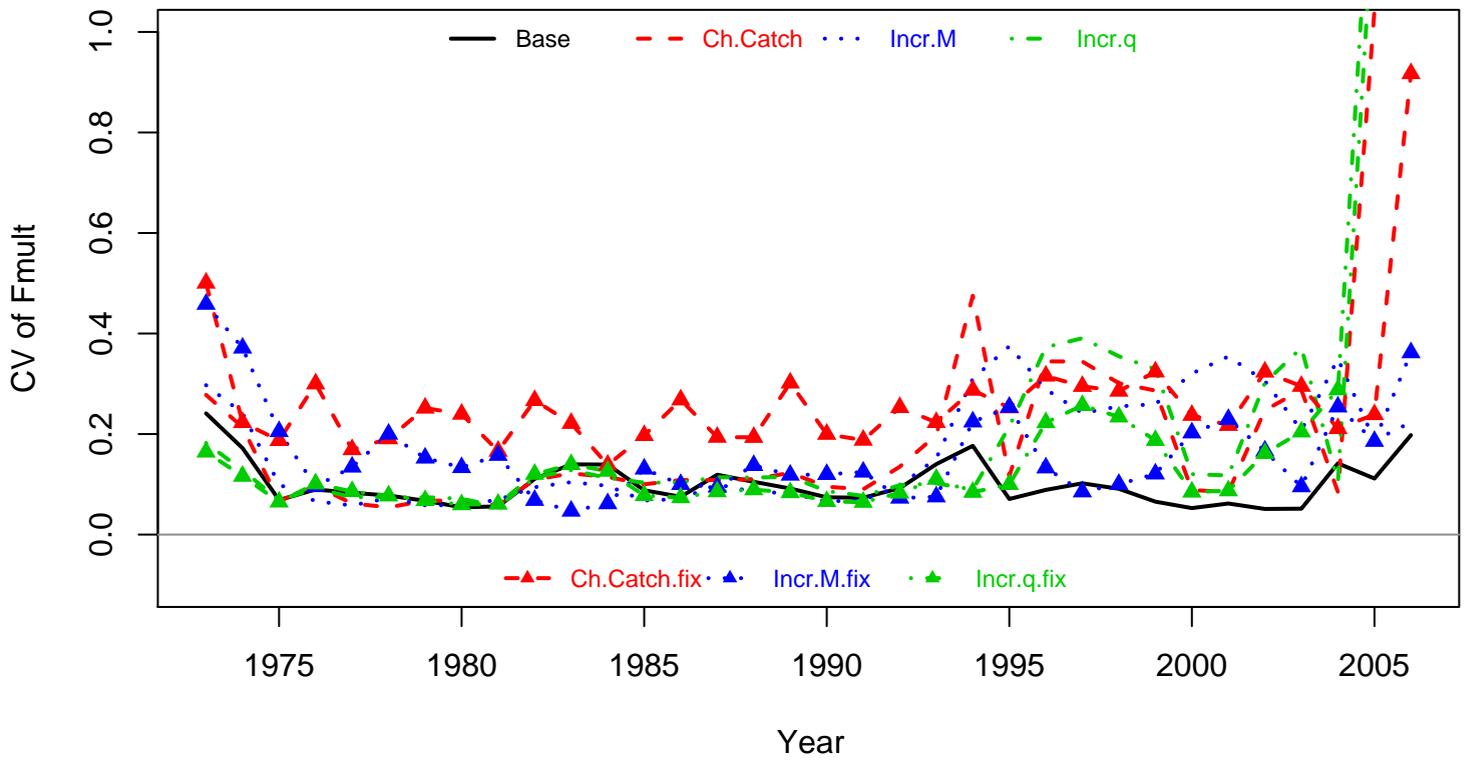
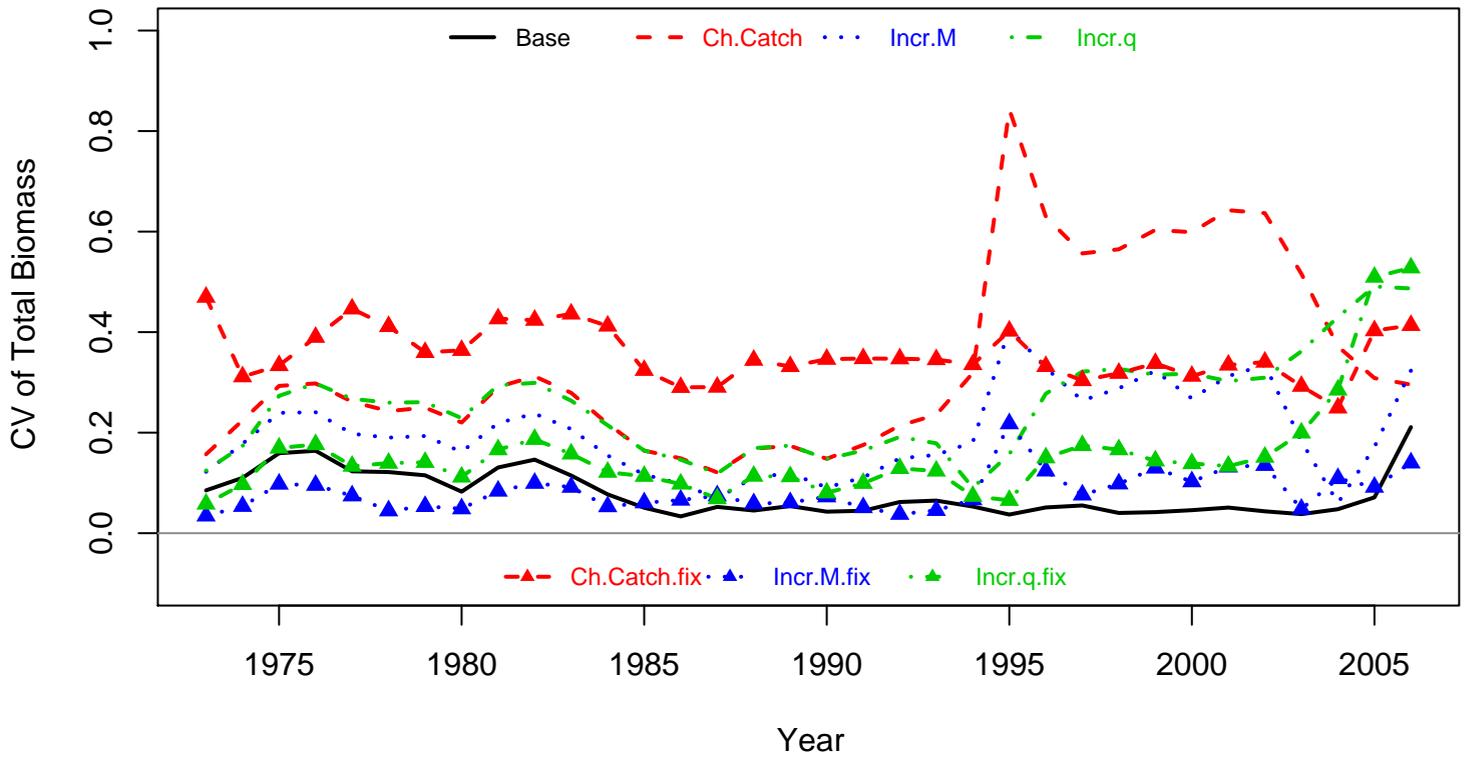


Figure GBYT.6 CV in total biomass and F multiplier for all simulated data sets using Scale. 18

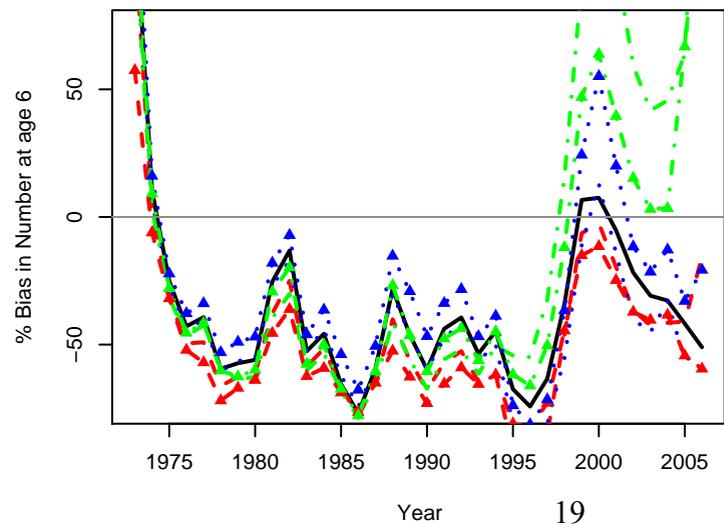
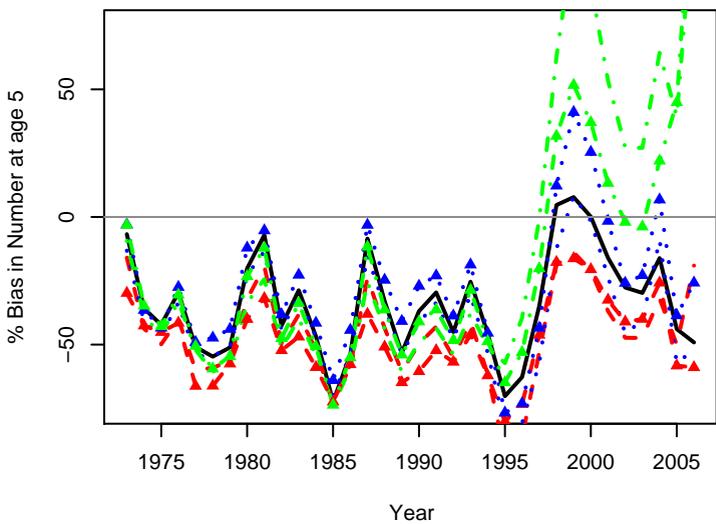
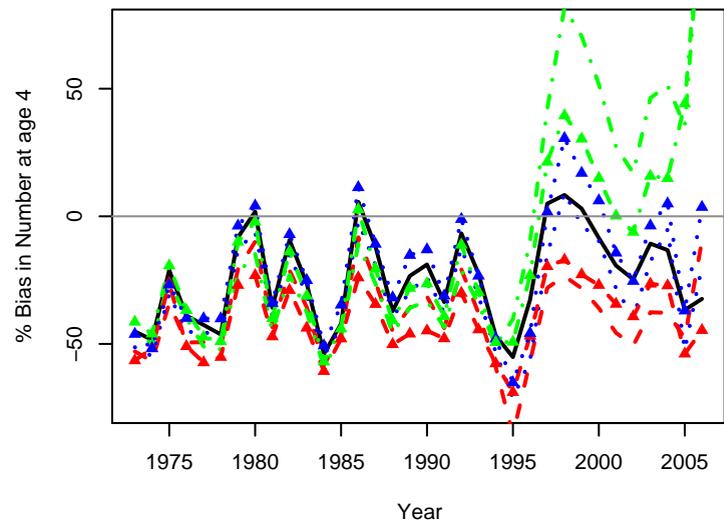
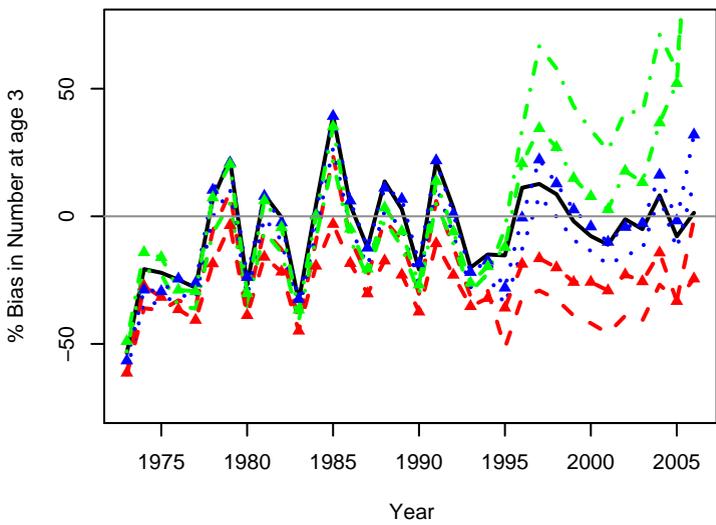
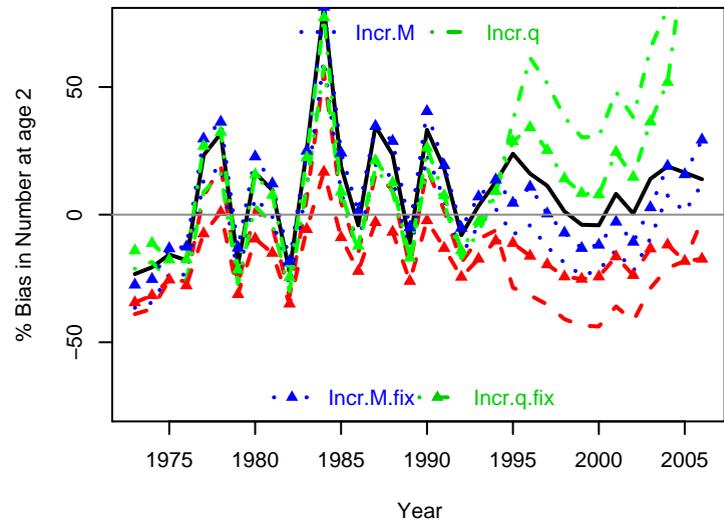
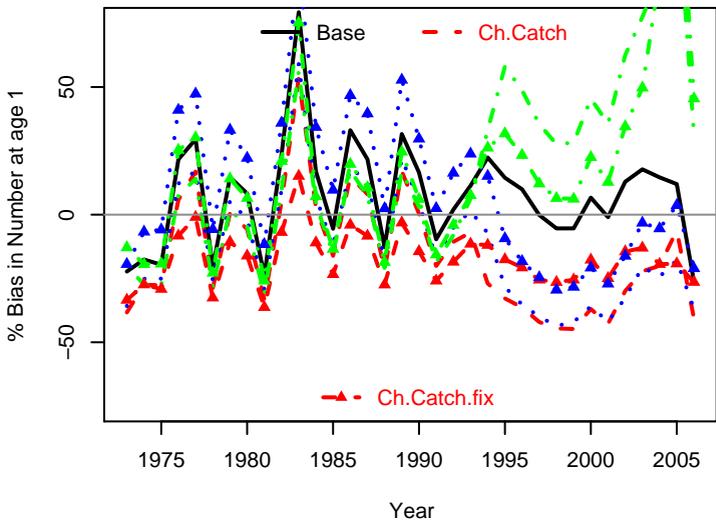


Figure GBYT.7 Bias in number at age for all simulated data sets using Scale.

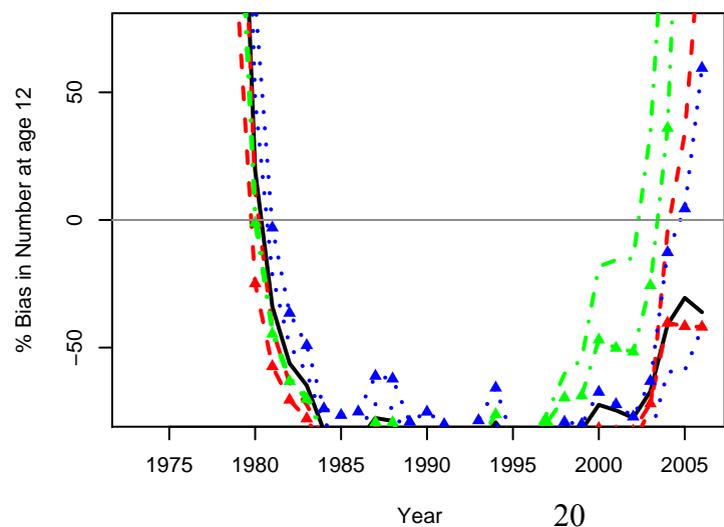
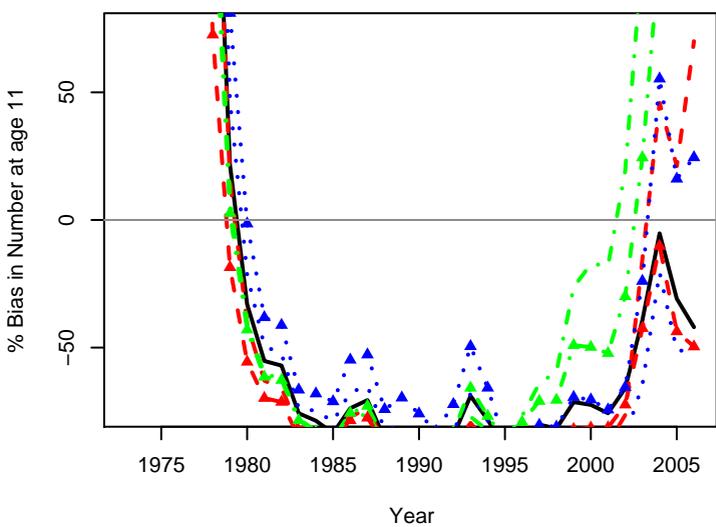
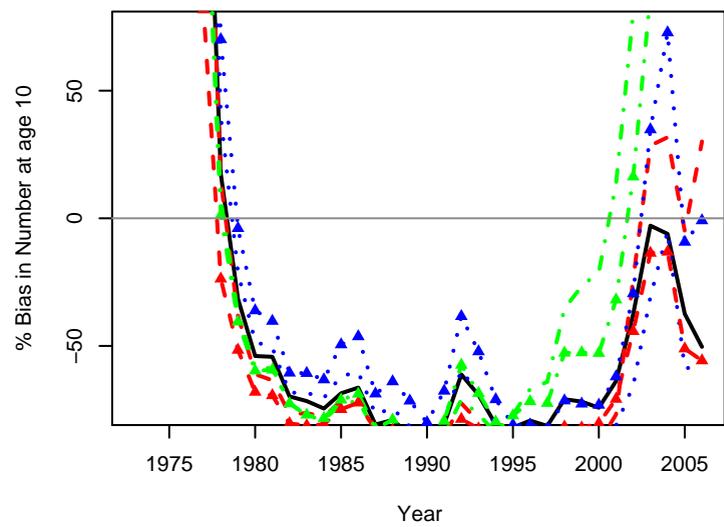
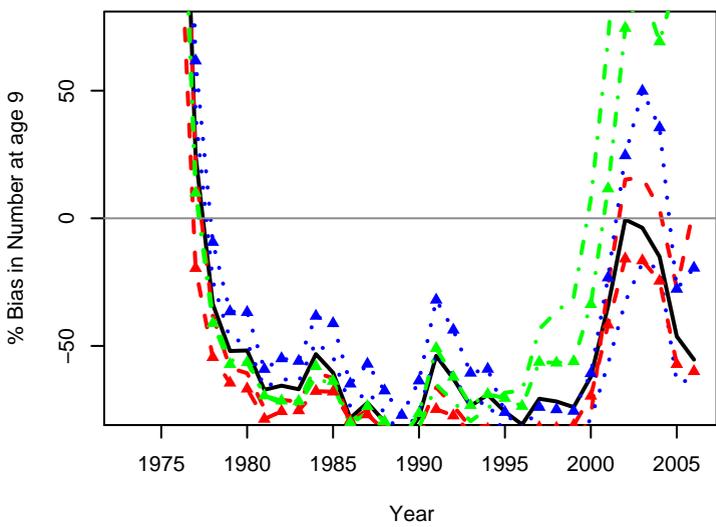
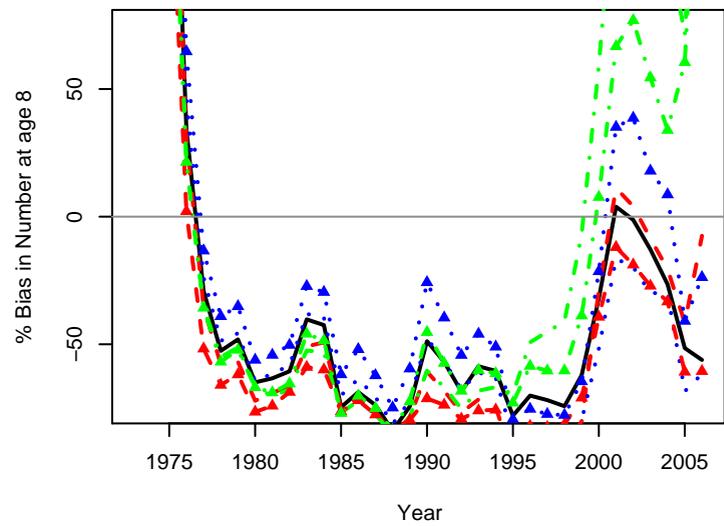
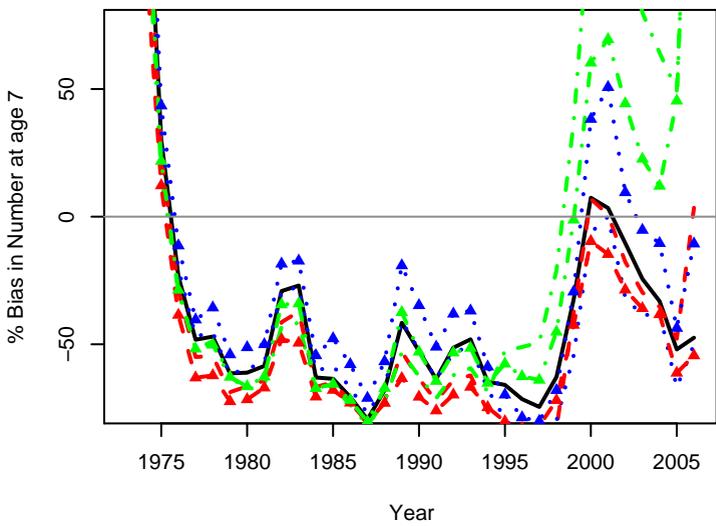
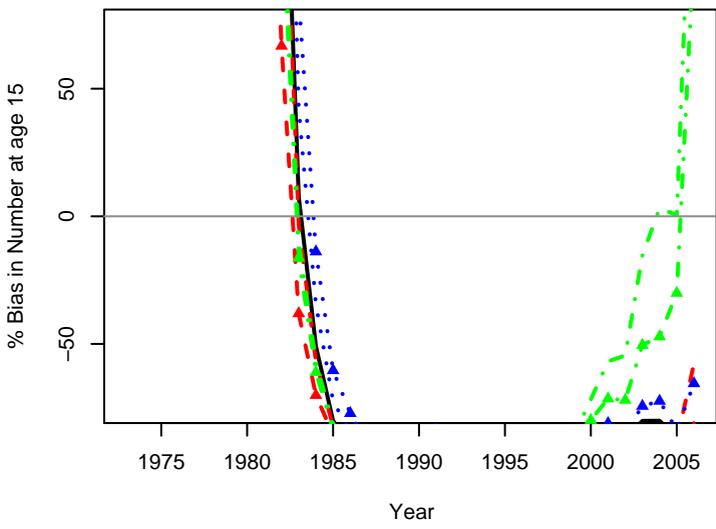
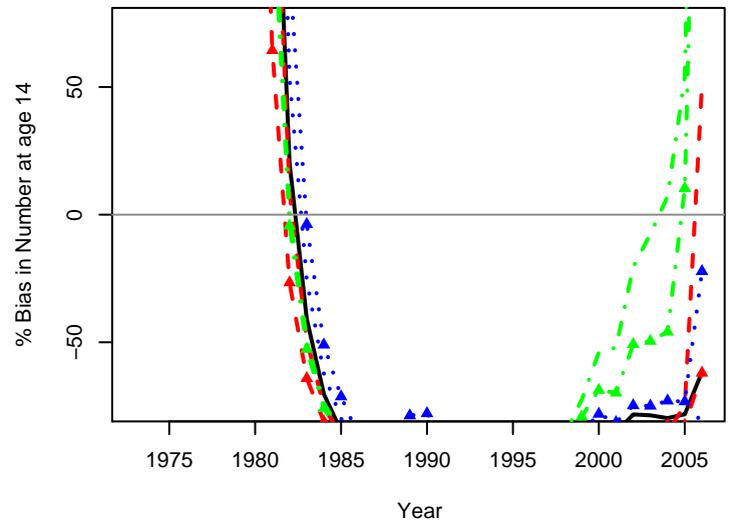
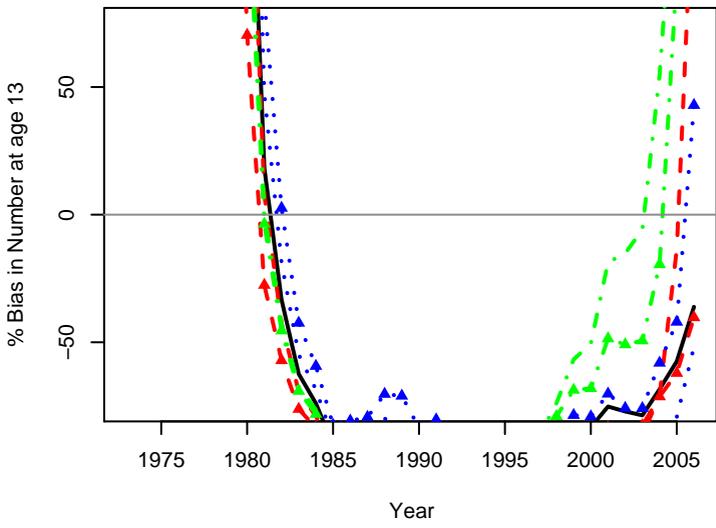


Figure GBYT.7 (cont.)



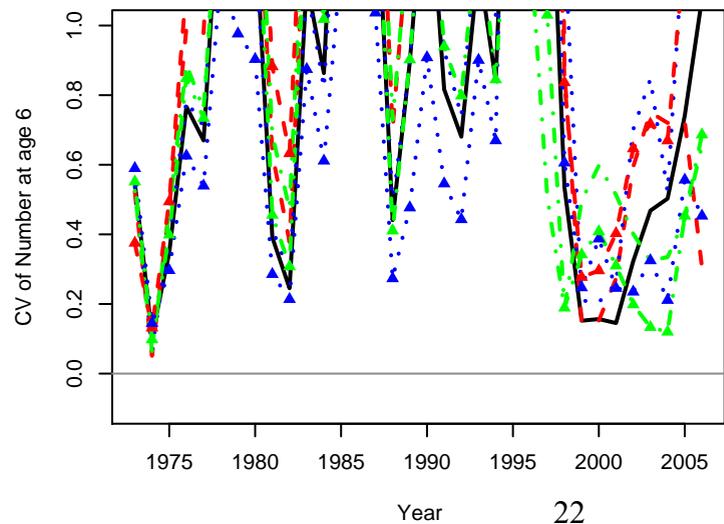
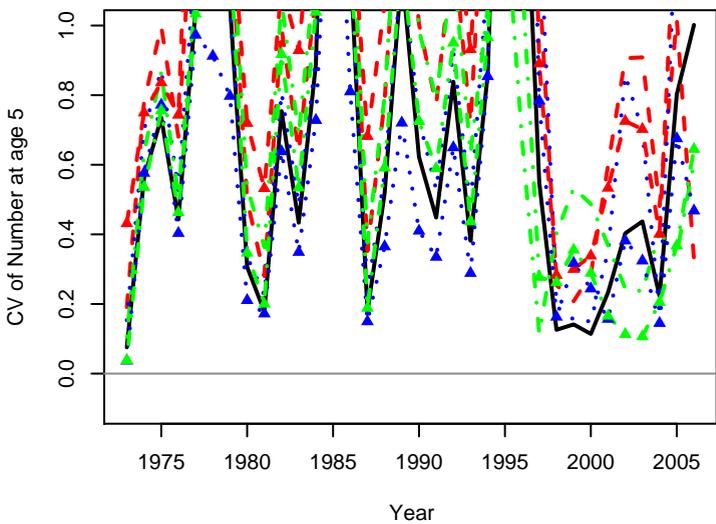
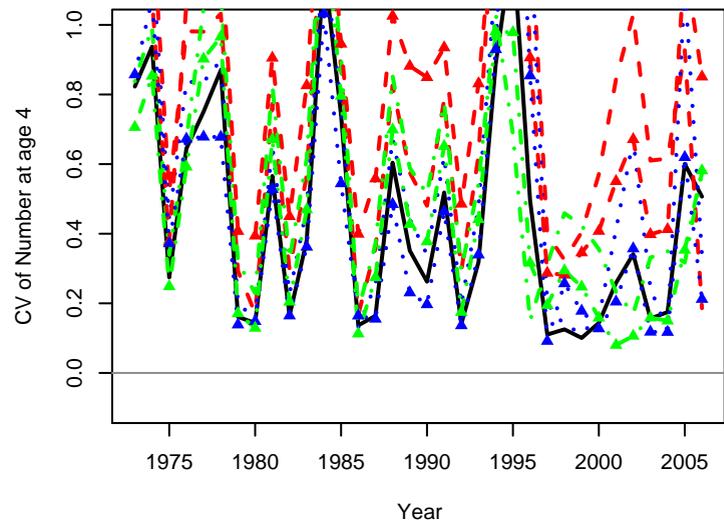
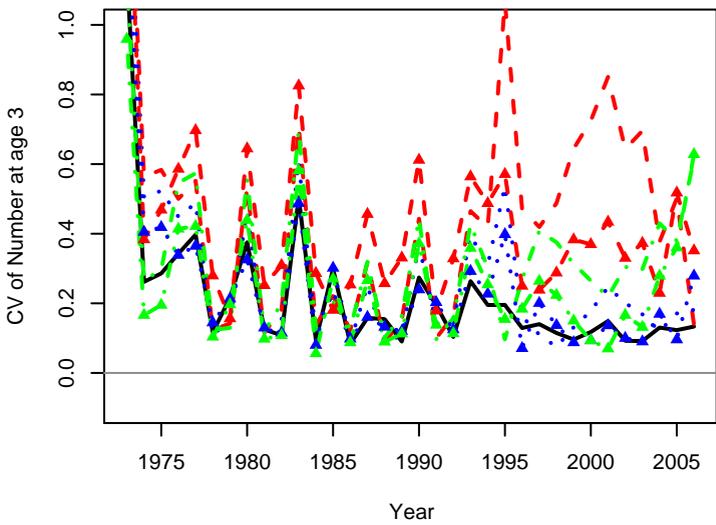
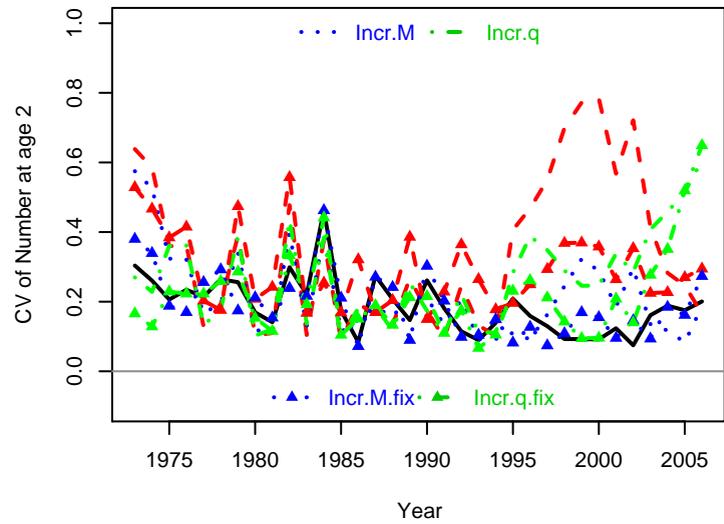
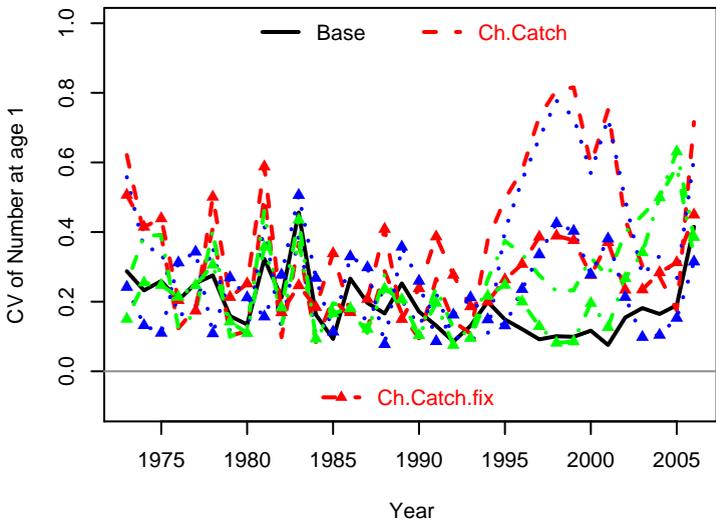


Figure GBYT.8 CV in number at age for all simulated data sets using Scale.

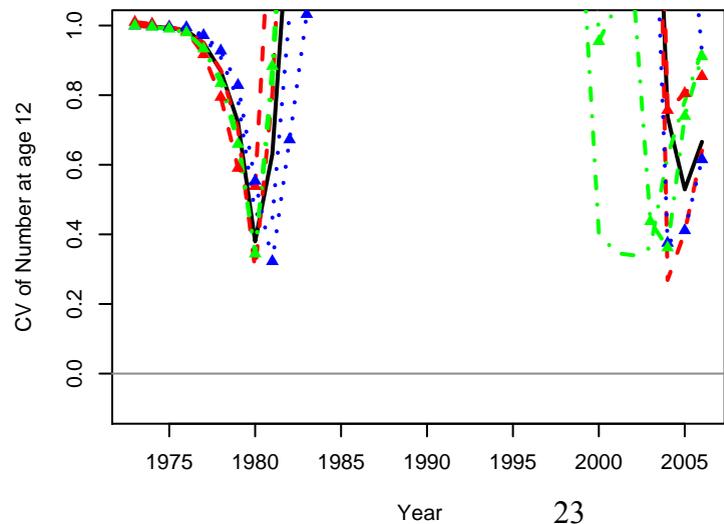
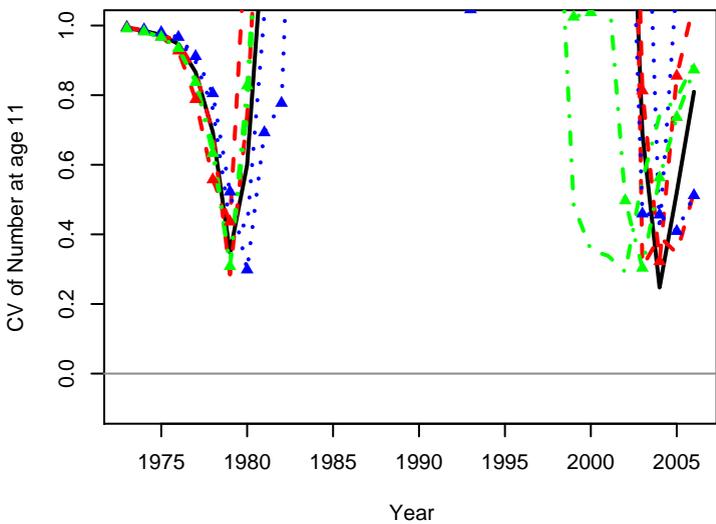
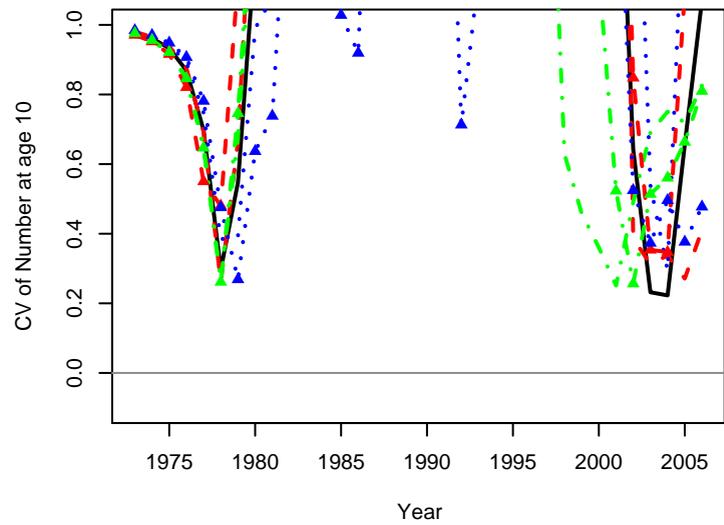
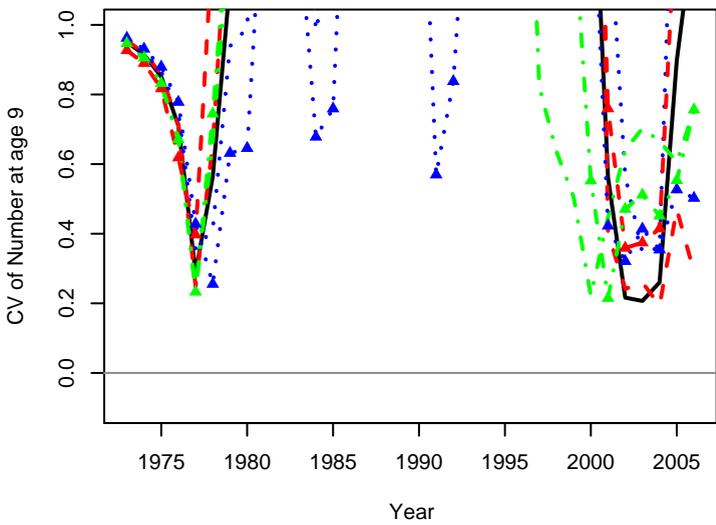
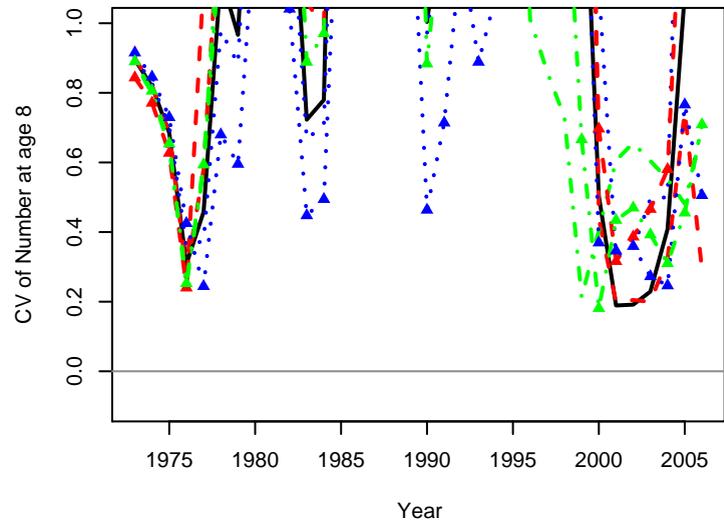
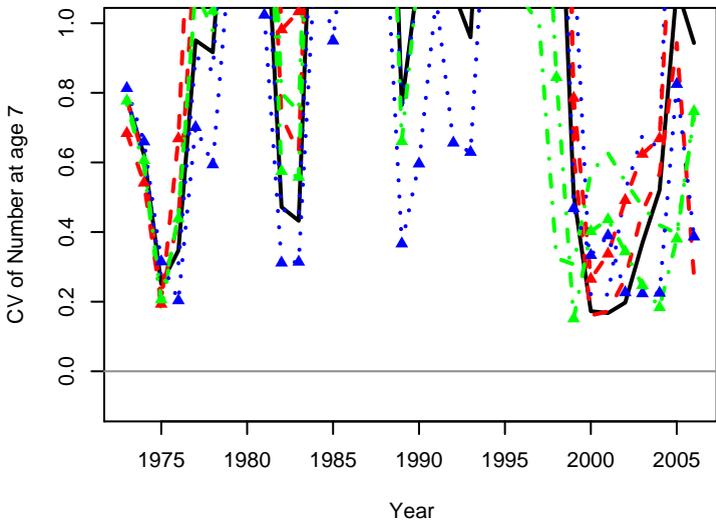


Figure GBYT.8 (cont.)

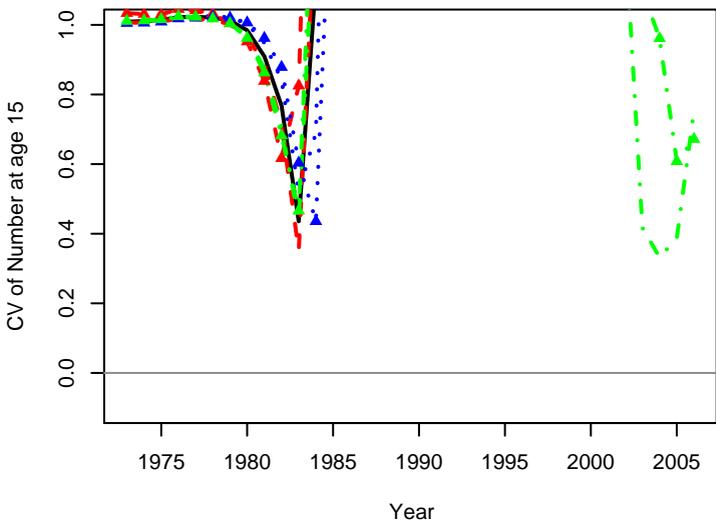
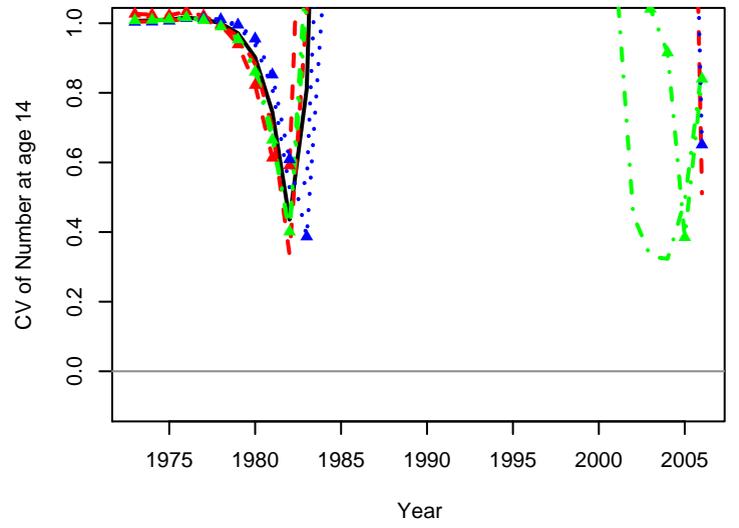
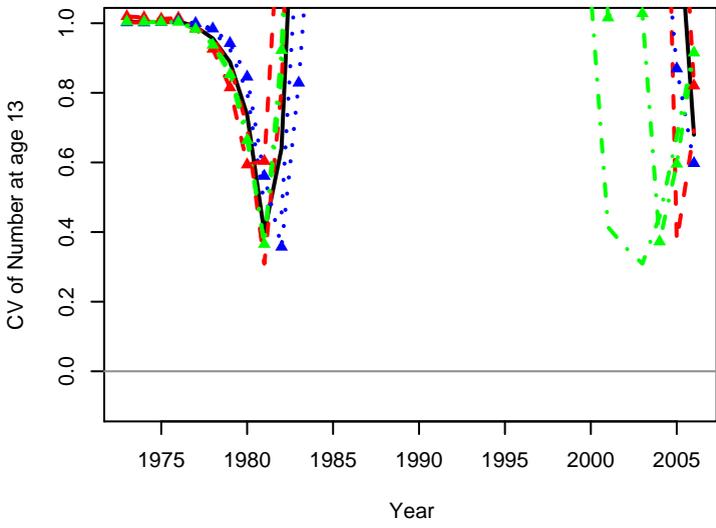


Figure GBYT.8 (cont.)

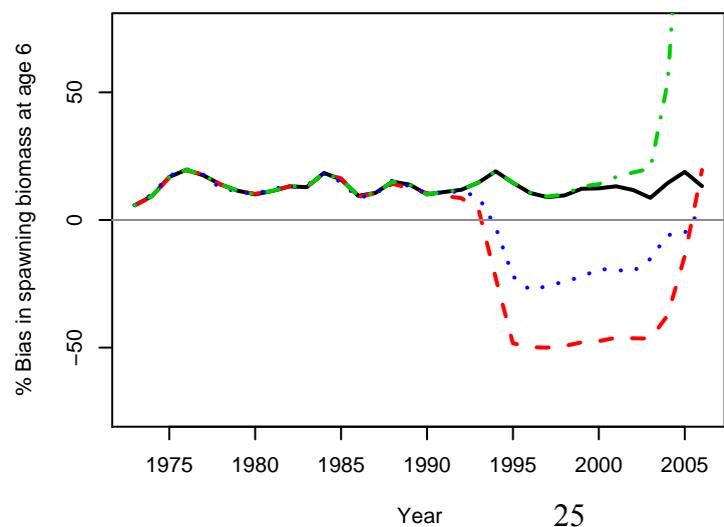
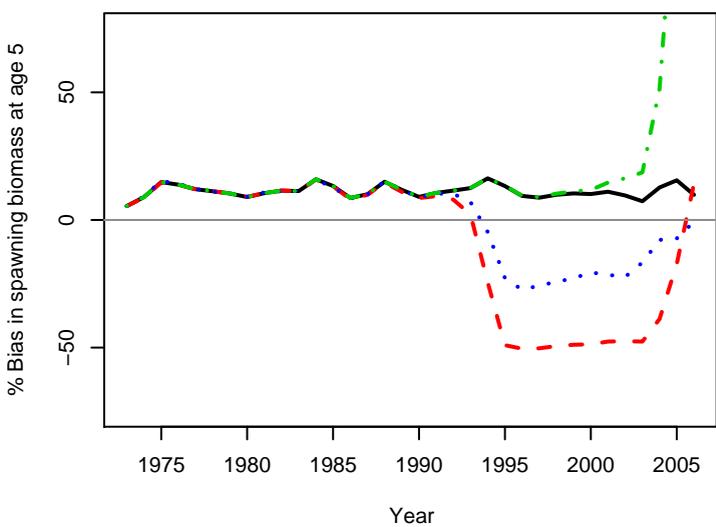
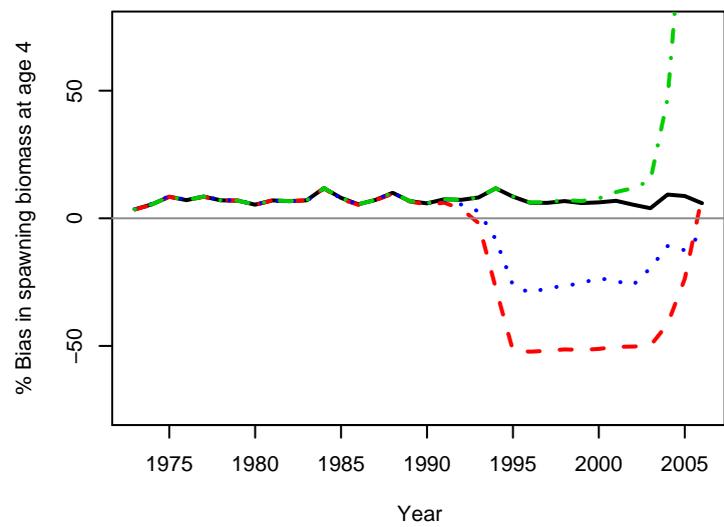
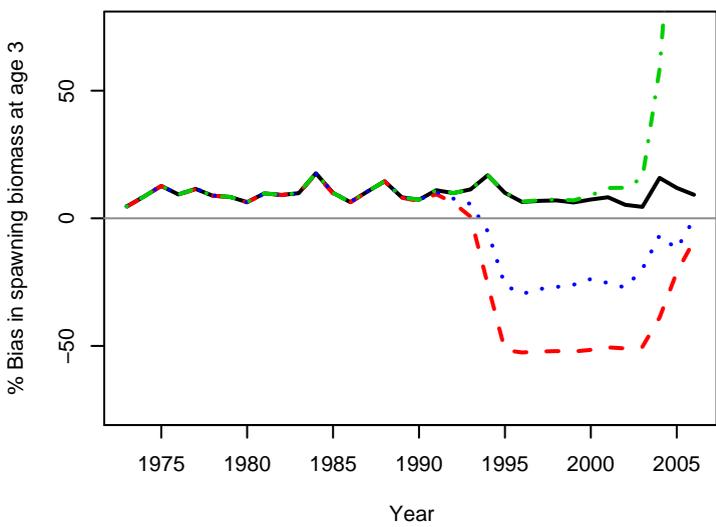
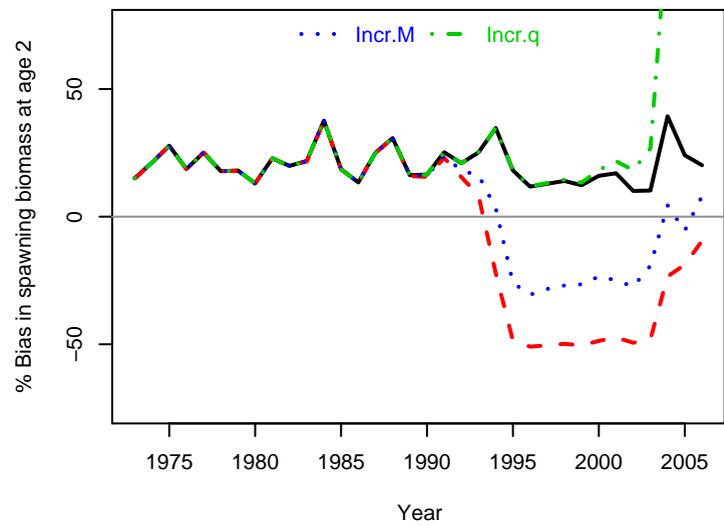
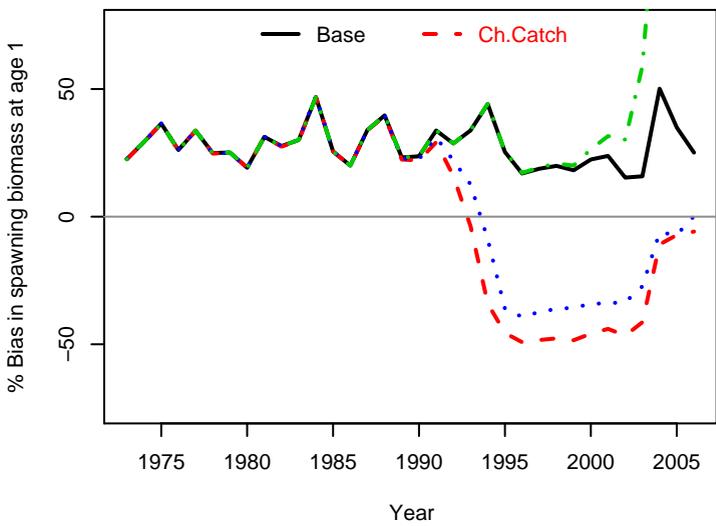


Figure GBYT.9 Bias in spawning biomass at age for all simulated data sets using VPA.

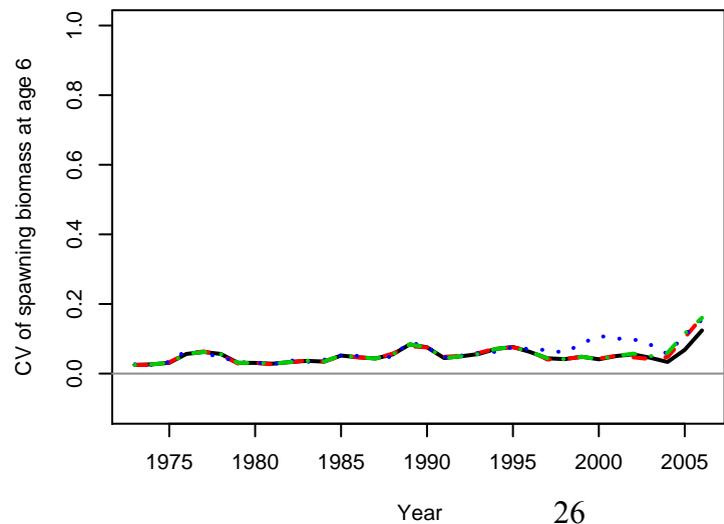
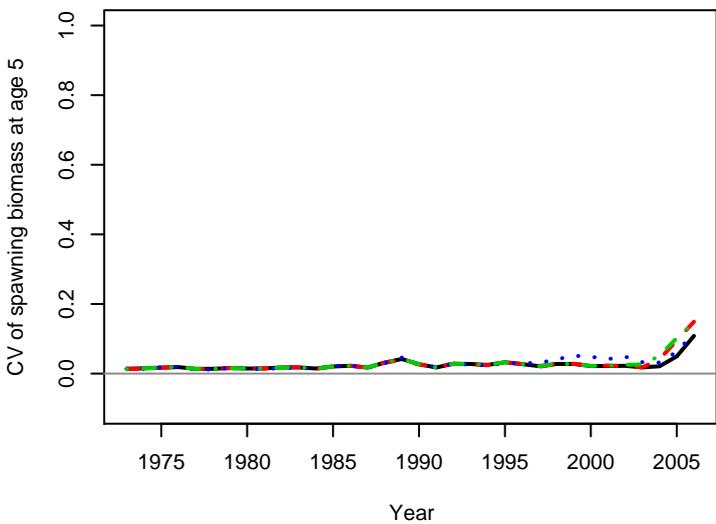
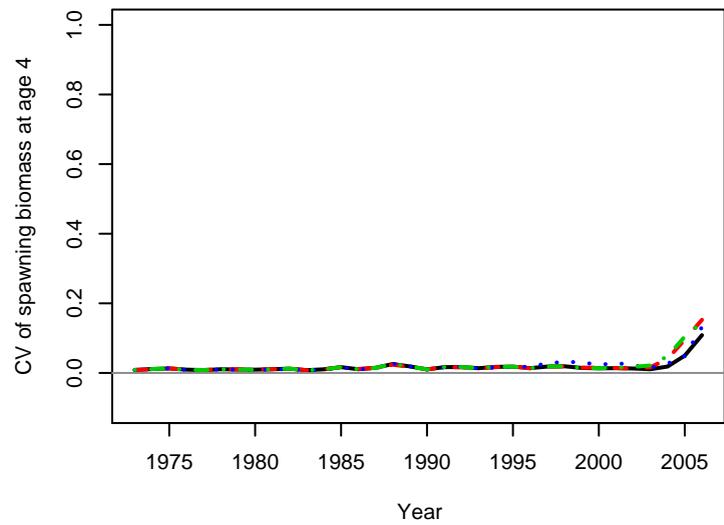
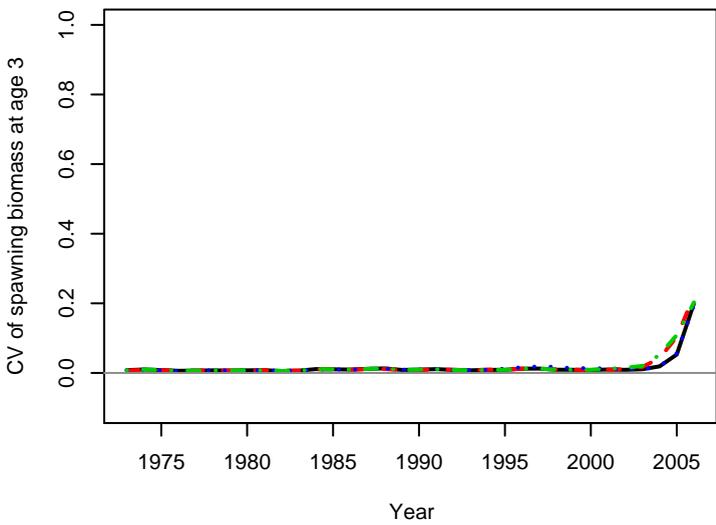
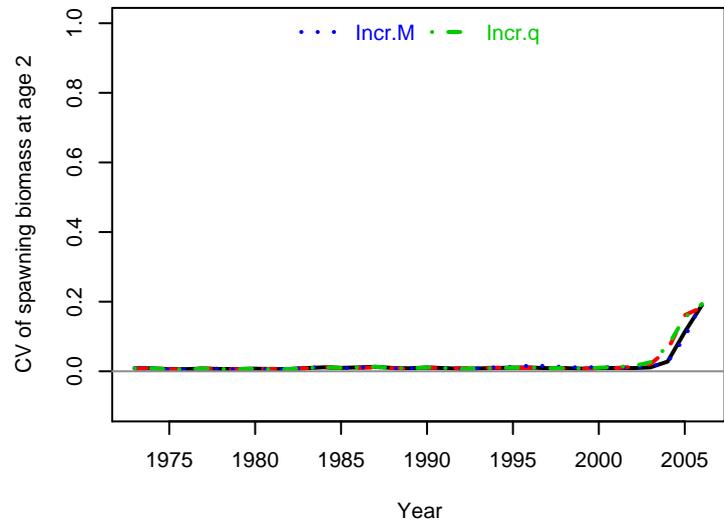
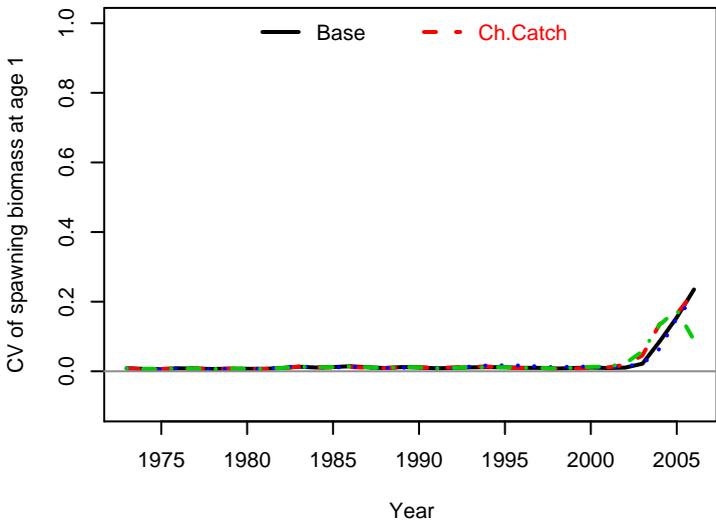


Figure GBYT.10 CV of spawning biomass at age for all simulated data sets using VPA.

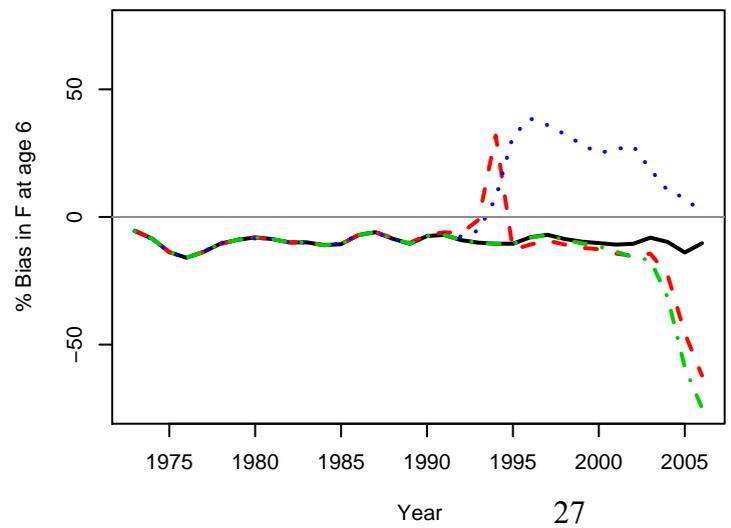
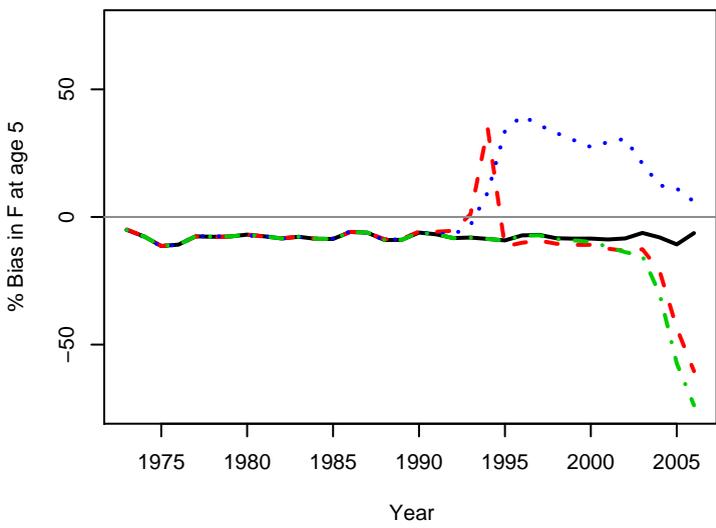
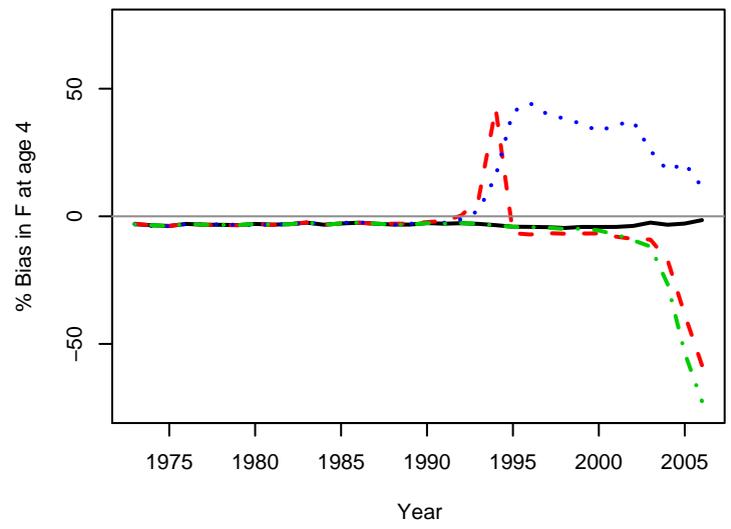
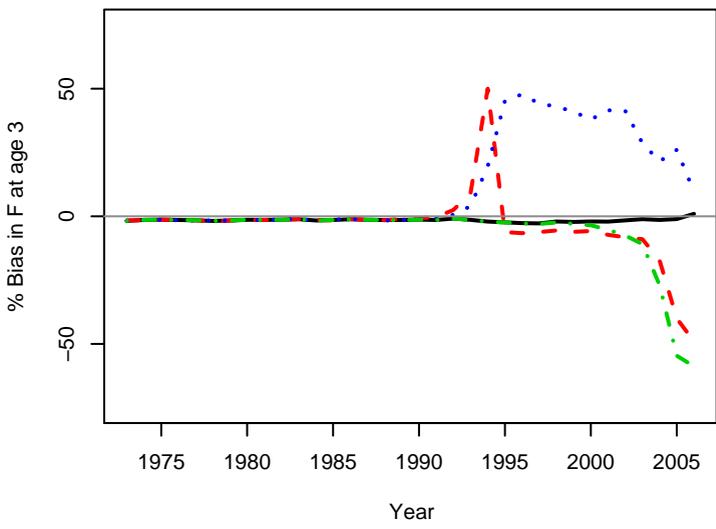
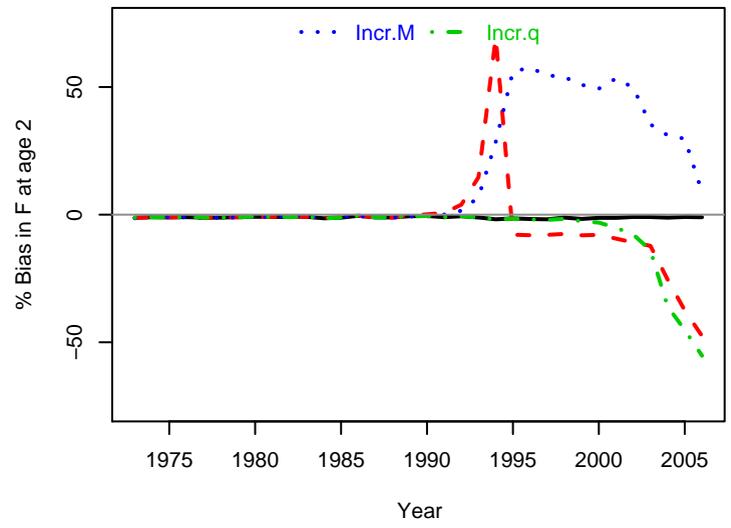
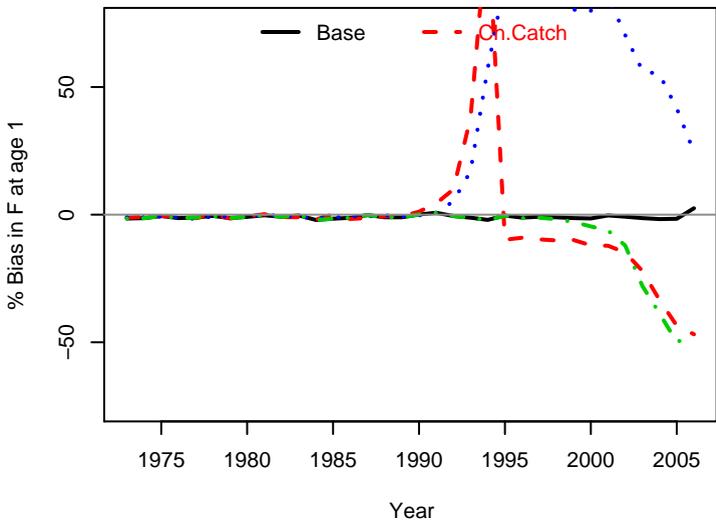


Figure GBYT.11 Bias in F at age for all simulated data sets using VPA.

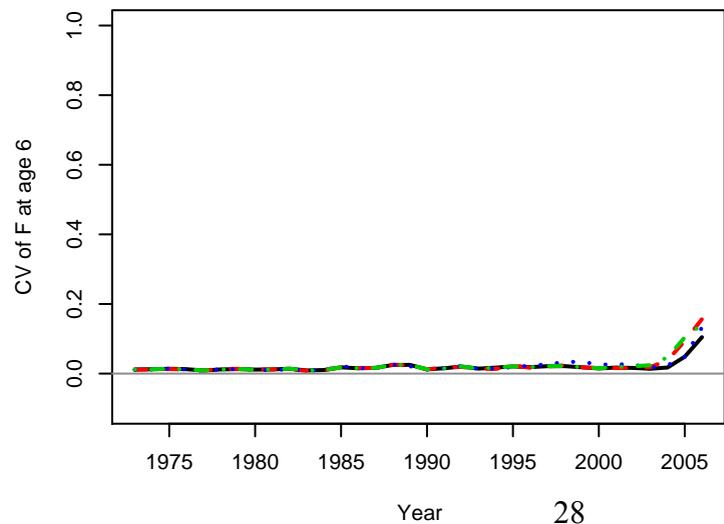
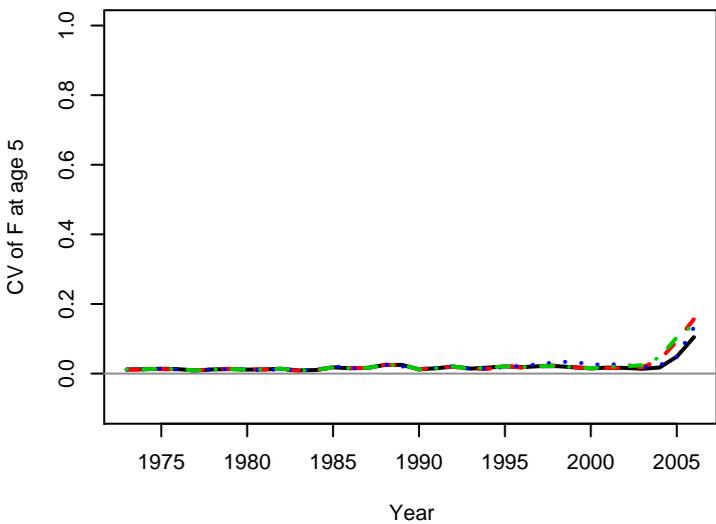
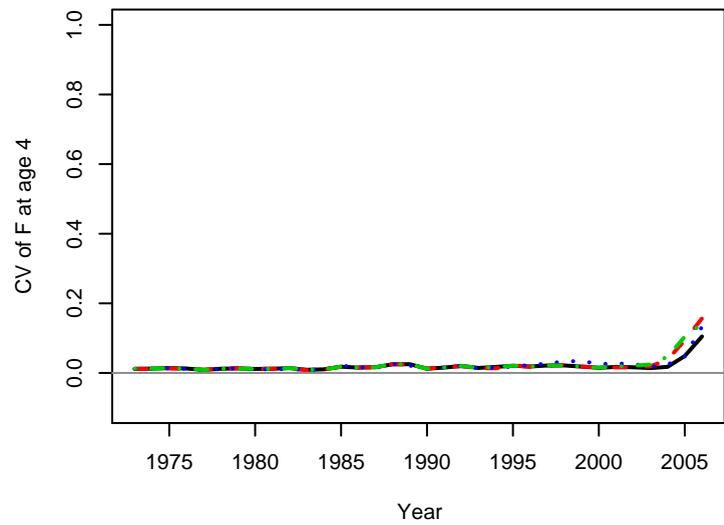
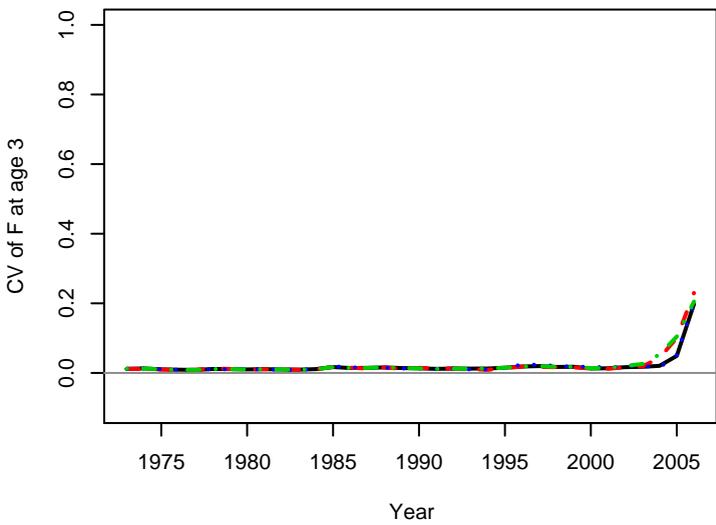
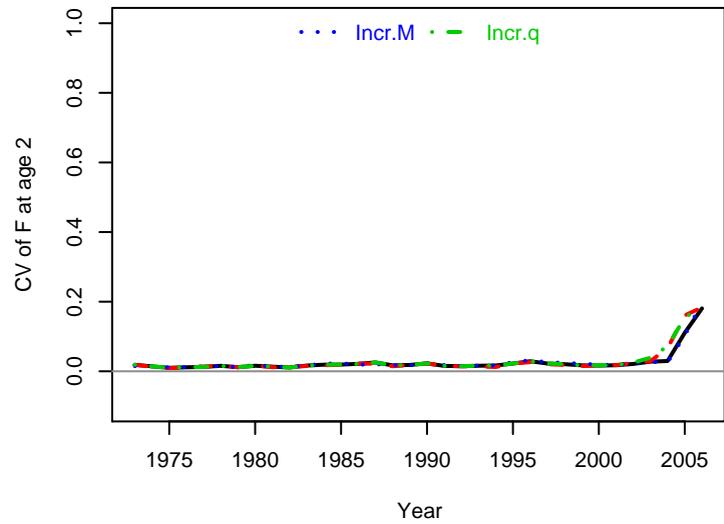
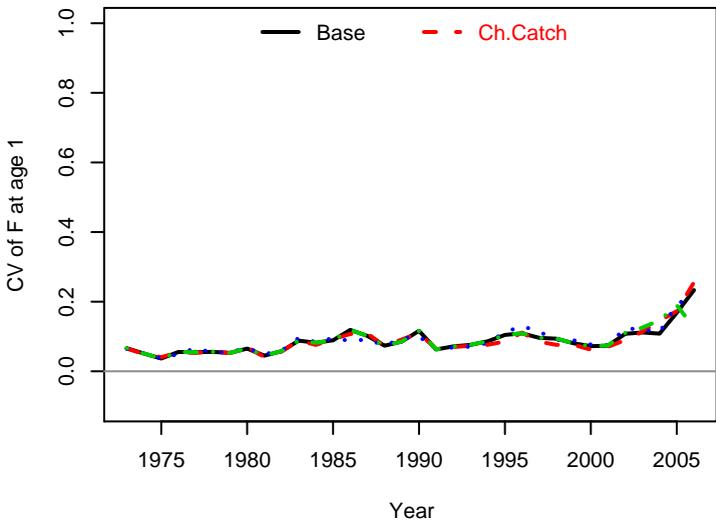


Figure GBYT.12 CV of F at age for all simulated data sets using VPA.

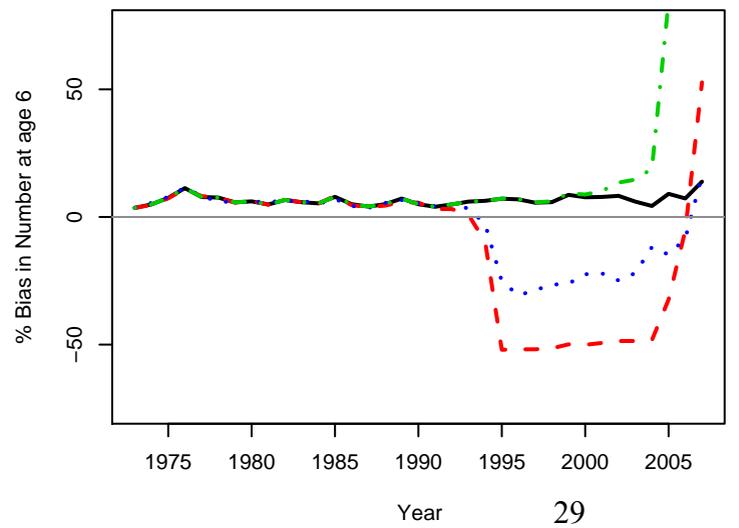
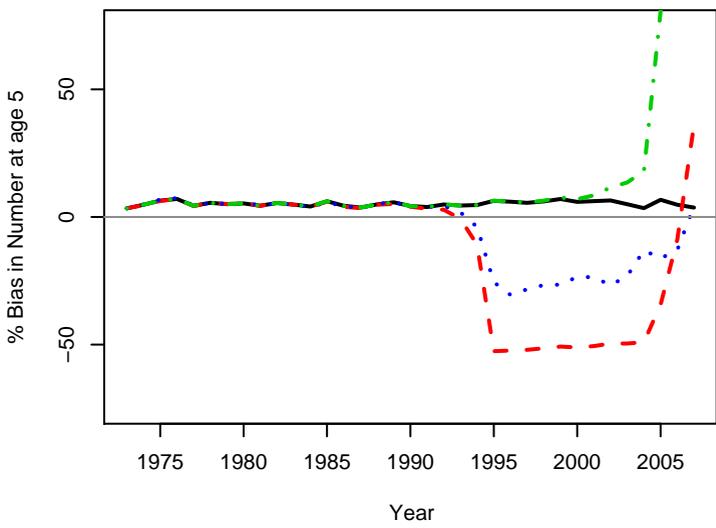
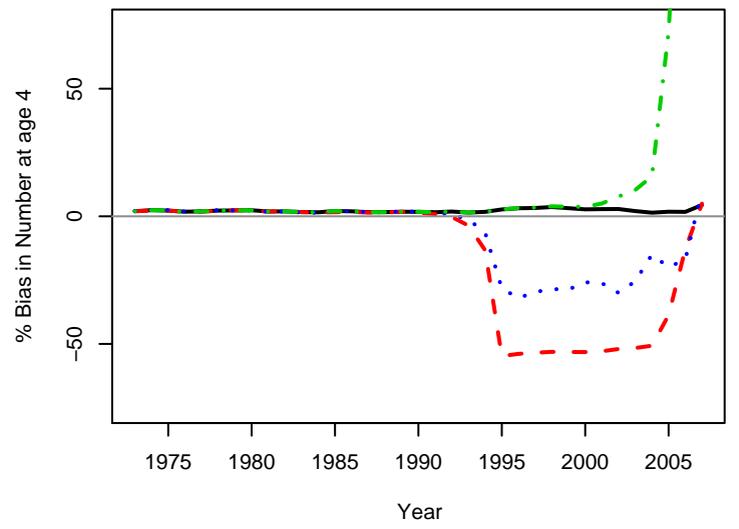
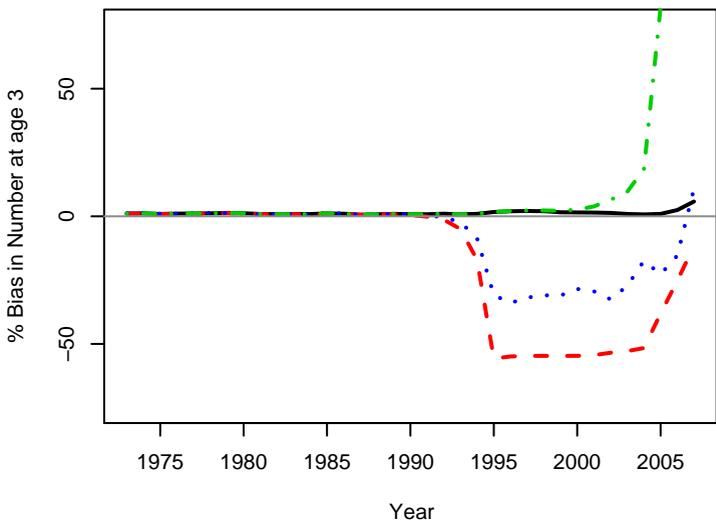
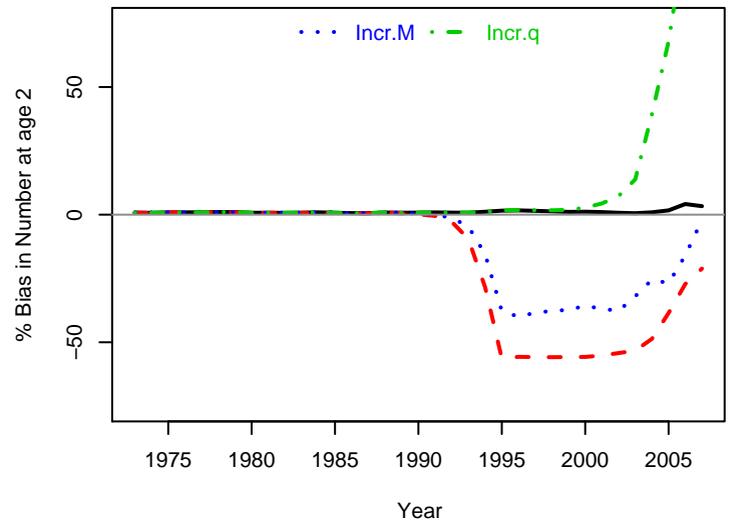
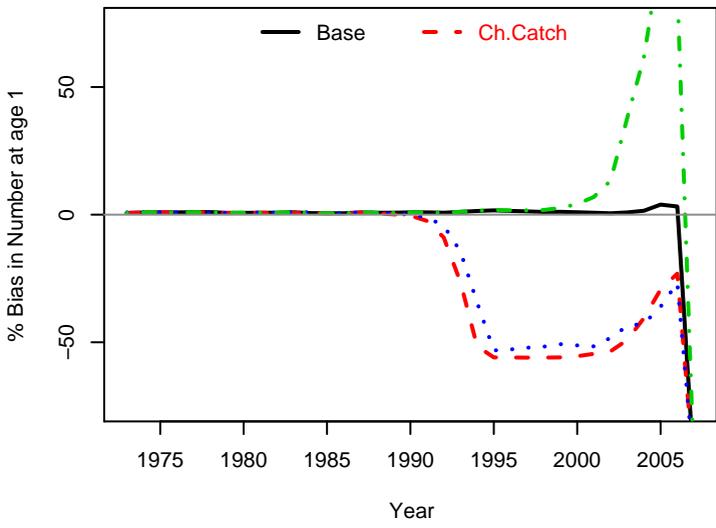


Figure GBYT.13 Bias in number at age for all simulated data sets using VPA.

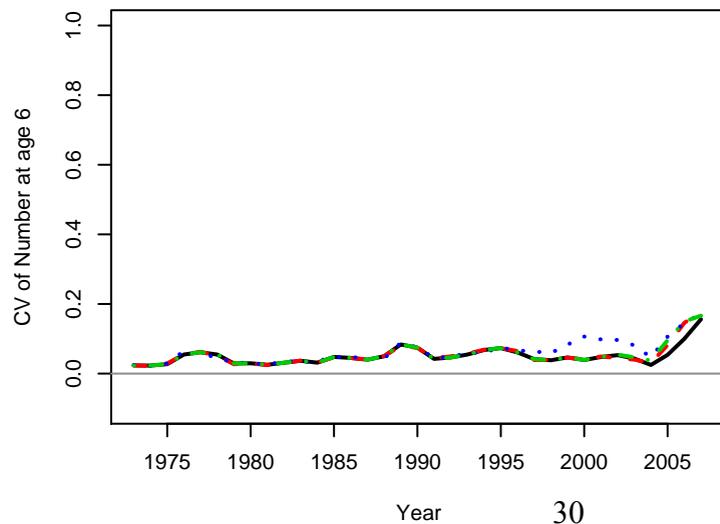
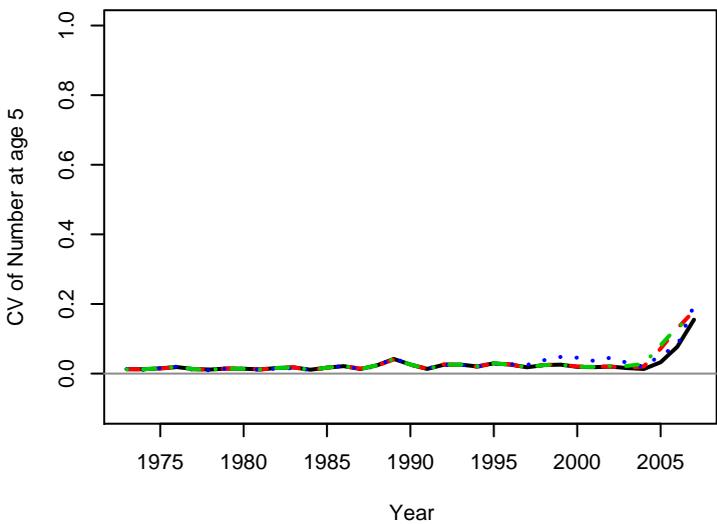
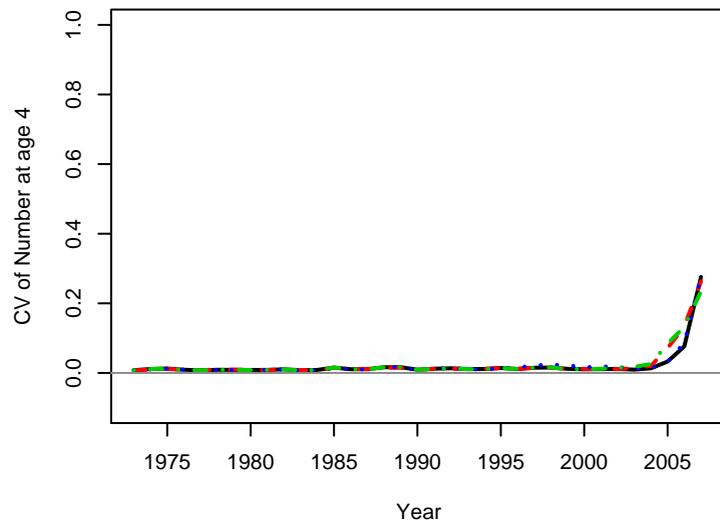
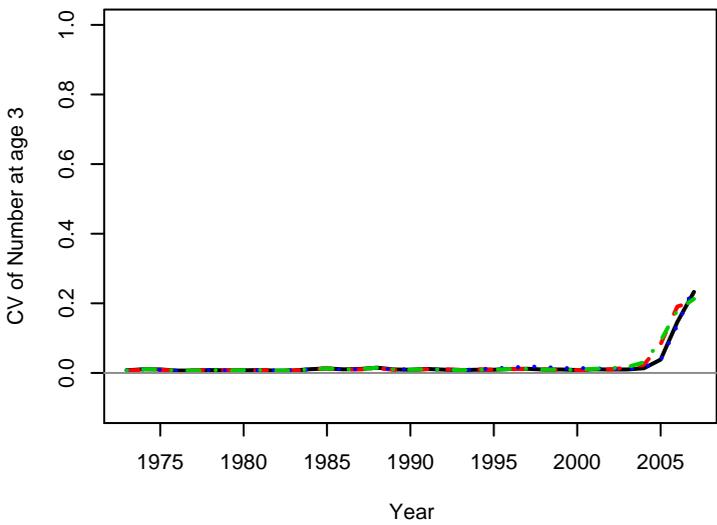
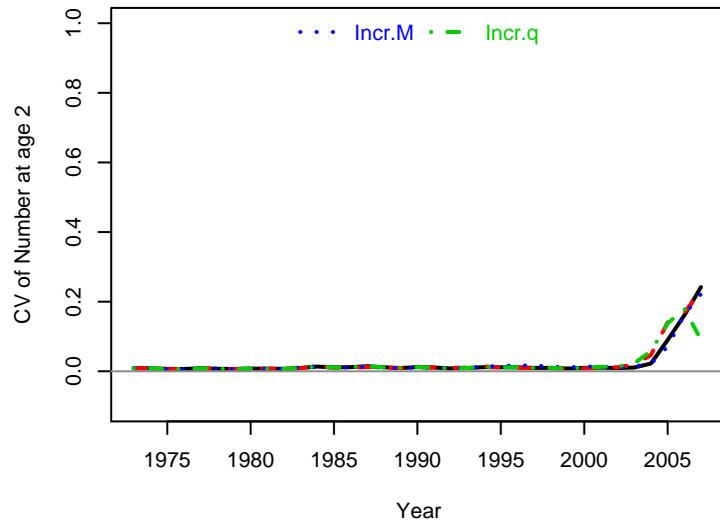
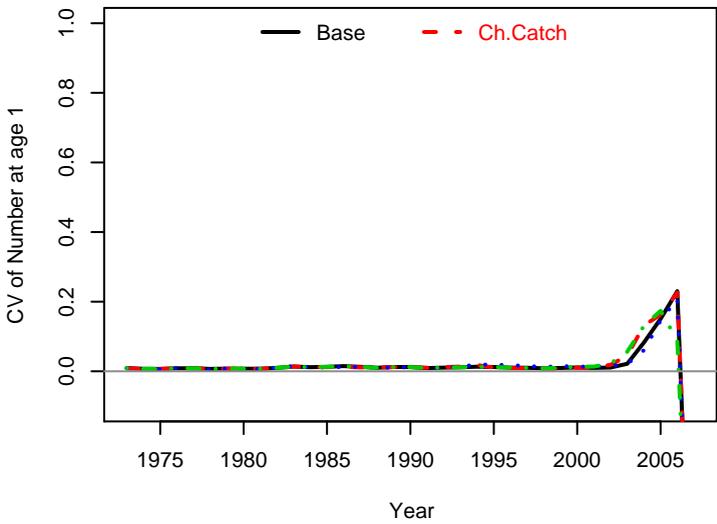


Figure GBYT.14 CV of number at age for all simulated data sets using VPA.

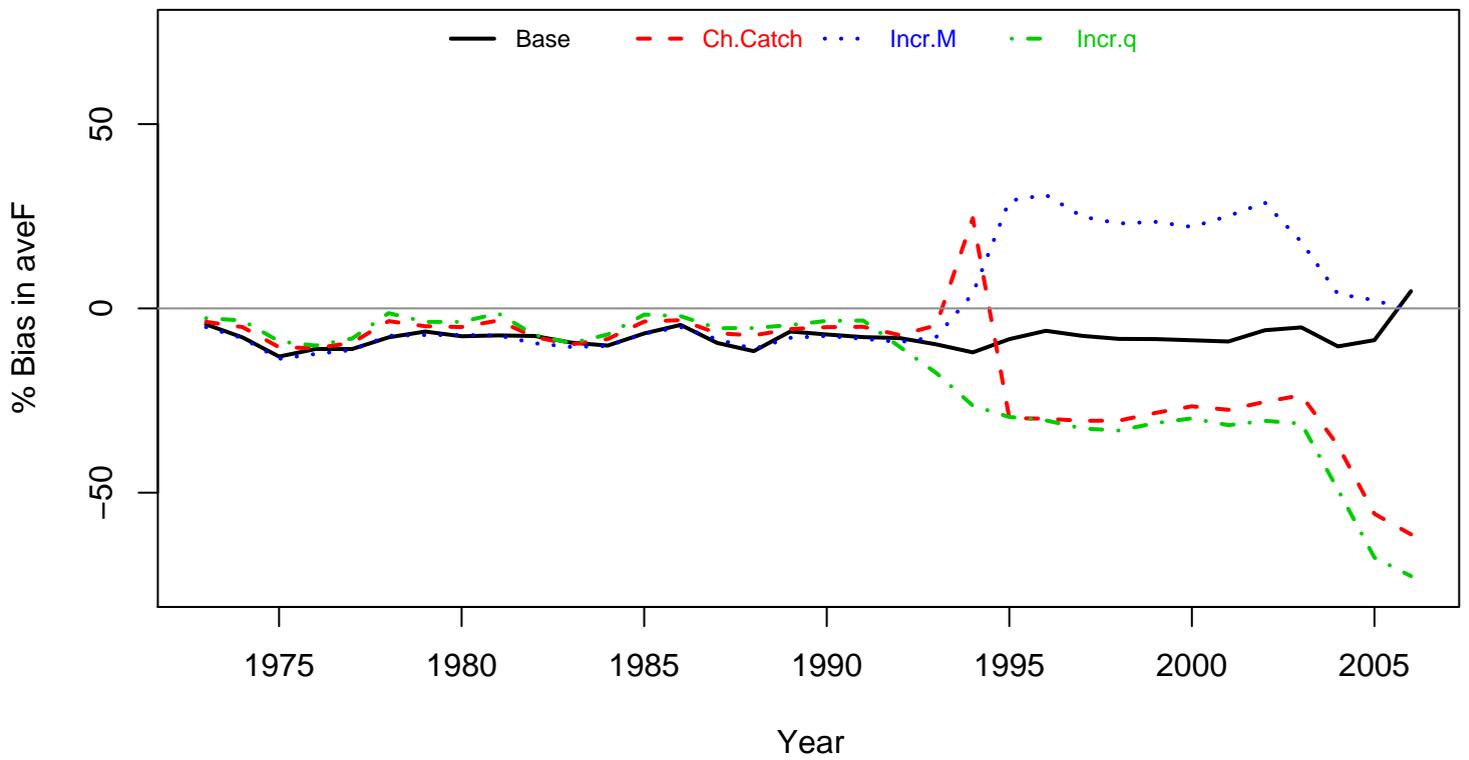
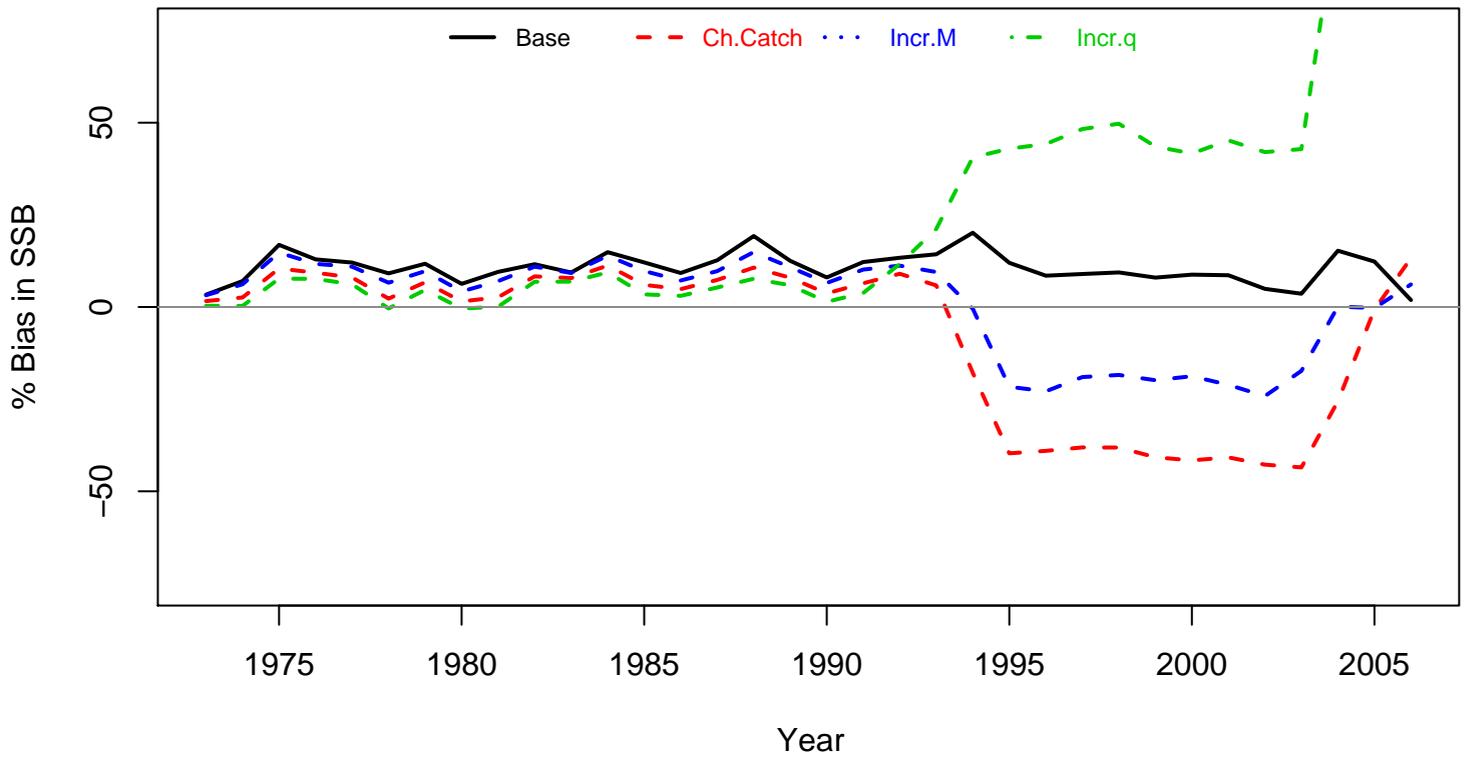


Figure GBYT.15 Bias in spawning biomass and average F for all simulated data sets using ASAP. ³¹

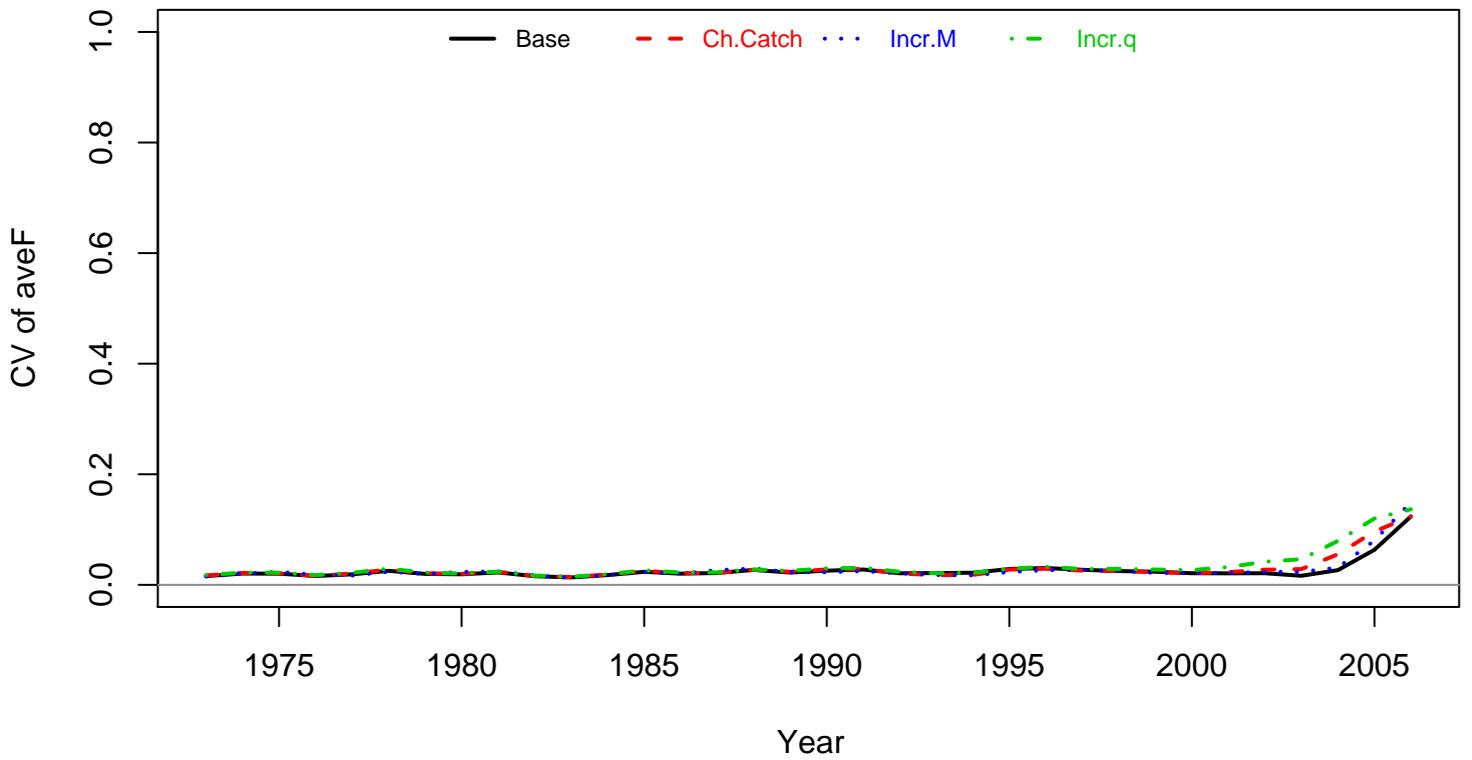
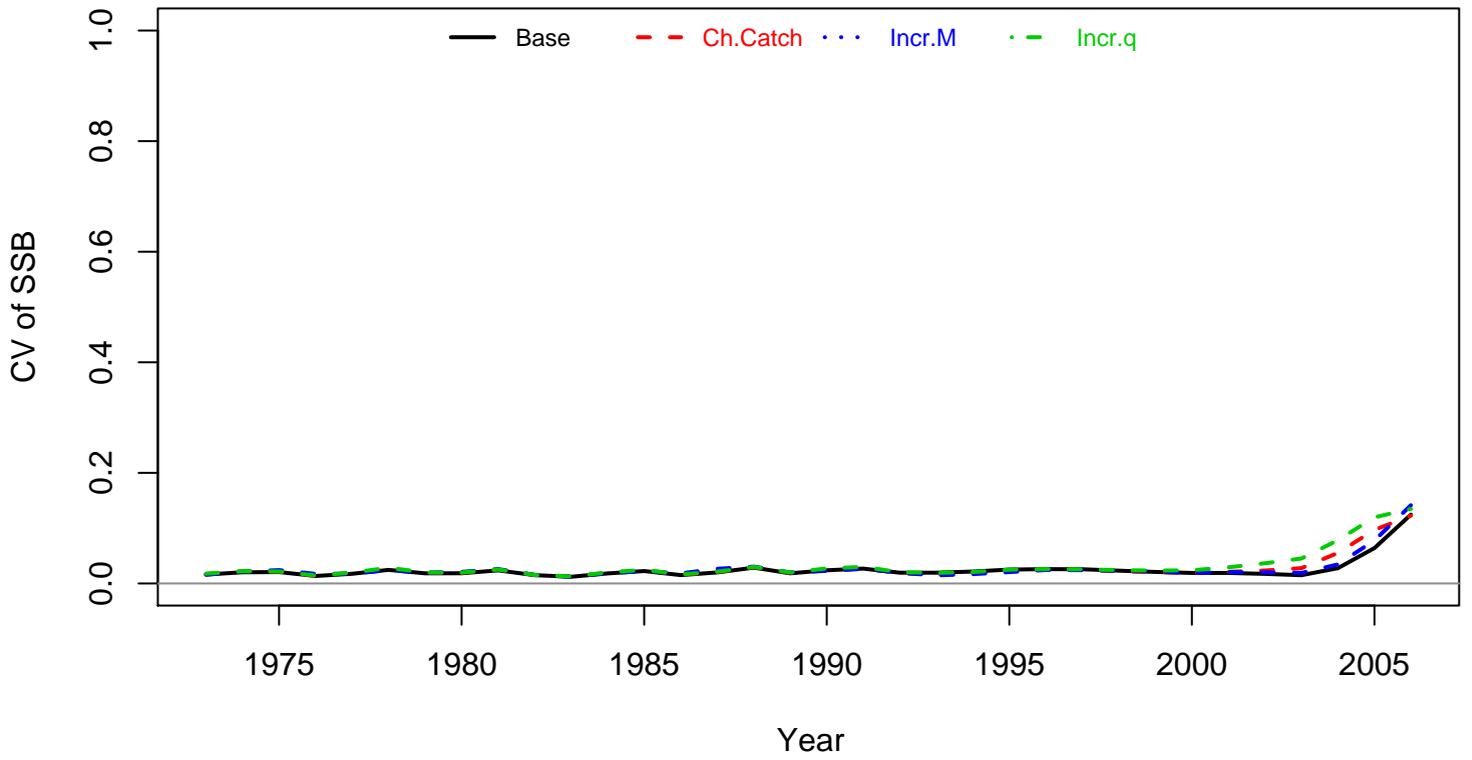


Figure GBYT.16 CV of spawning biomass and average F for all simulated data sets using ASAP.³²

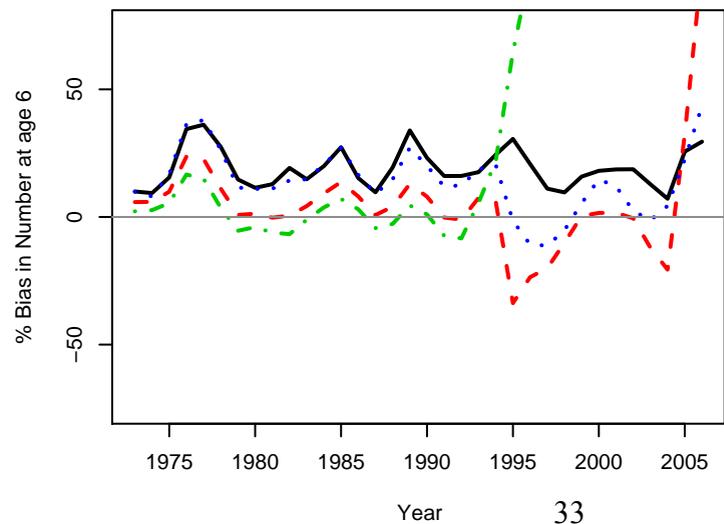
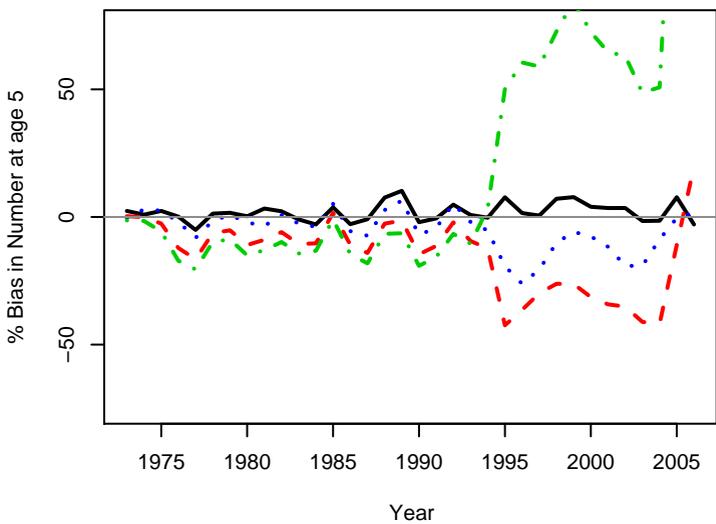
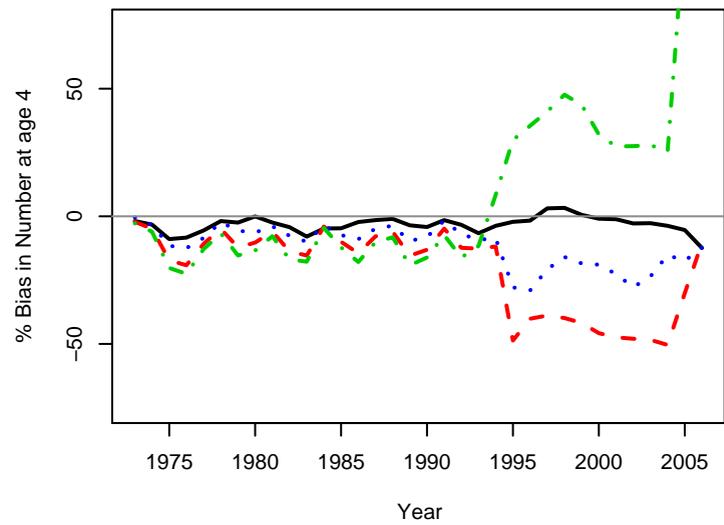
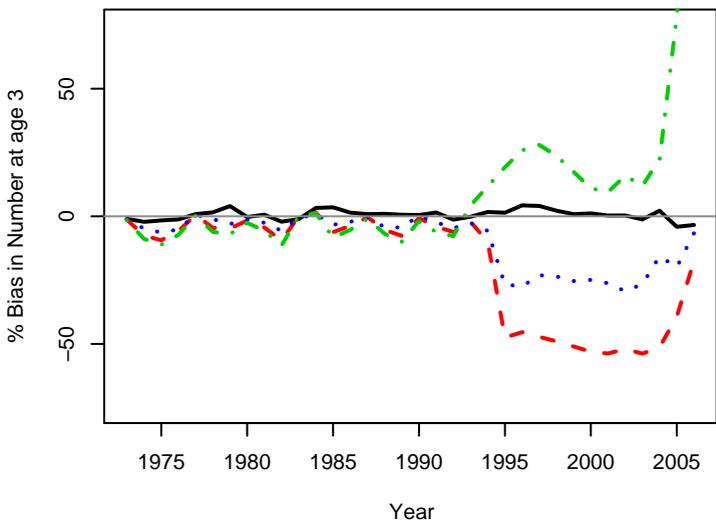
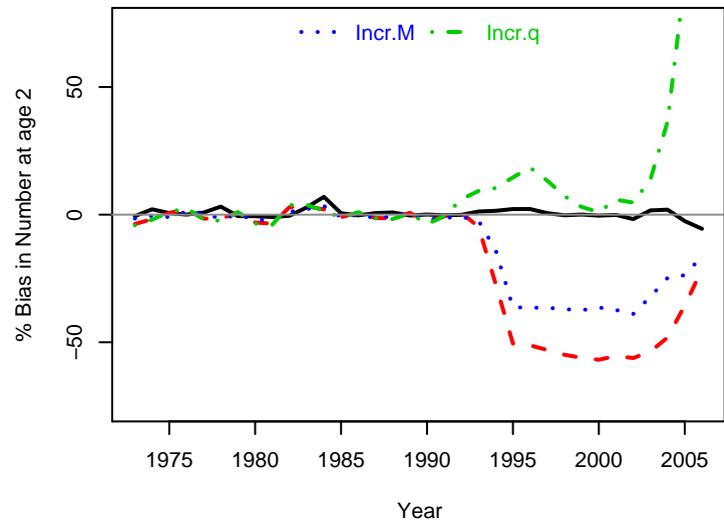
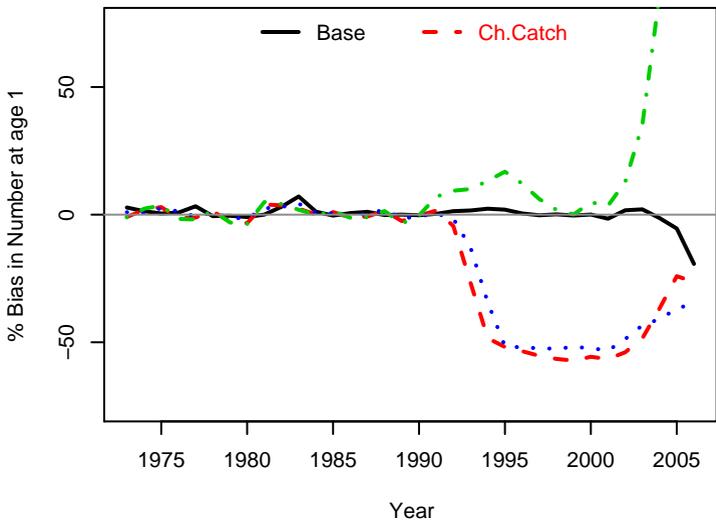


Figure GBYT.17 Bias in number at age for all simulated data sets using ASAP.

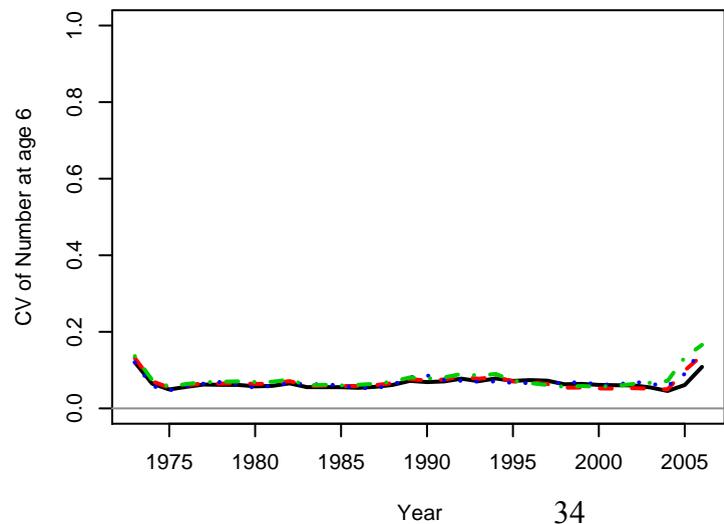
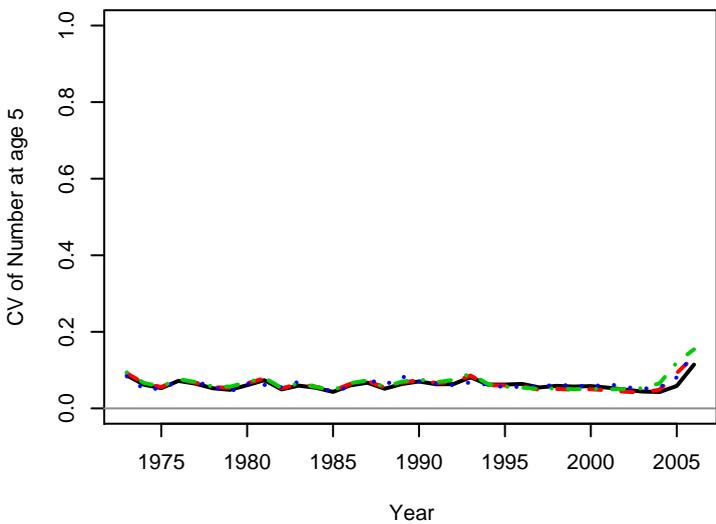
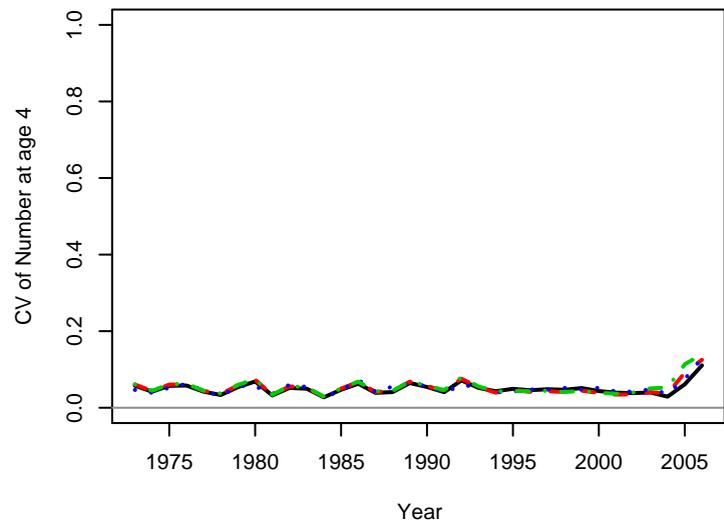
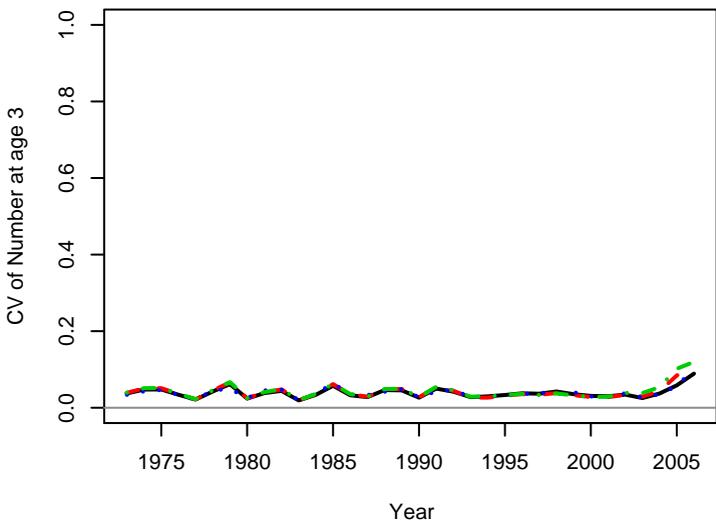
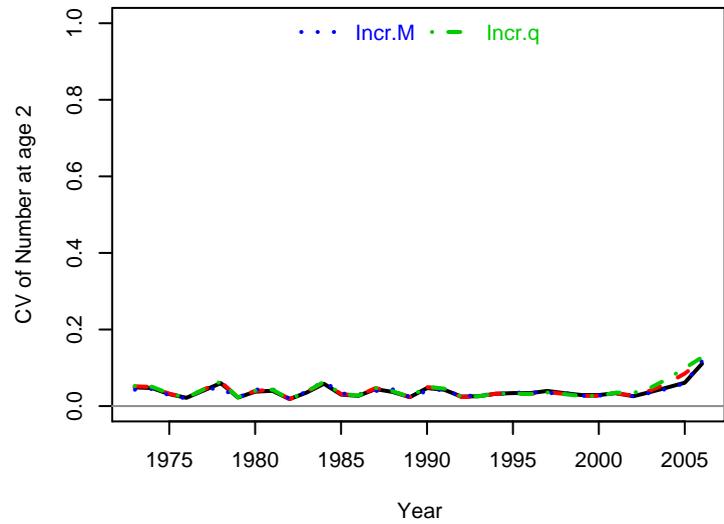
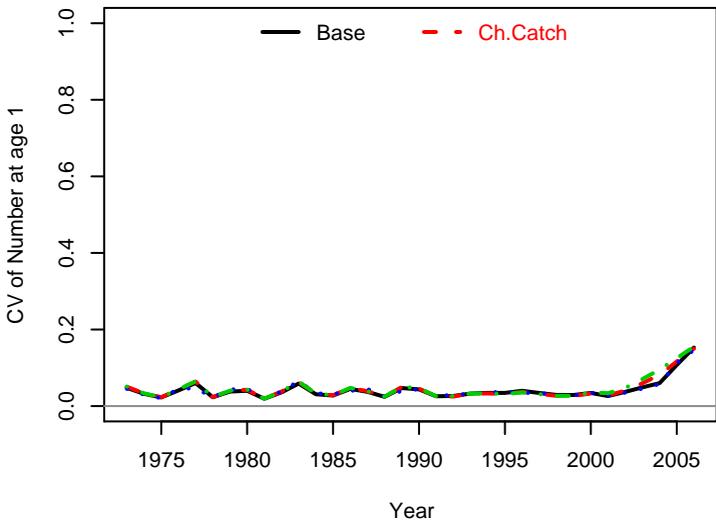


Figure GBYT.18 CV of number at age for all simulated data sets using ASAP.

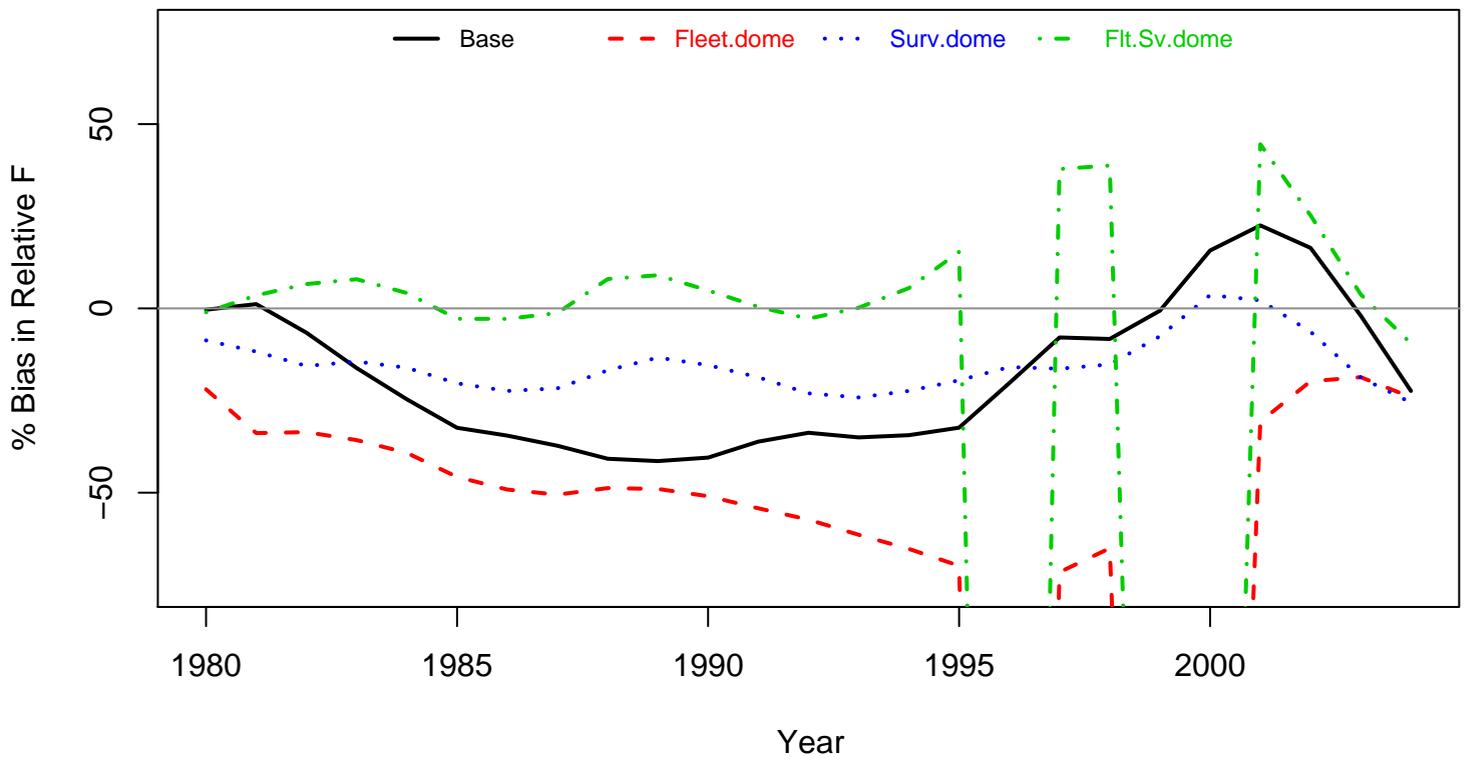
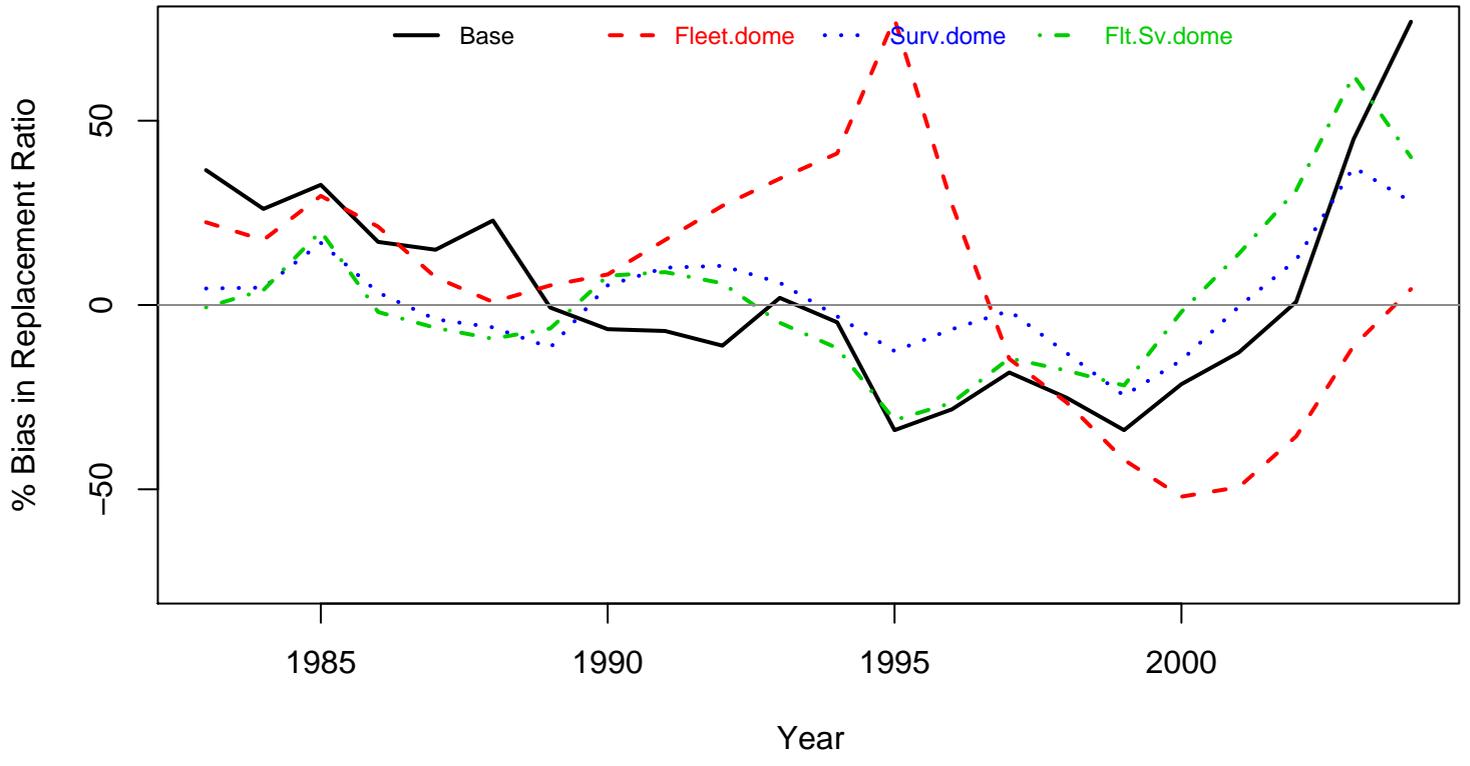


Figure GBCOD.1 Bias in replacement ratio and relative F for all simulated data sets using AIM. ³⁵

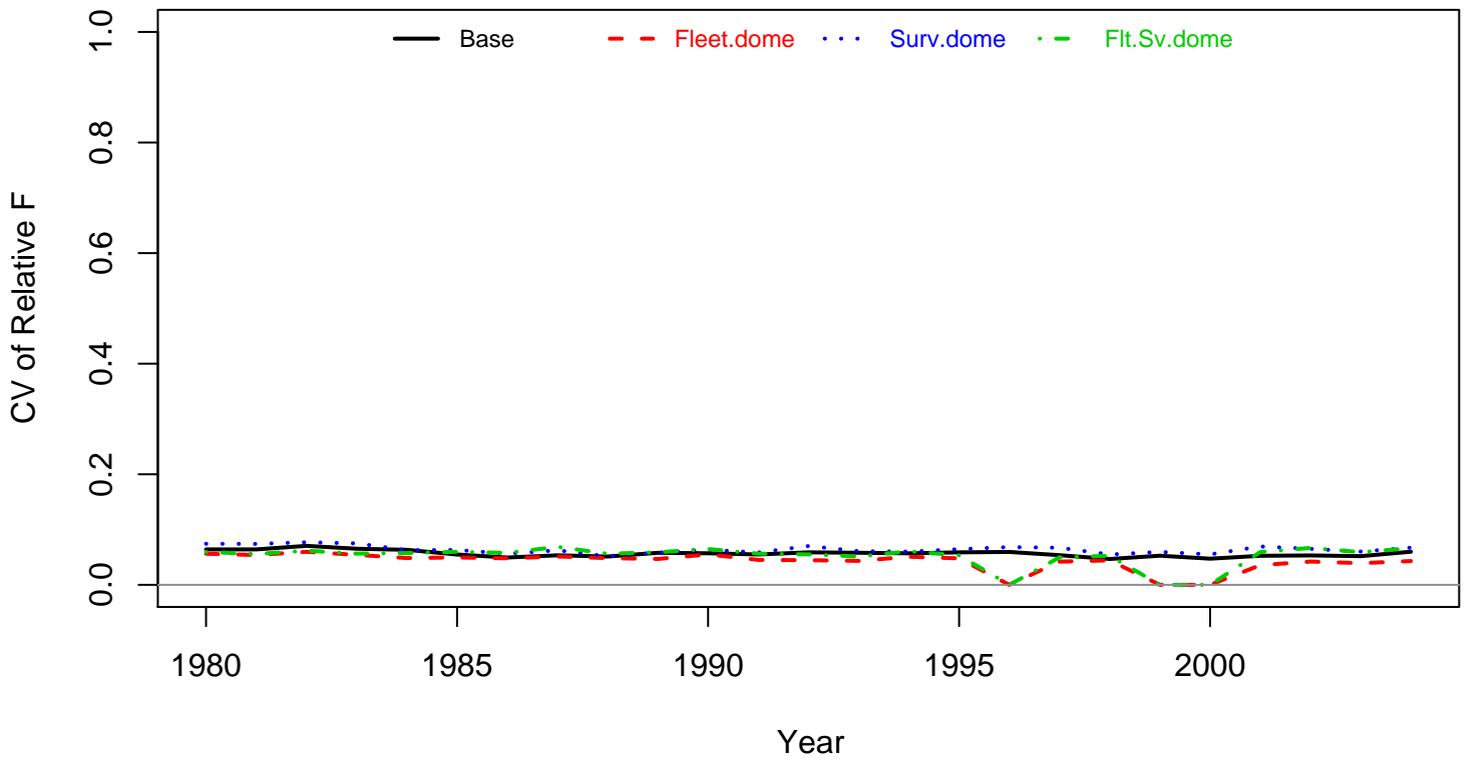
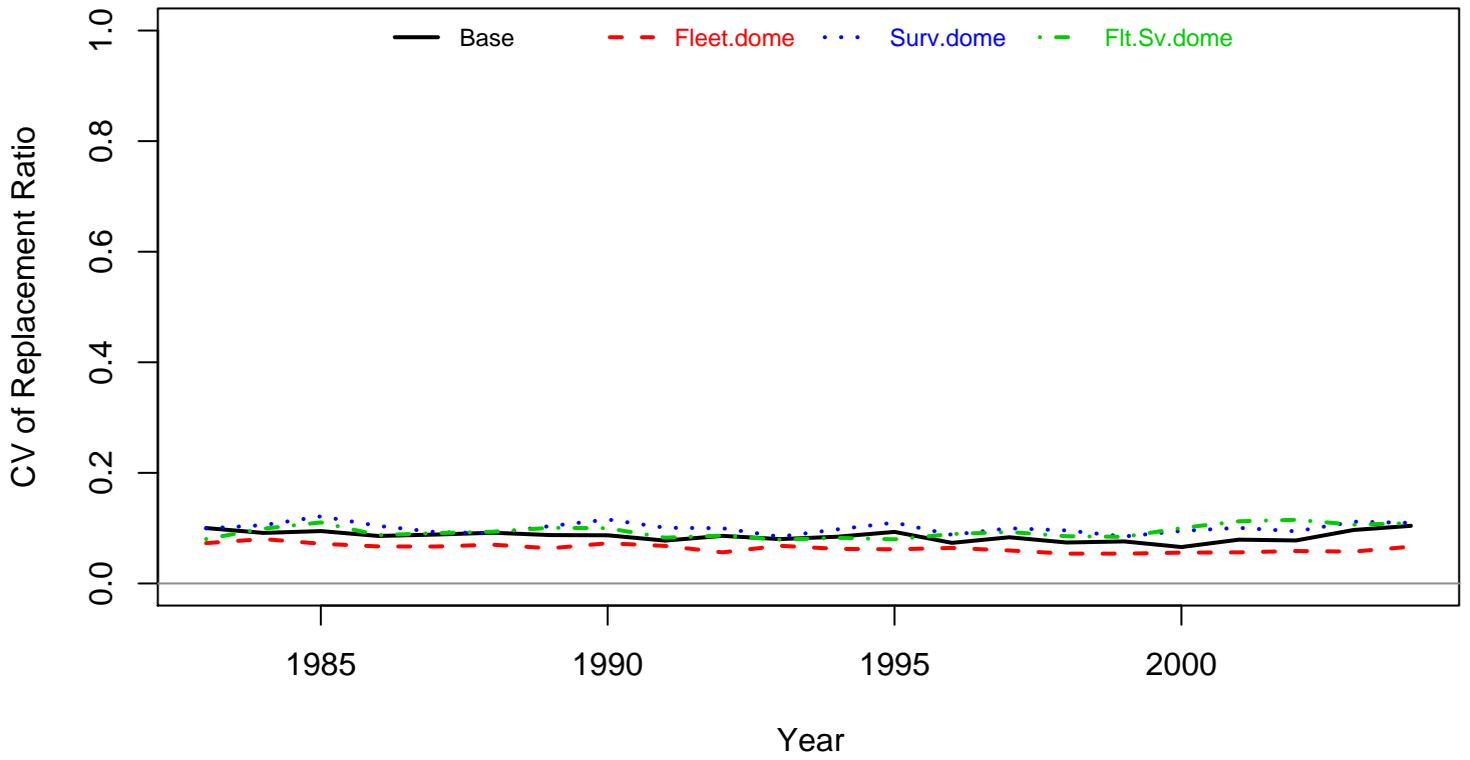


Figure GBCOD.2 CV in replacement ratio and relative F for all simulated data sets using AIM.³⁶

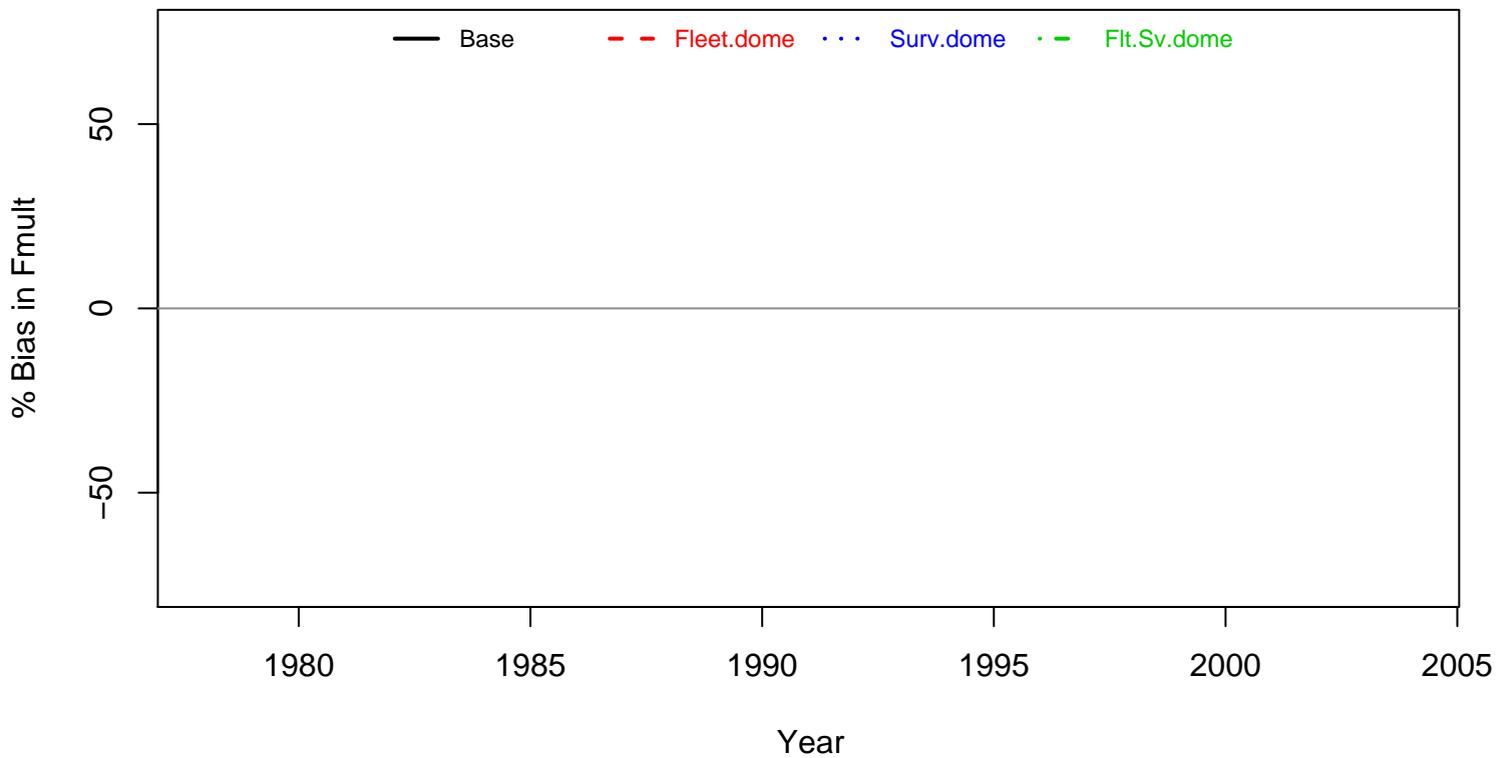
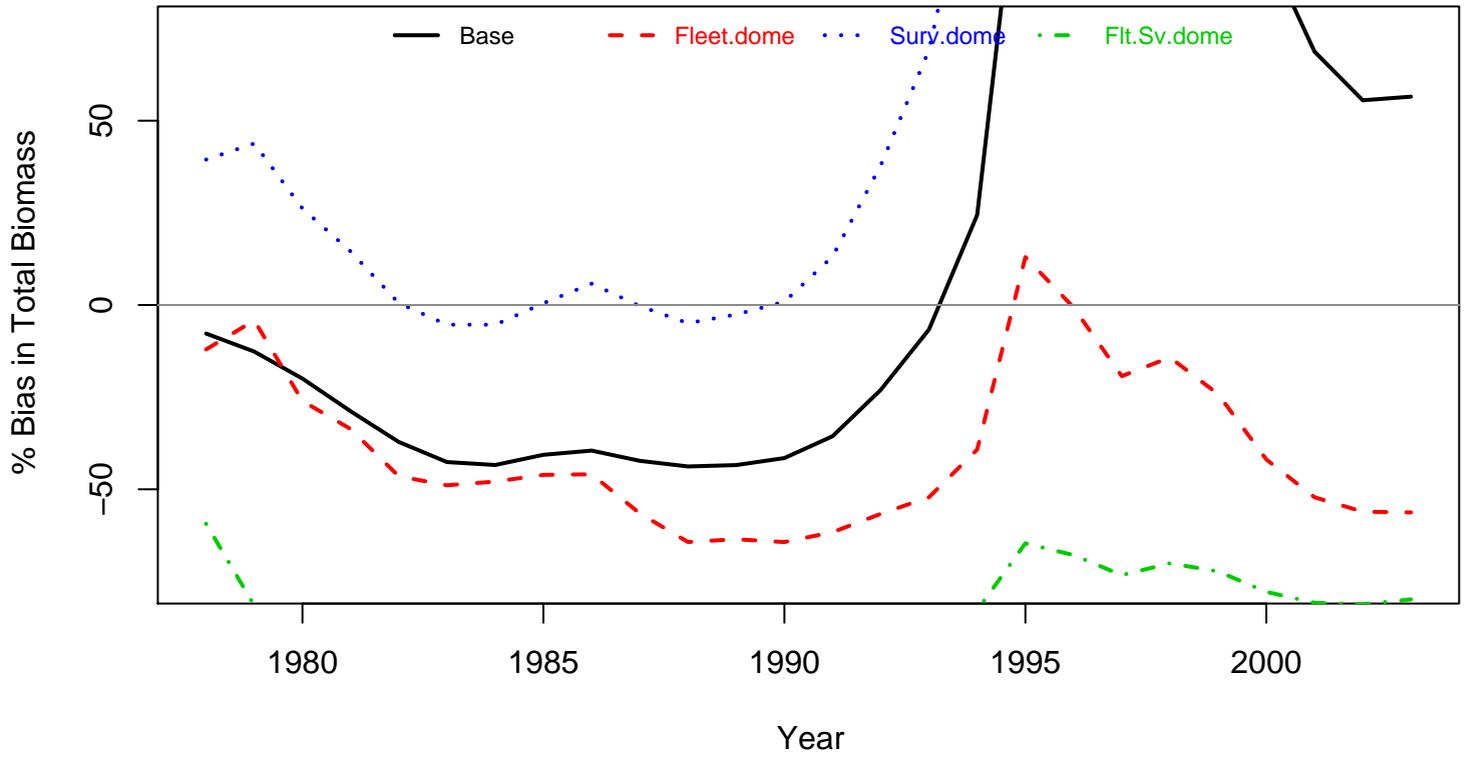


Figure GBCOD.3 Bias in total biomass and F for all simulated data sets using ASPIC.

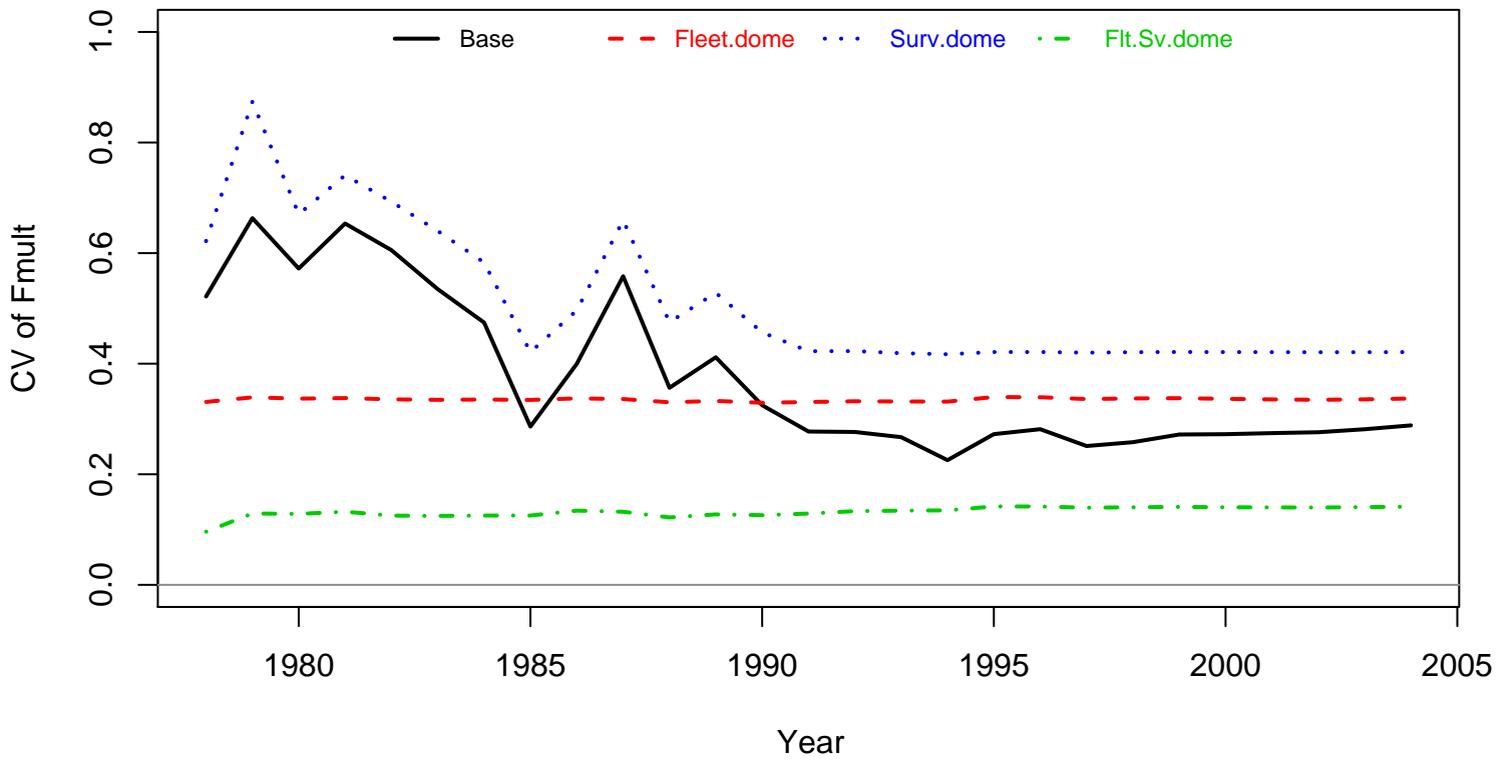
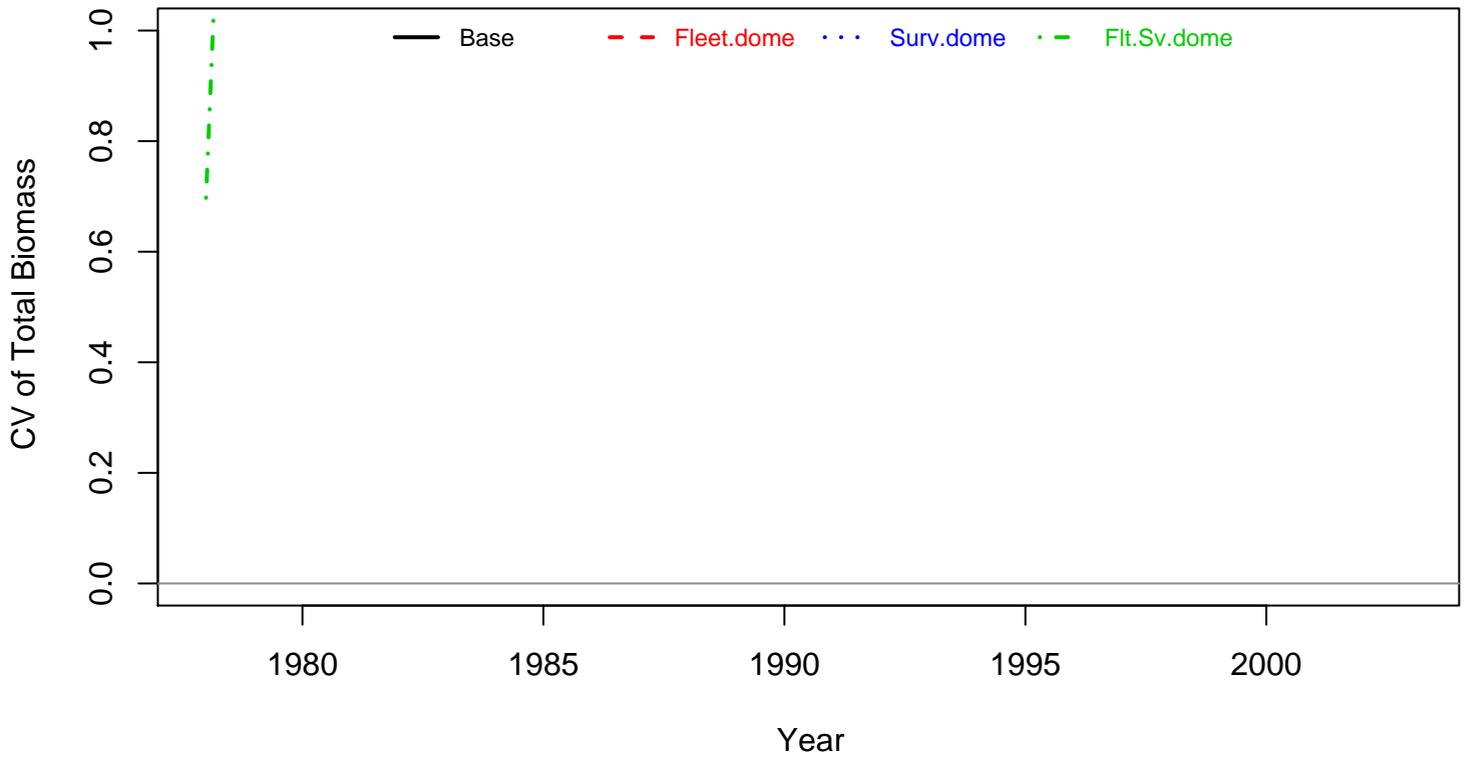


Figure GBCOD.4 CV in total biomass and F for all simulated data sets using ASPIC.

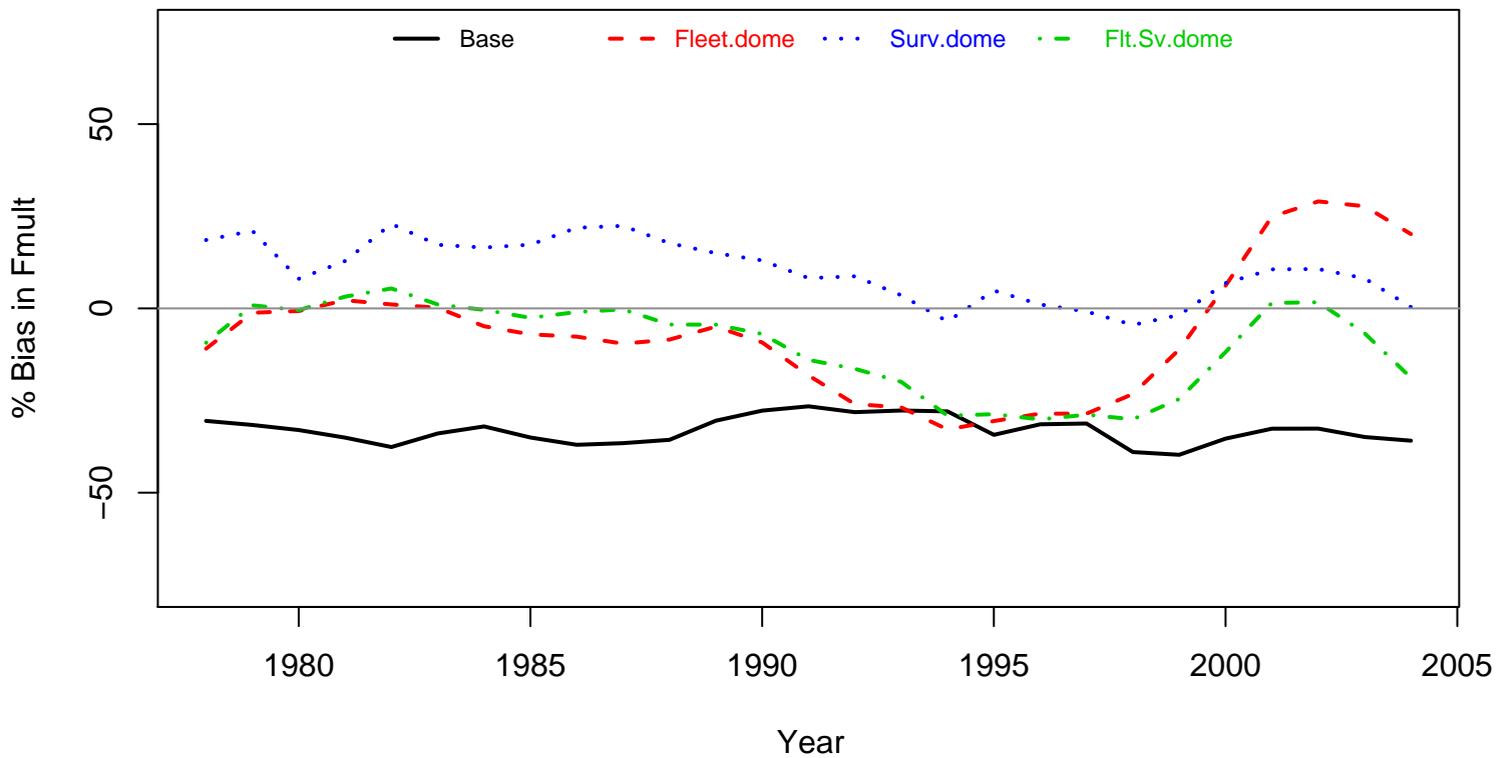
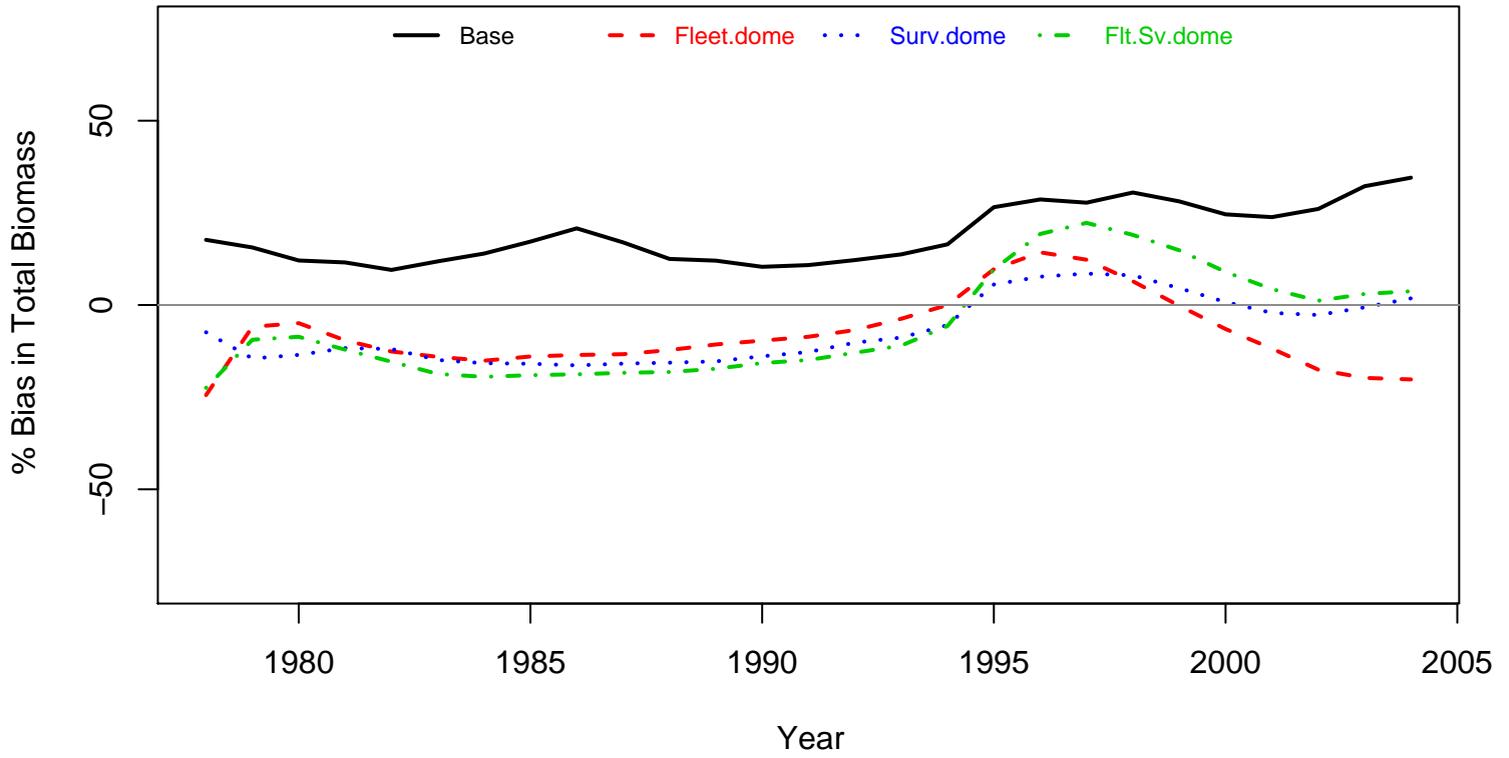


Figure GBCOD.5 Bias in total biomass and F multiplier for all simulated data sets using Scale. ³⁹

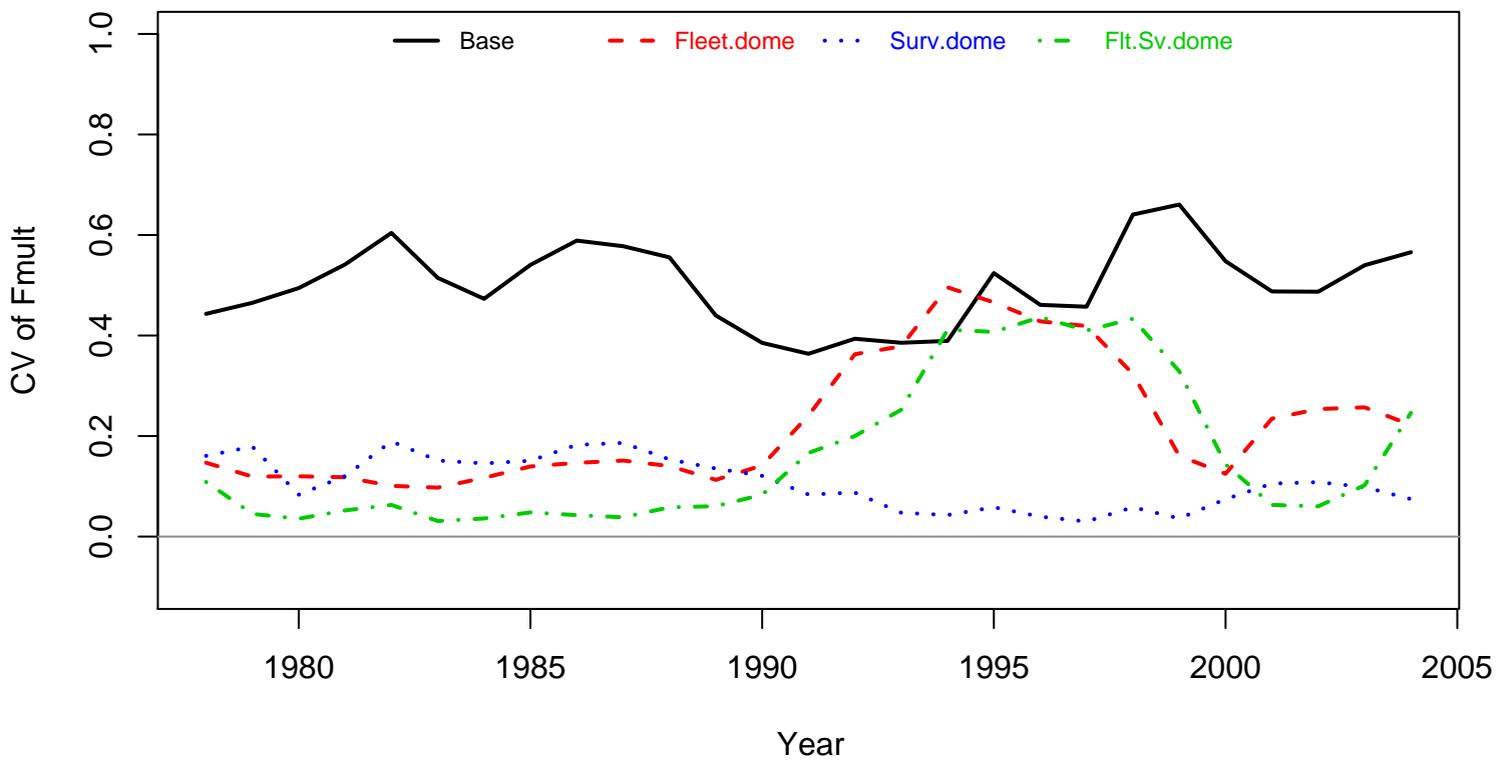
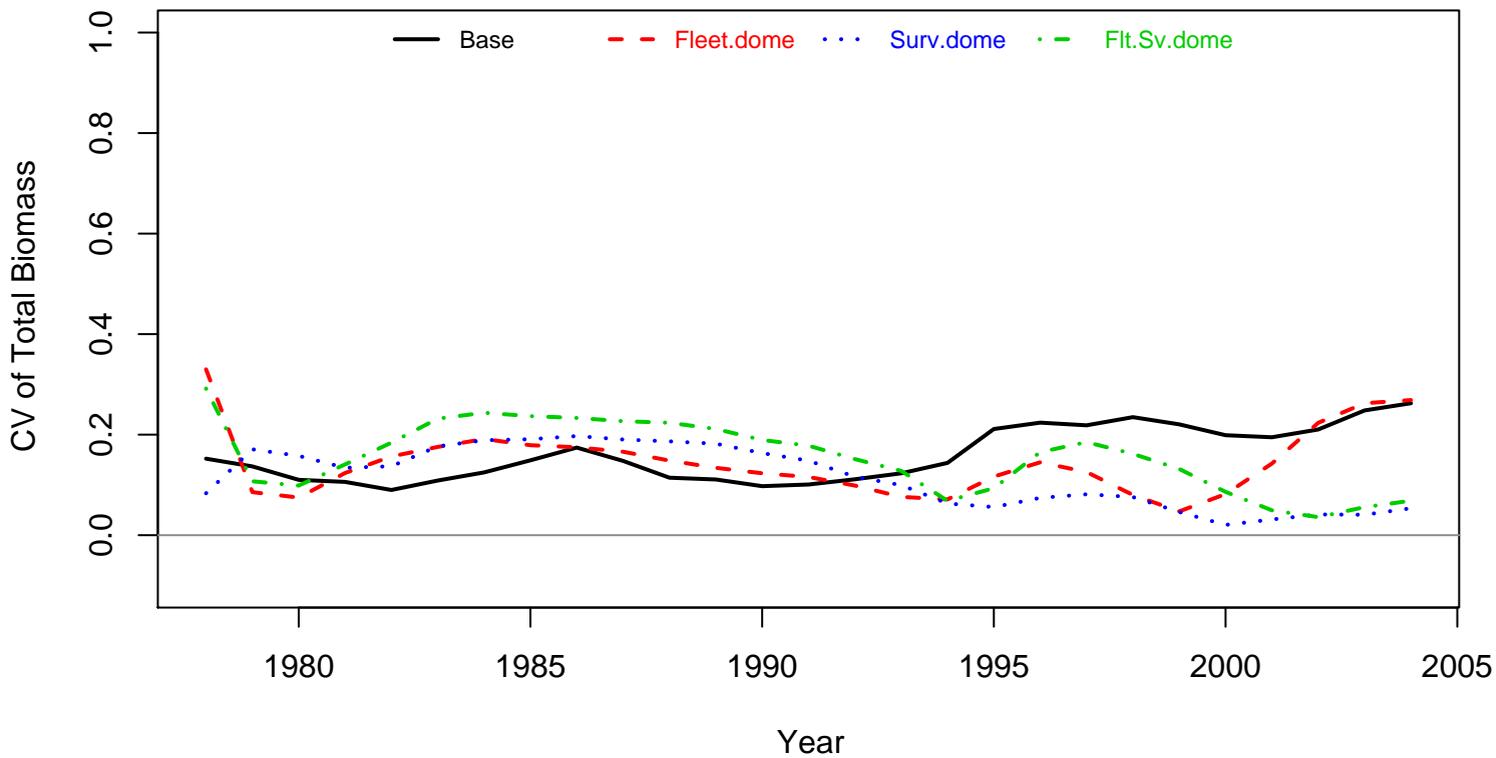


Figure GBCOD.6 CV in total biomass and F multiplier for all simulated data sets using Scale. ⁴⁰

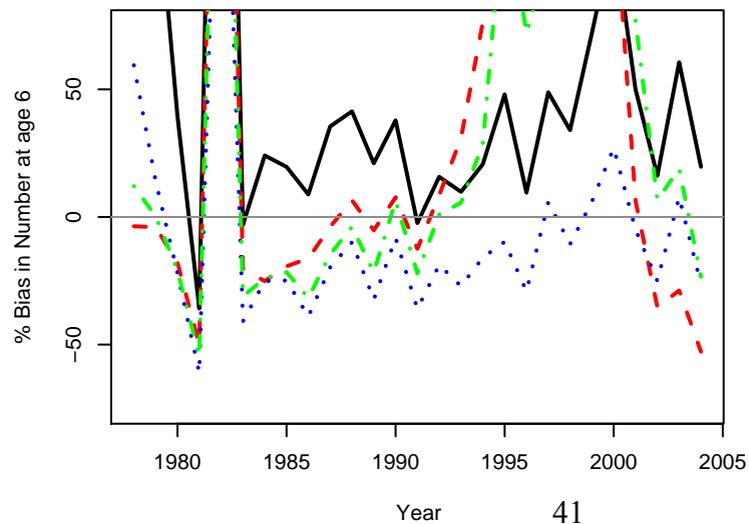
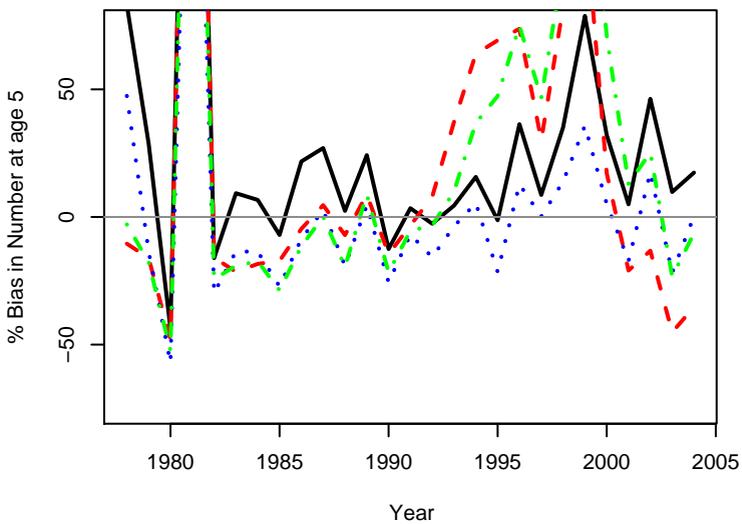
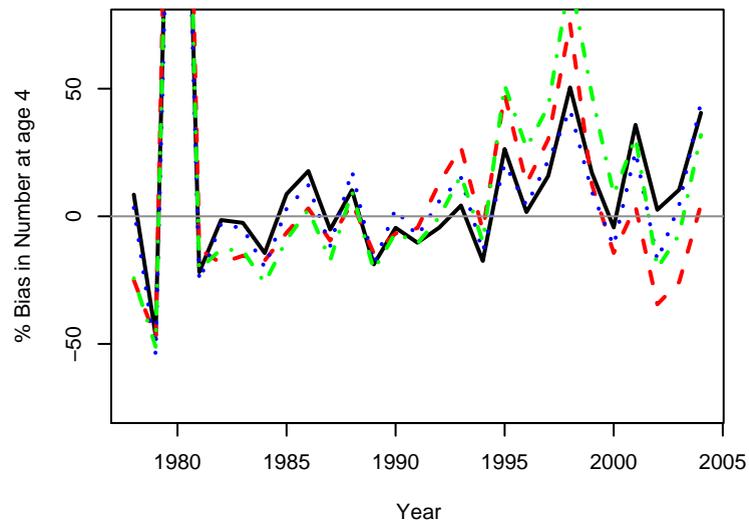
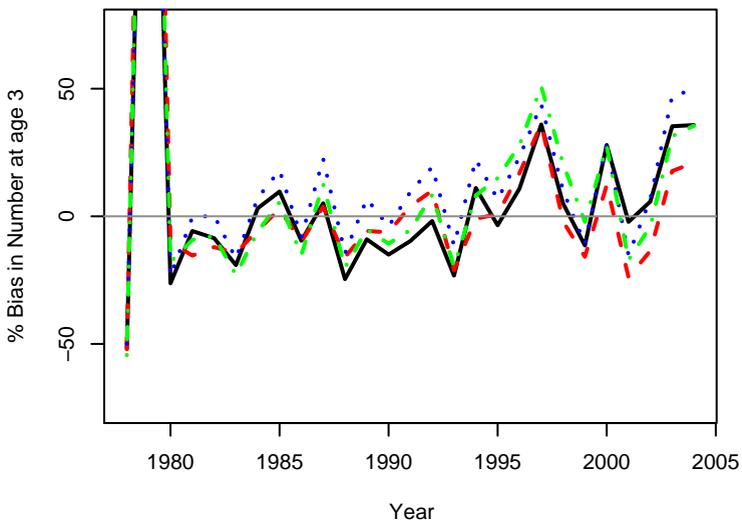
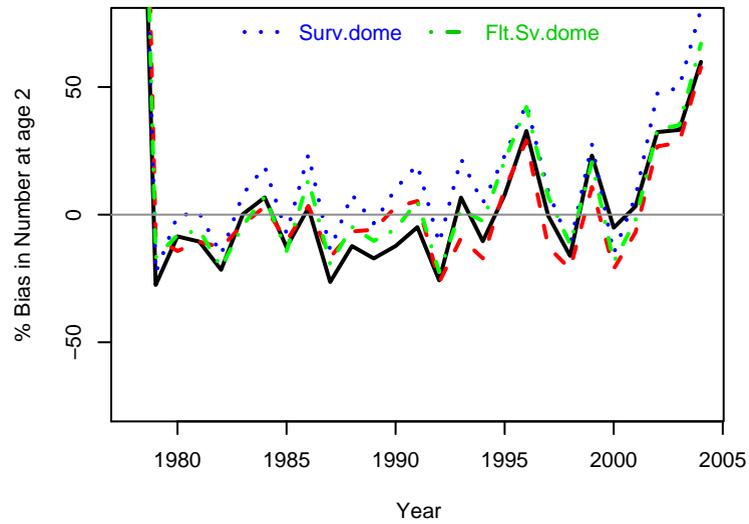
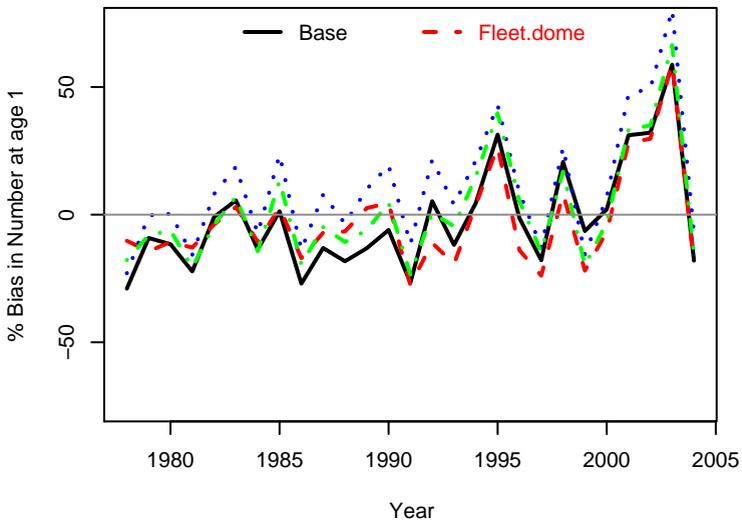


Figure GBCOD.7 Bias in number at age for all simulated data sets using Scale.

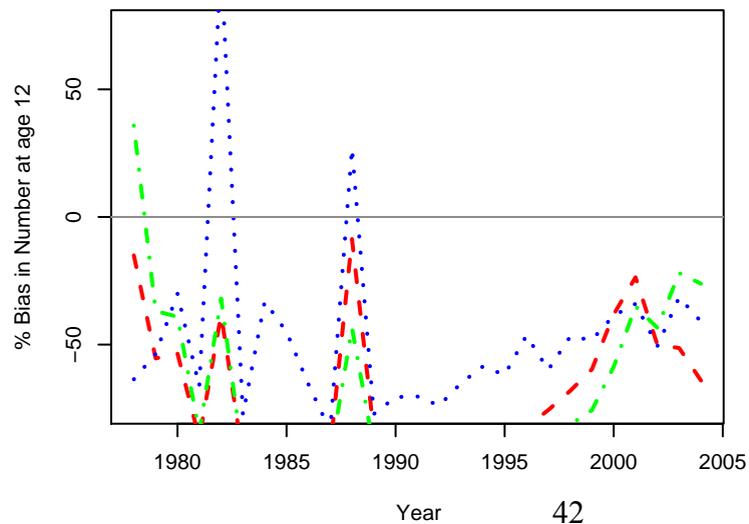
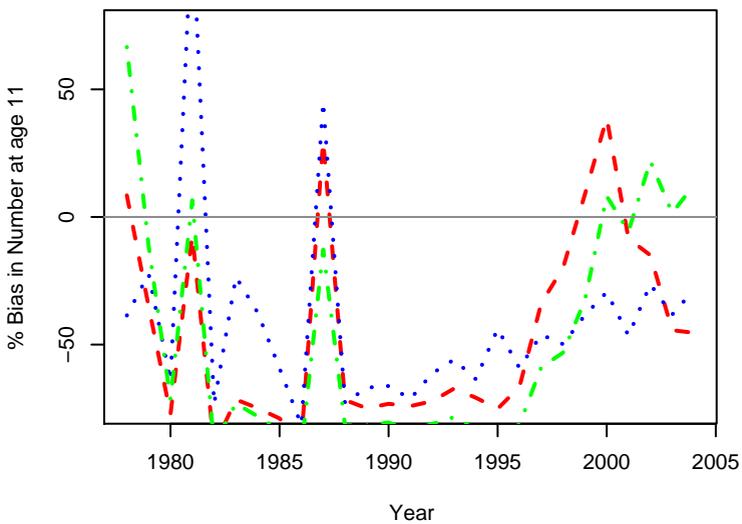
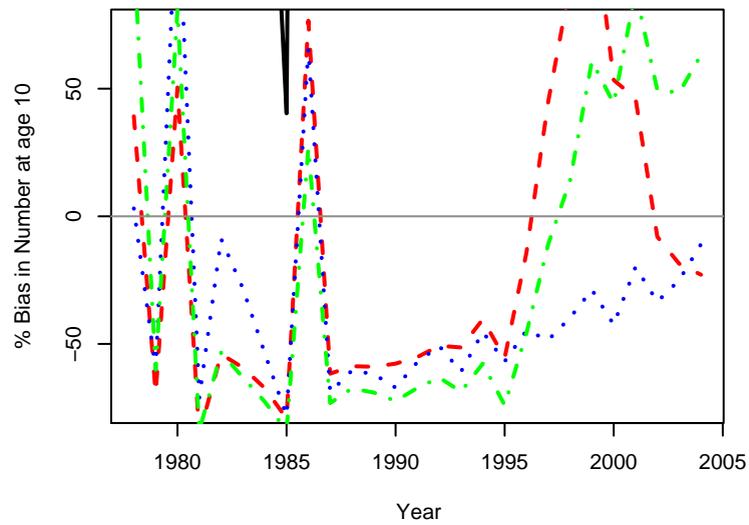
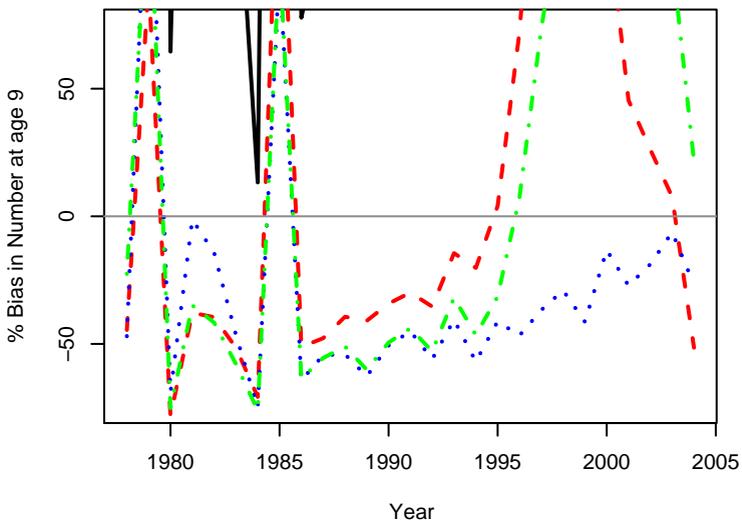
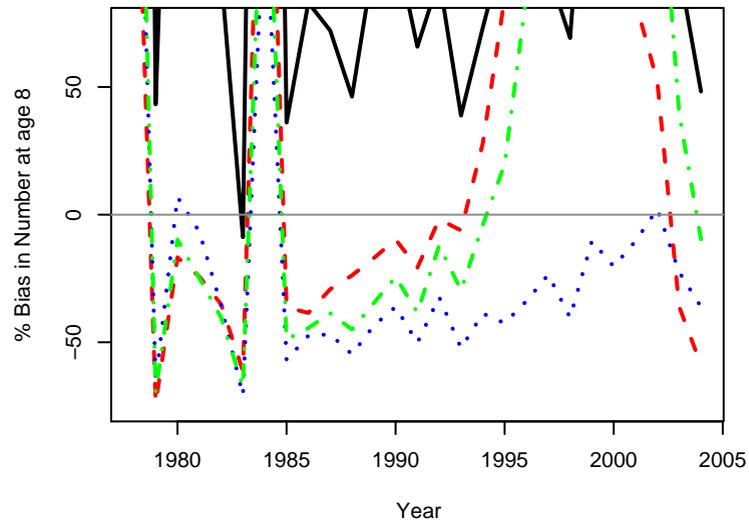
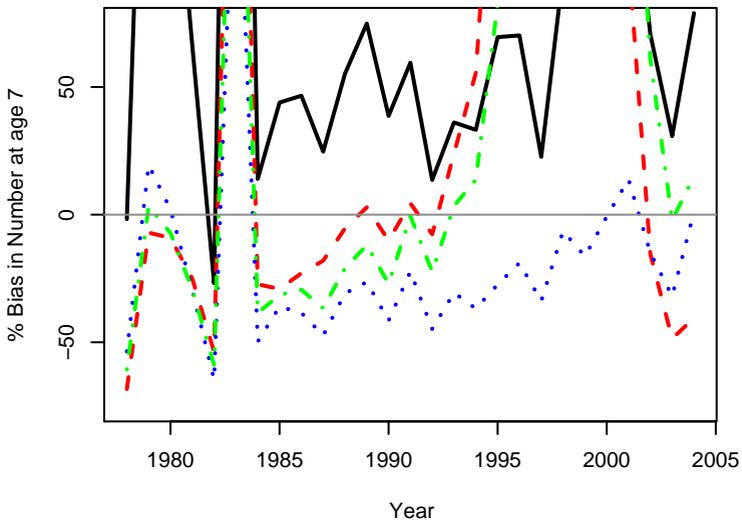


Figure GBCOD.7 (cont.)

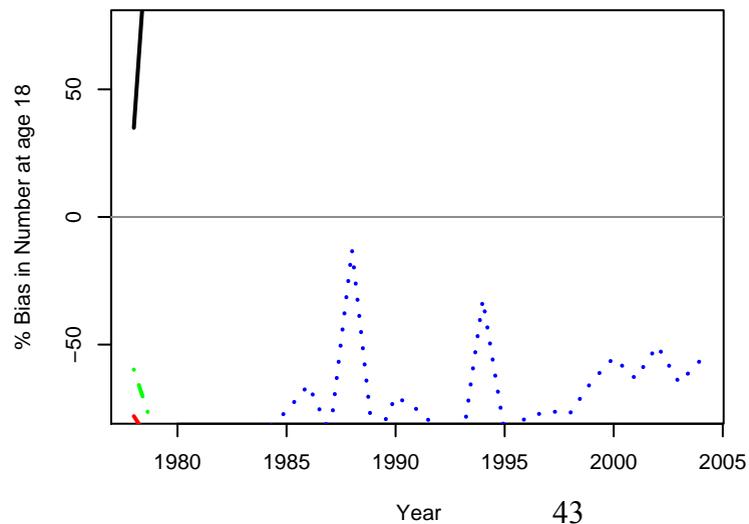
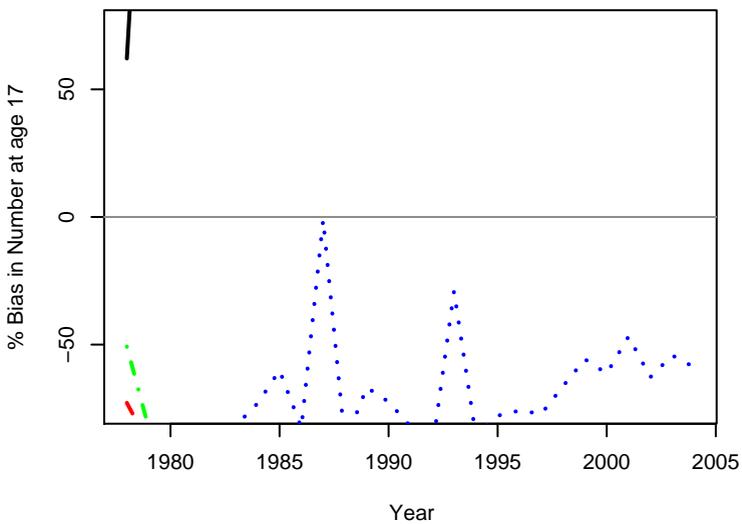
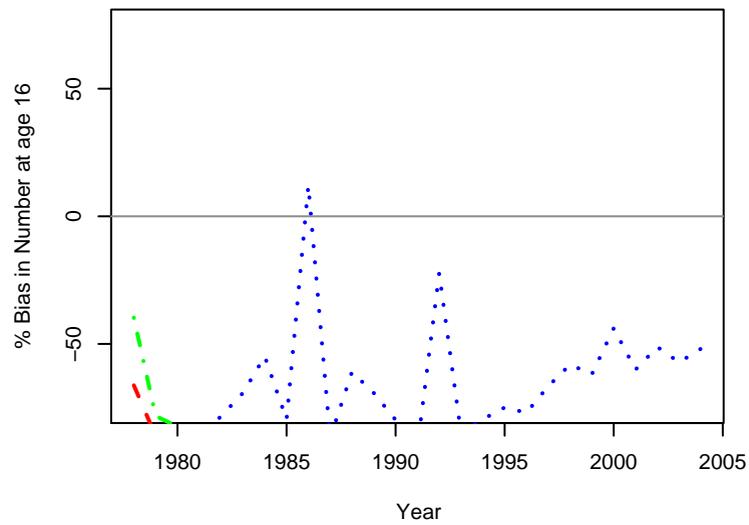
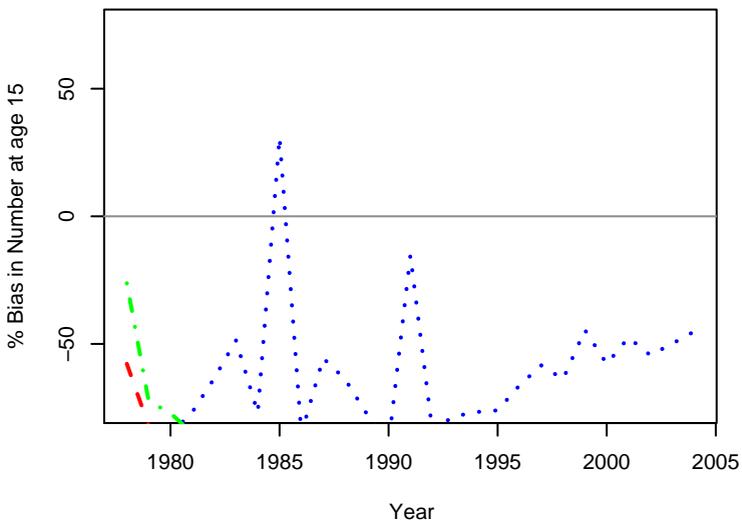
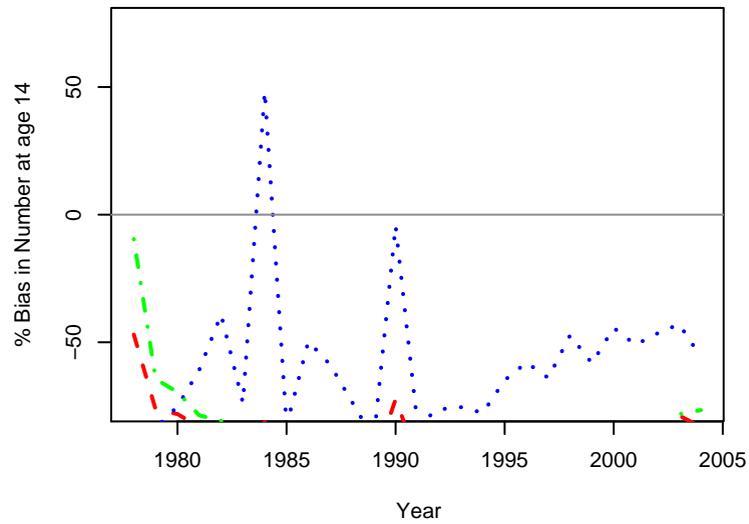
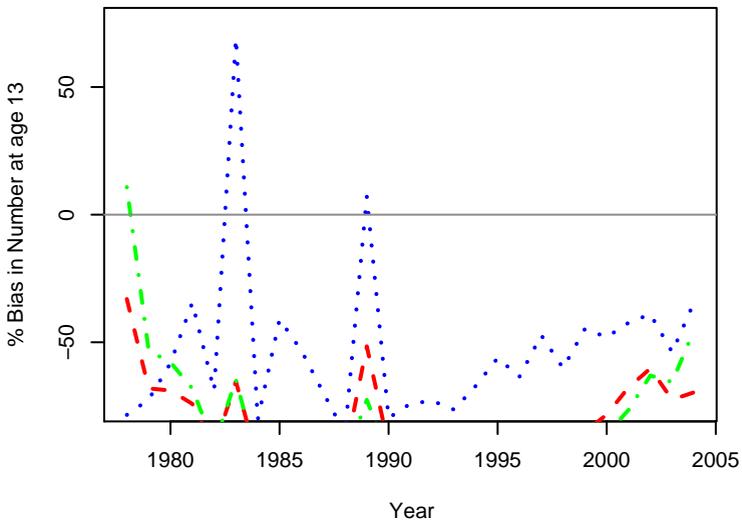


Figure GBCOD.7 (cont.)

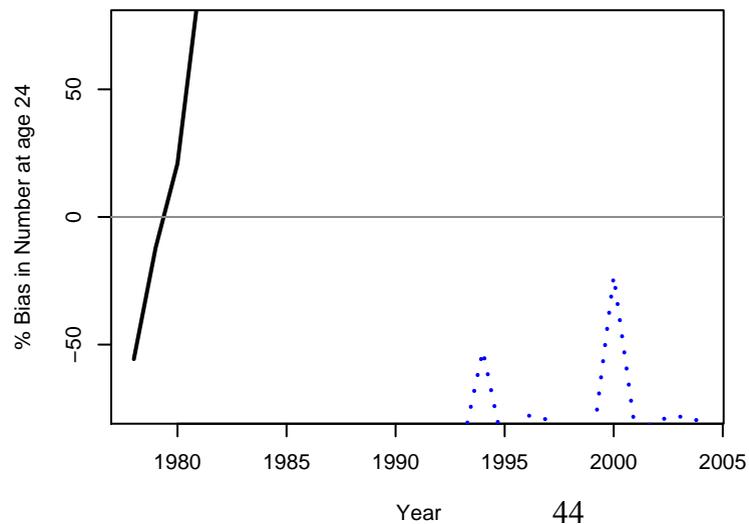
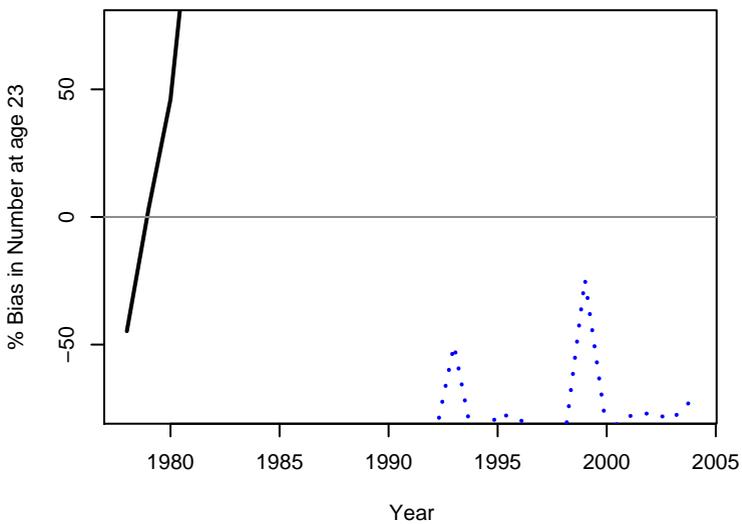
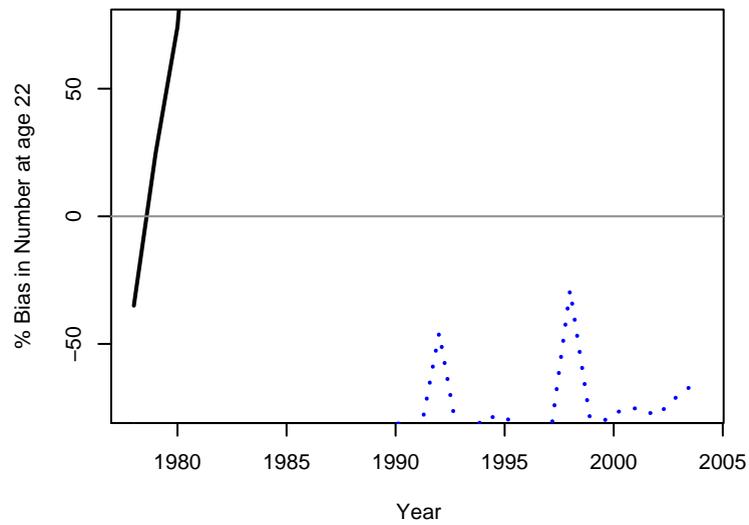
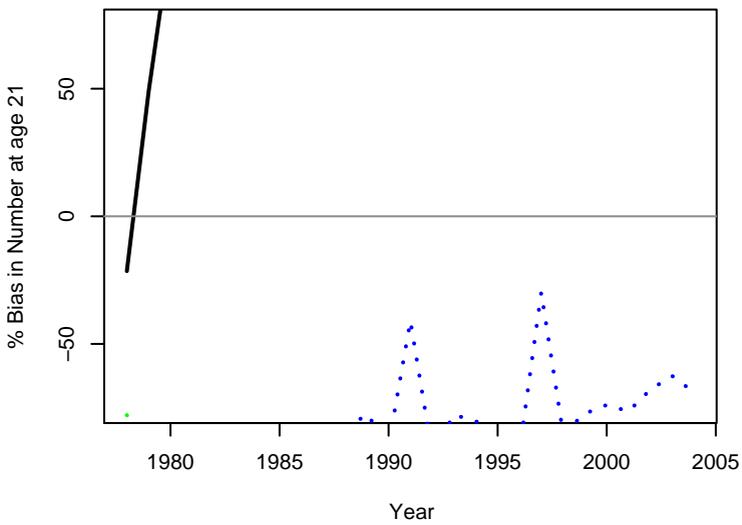
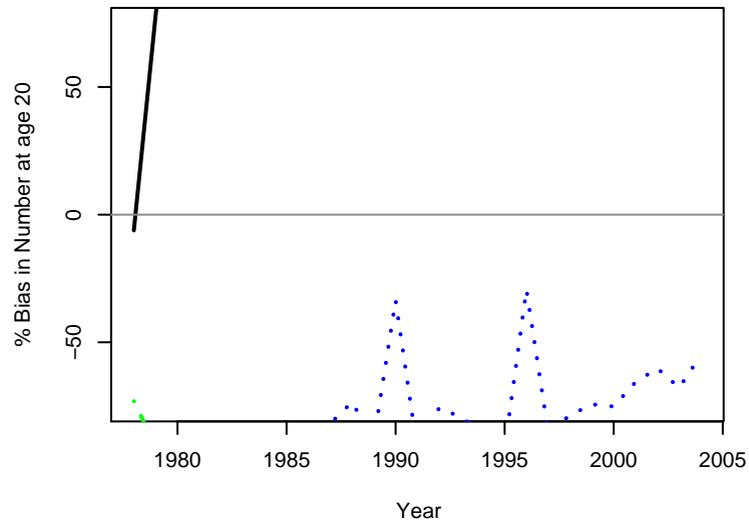
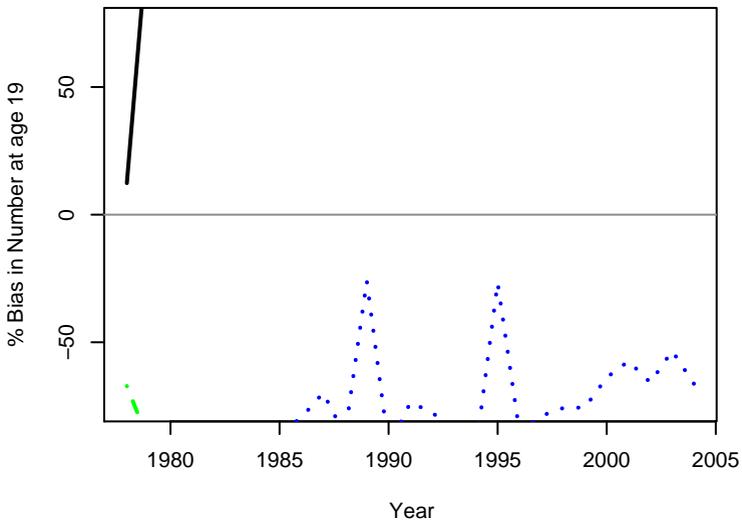


Figure GBCOD.7 (cont.)

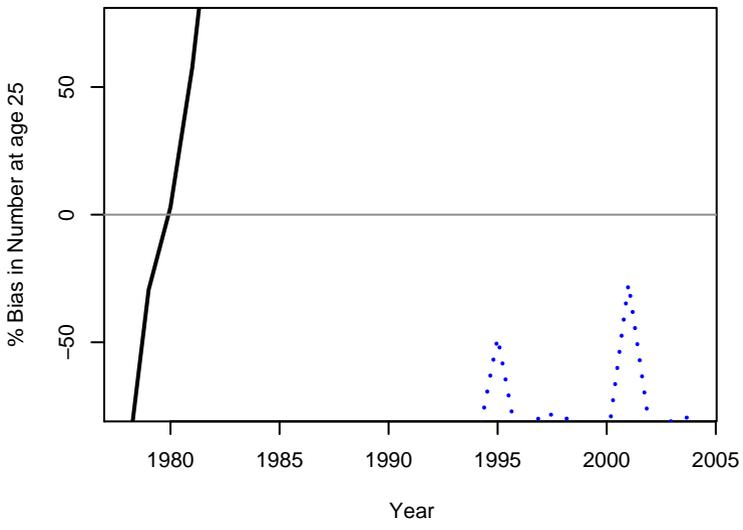


Figure GBCOD.7 (cont.)

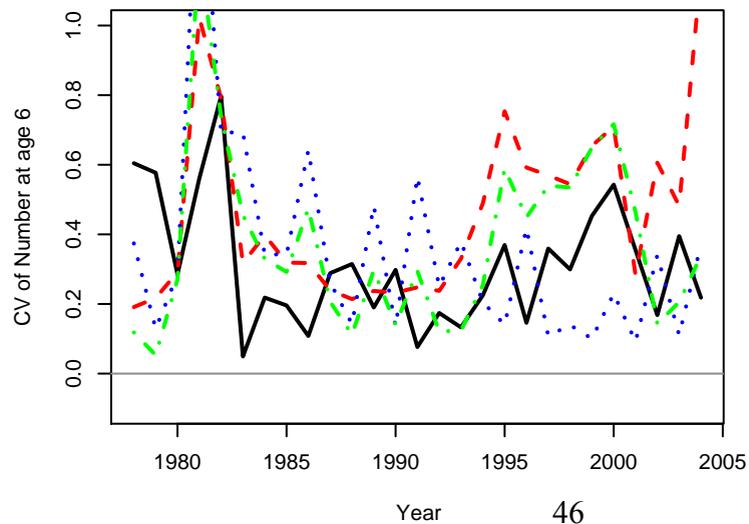
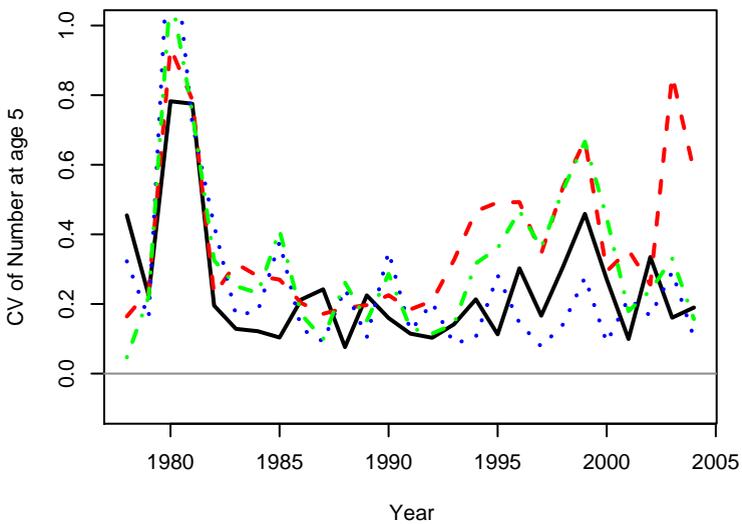
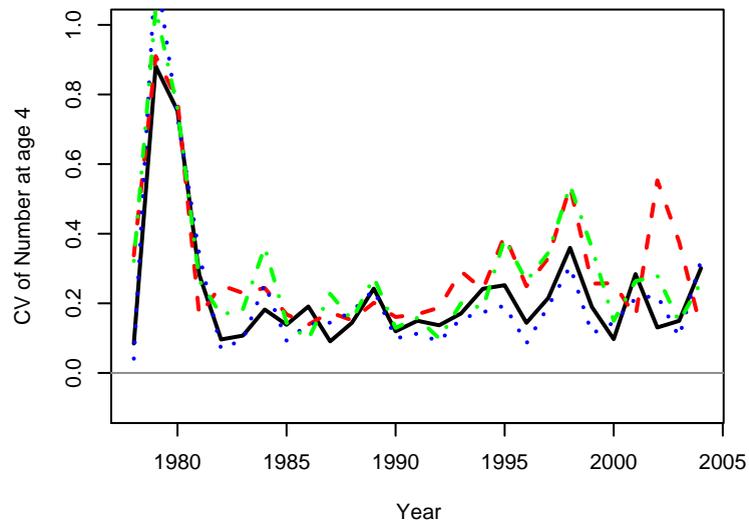
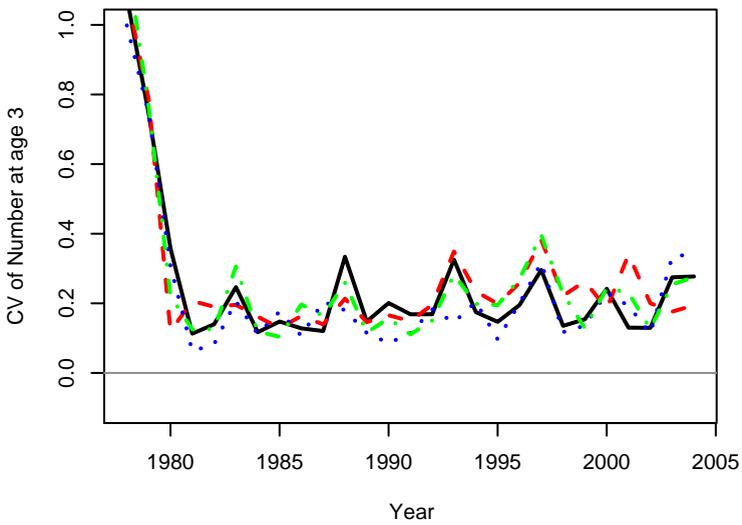
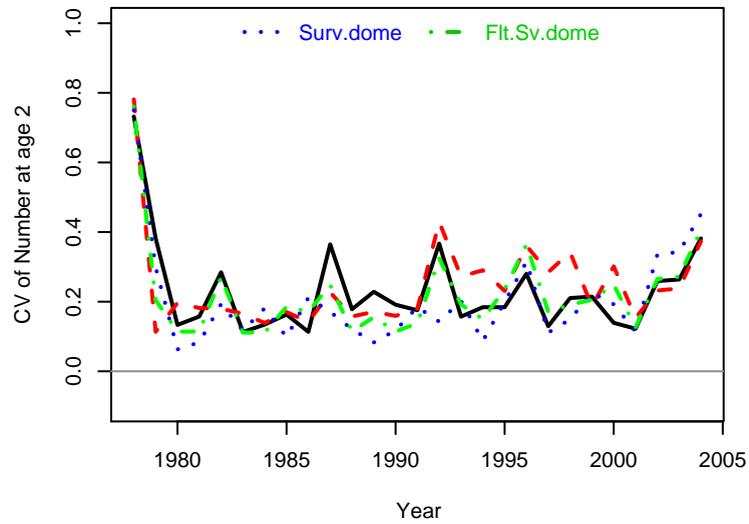
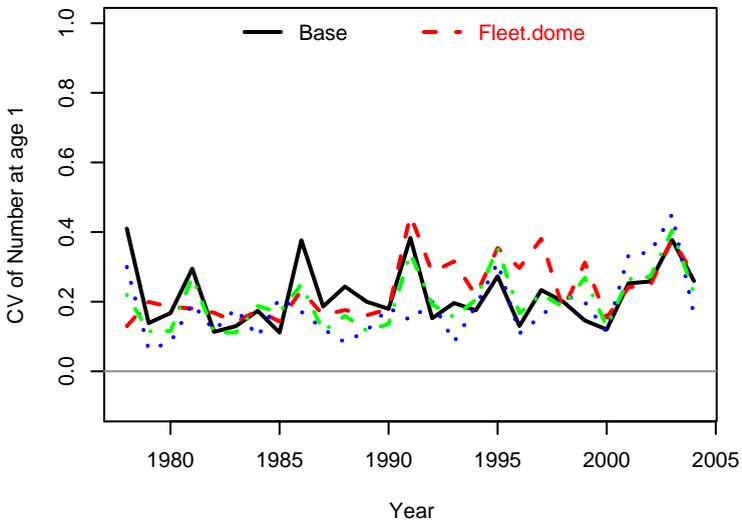


Figure GBCOD.8 CV in number at age for all simulated data sets using Scale.

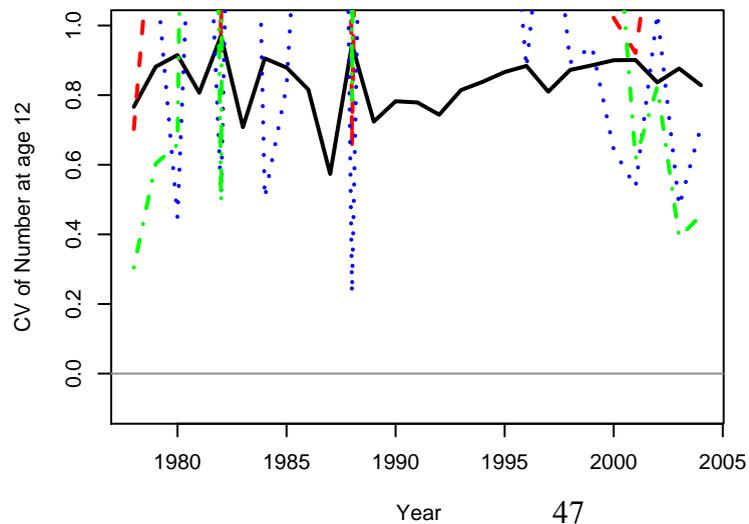
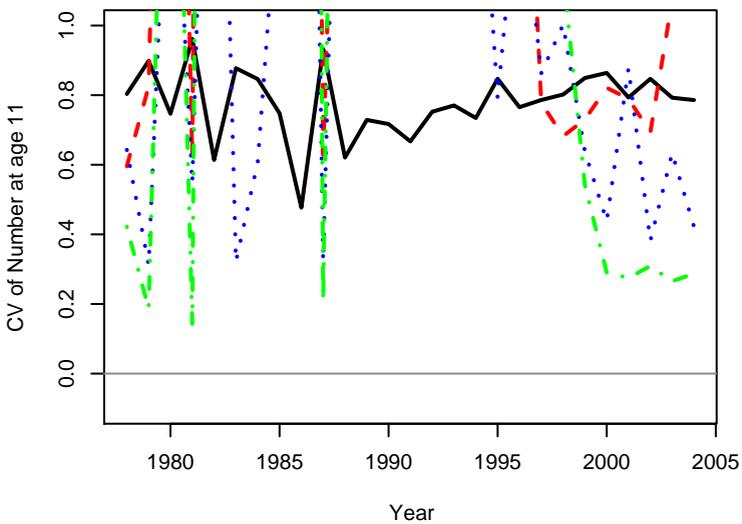
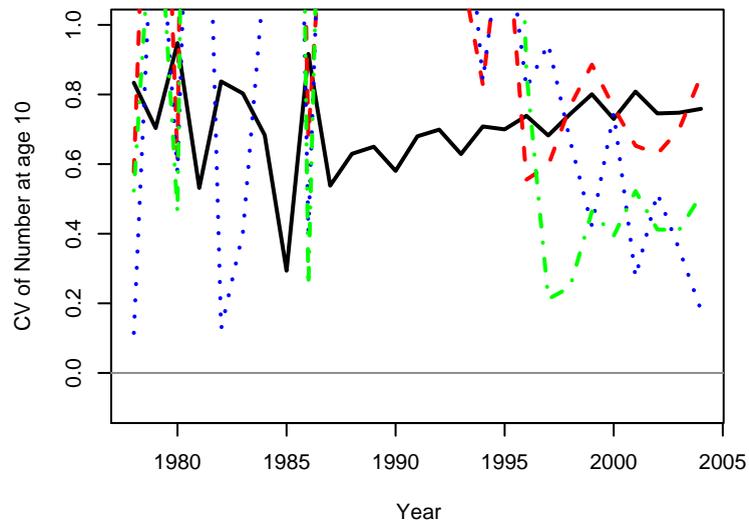
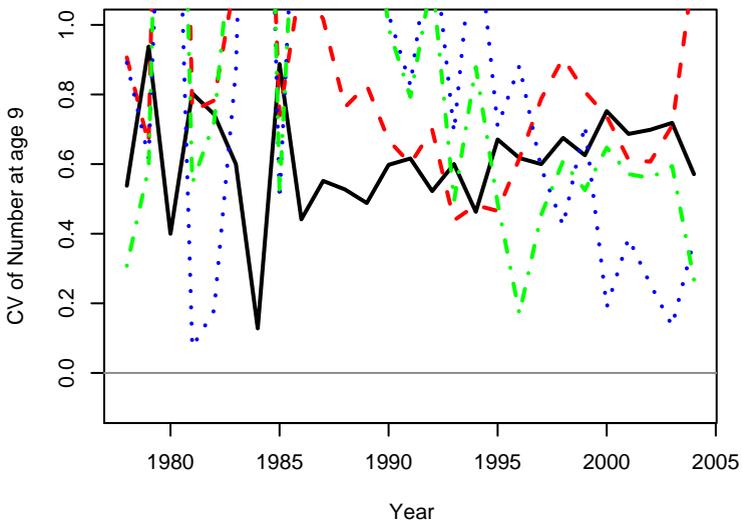
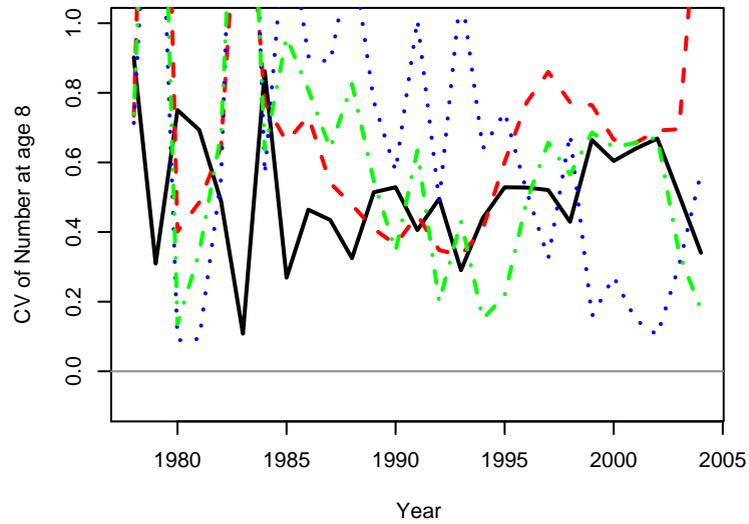
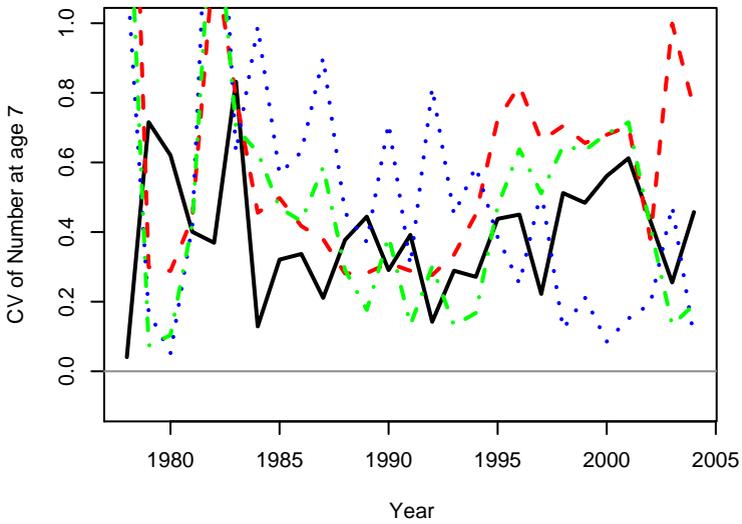


Figure GBCOD.8 (cont.)

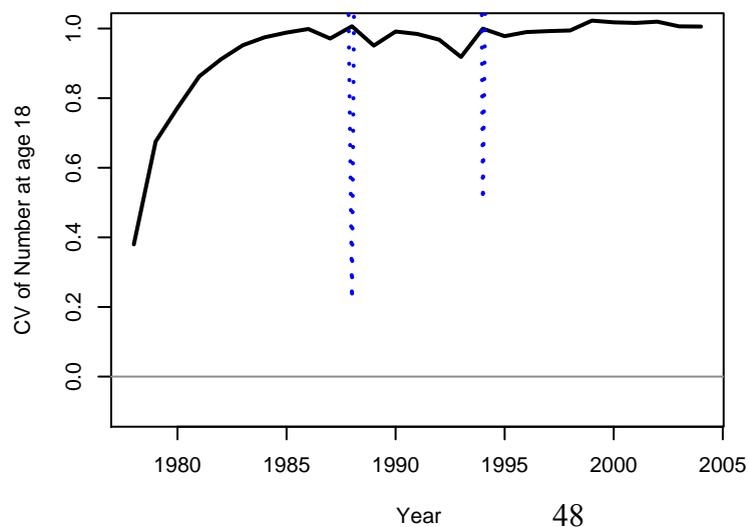
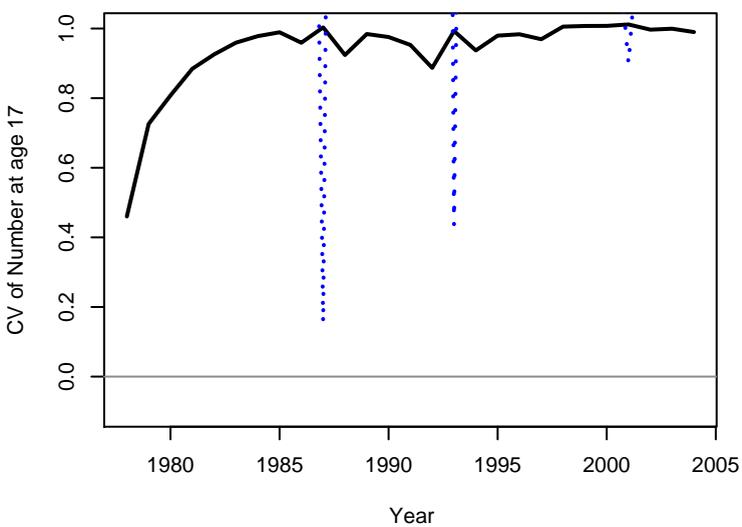
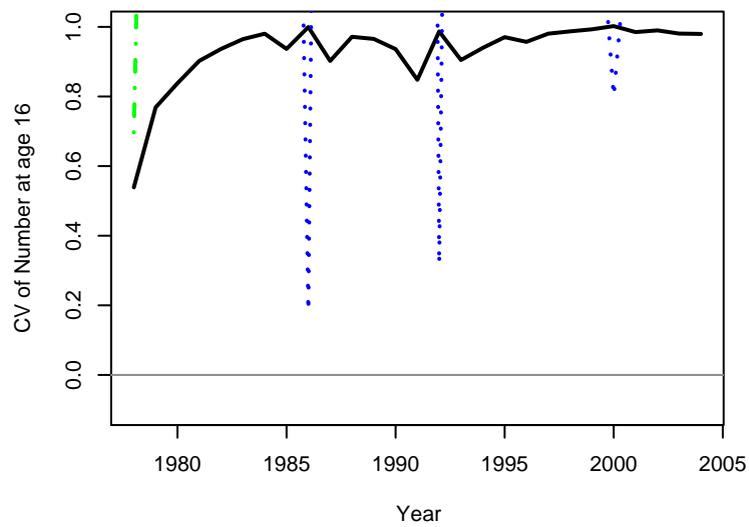
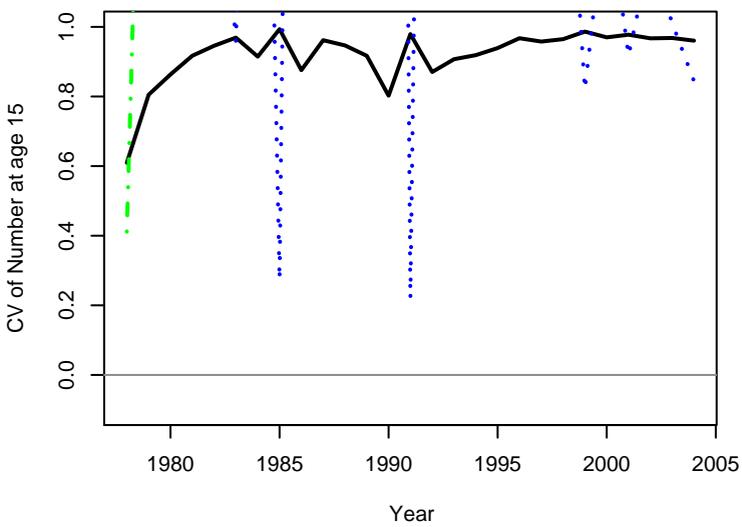
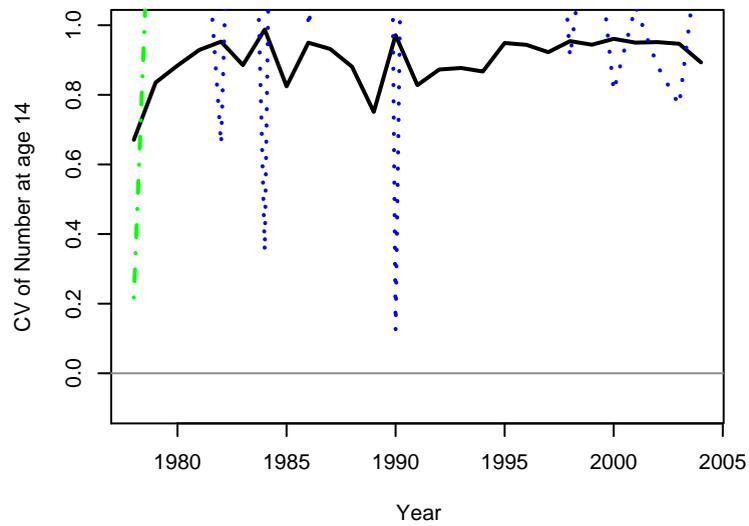
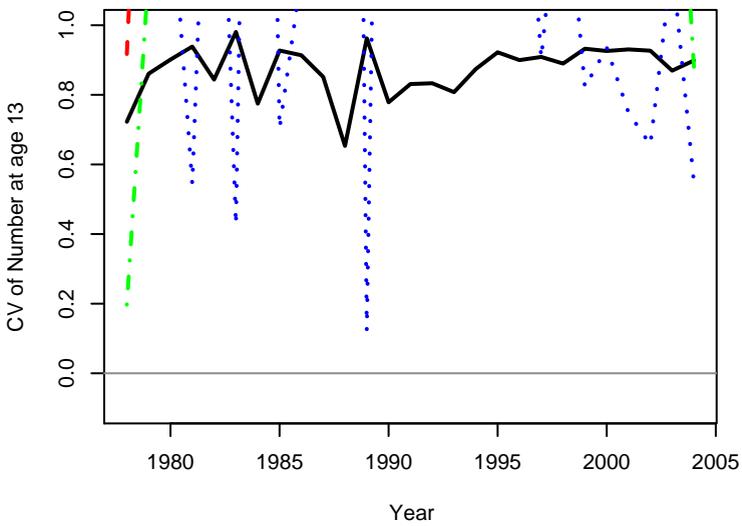


Figure GBCOD.8 (cont.)

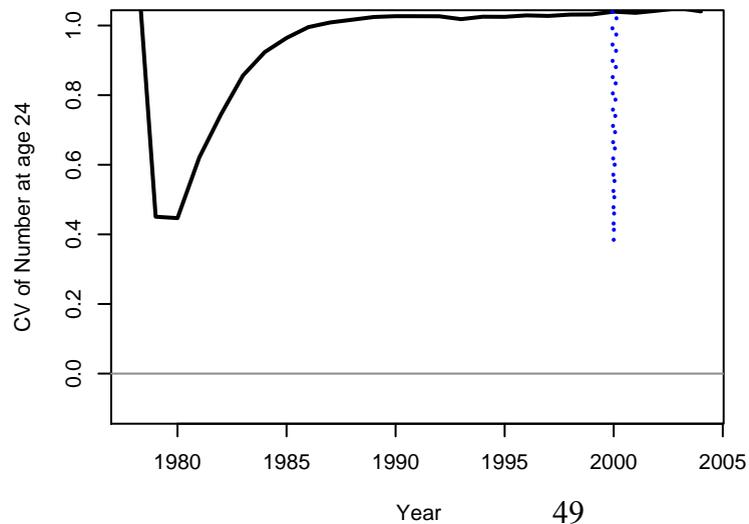
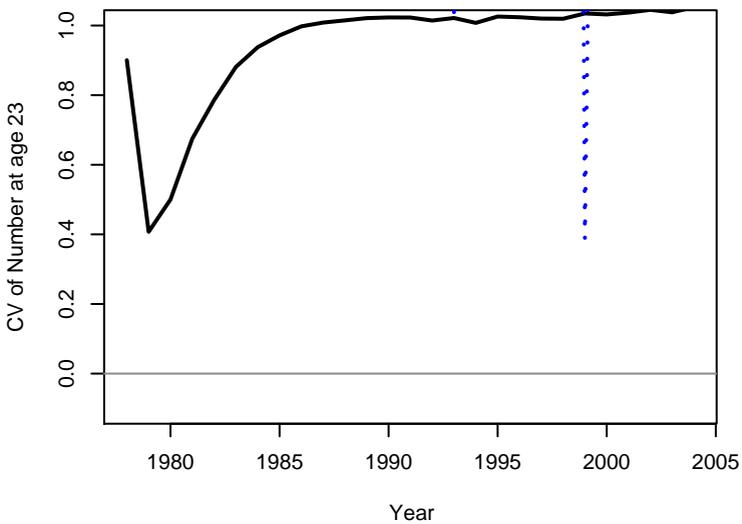
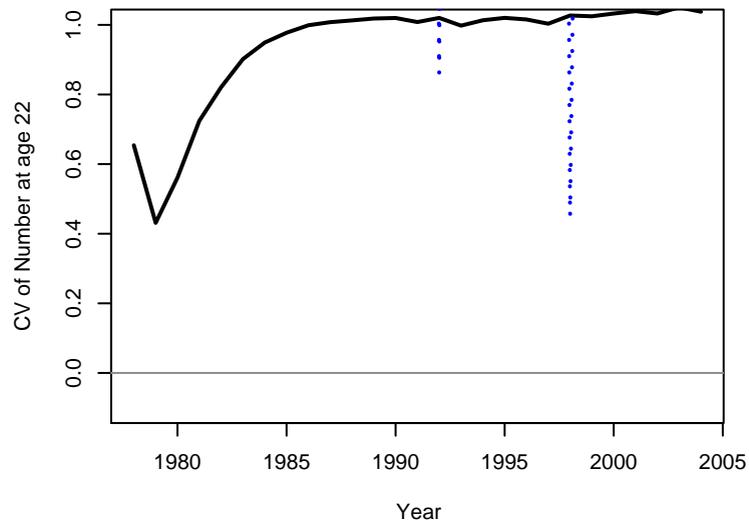
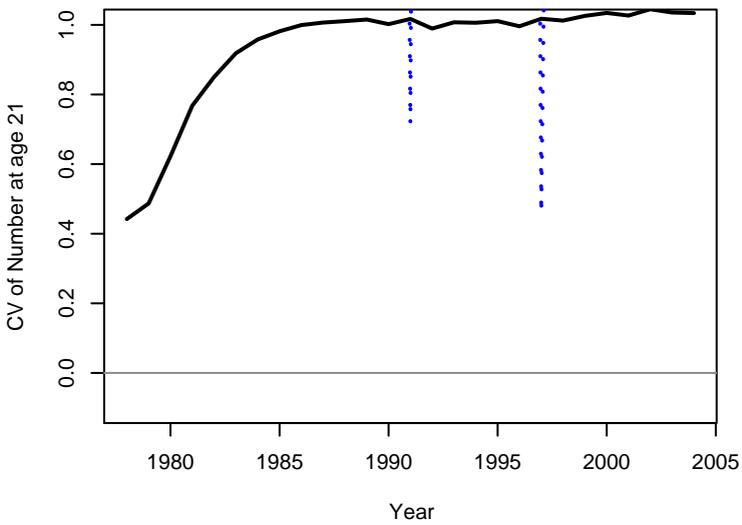
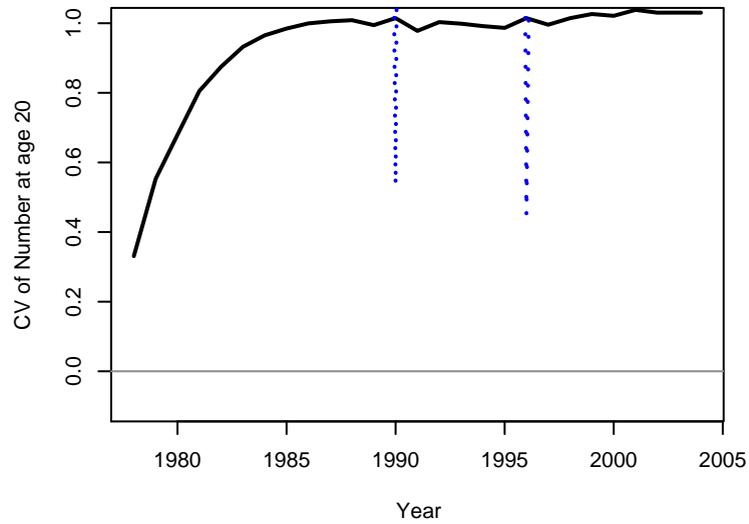
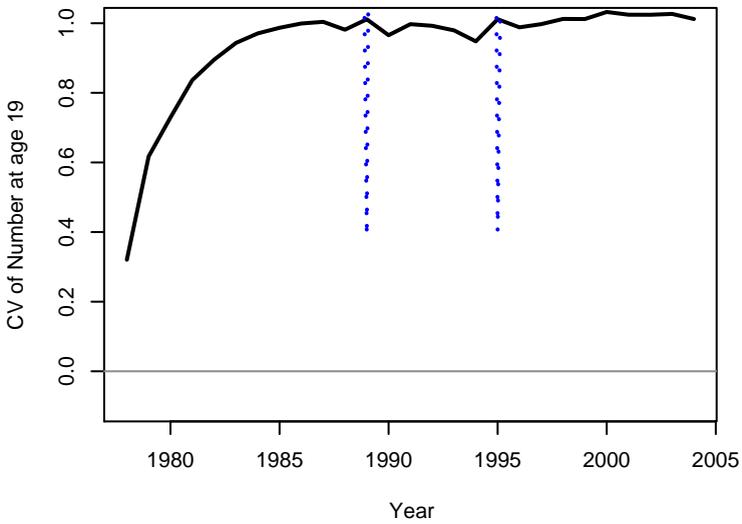
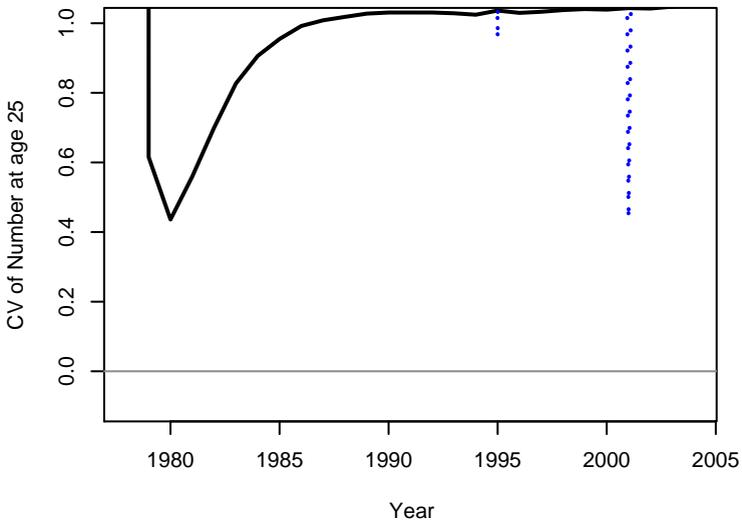


Figure GBCOD.8 (cont.)



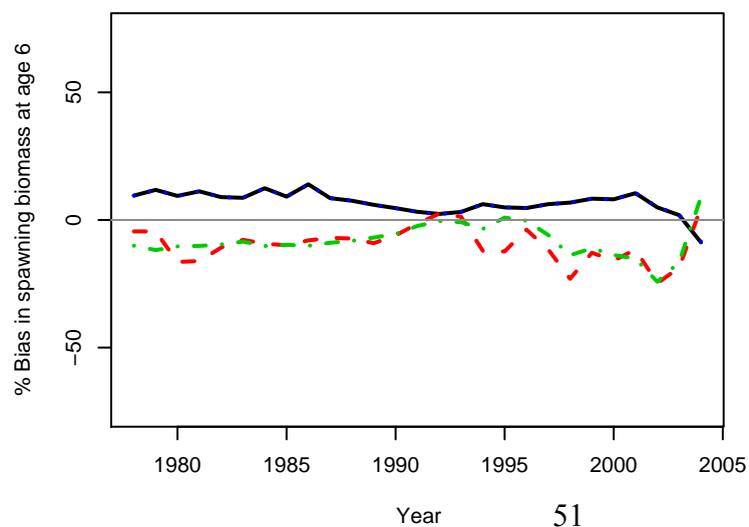
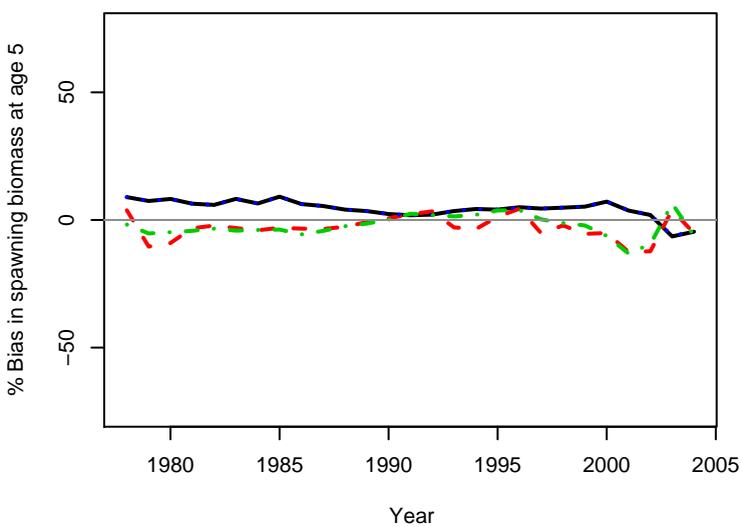
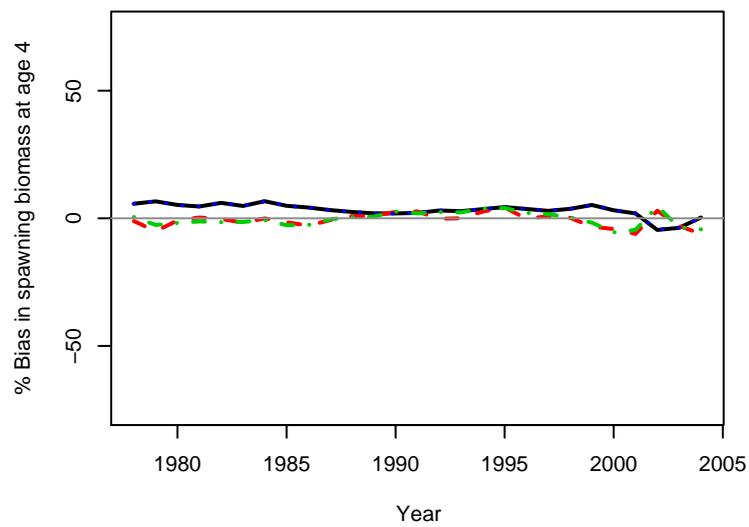
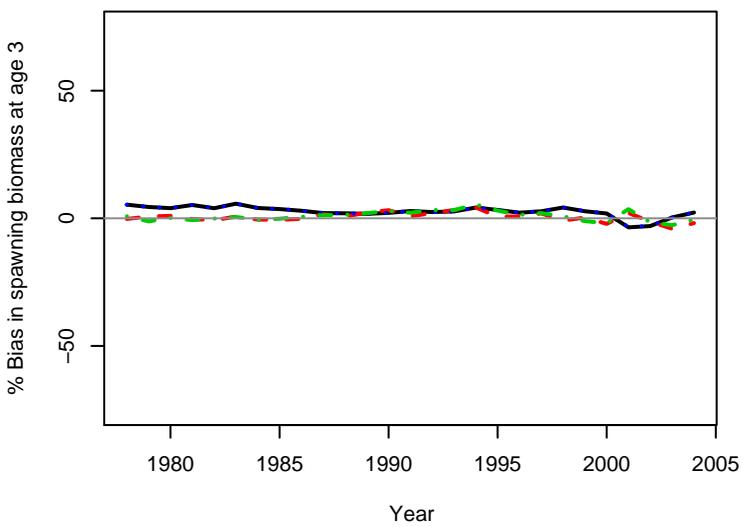
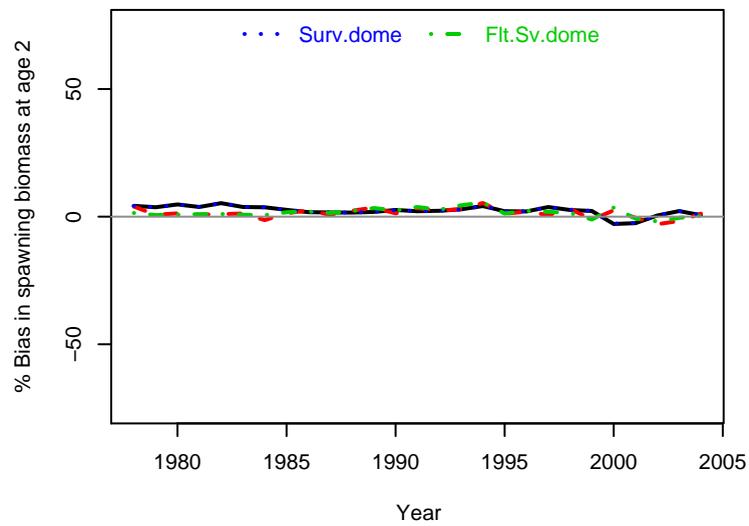
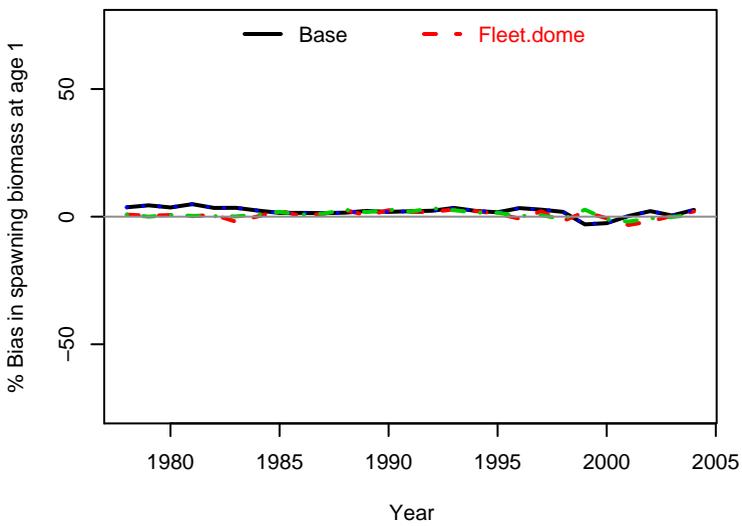


Figure GBCOD.9 Bias in spawning biomass at age for all simulated data sets using VPA.

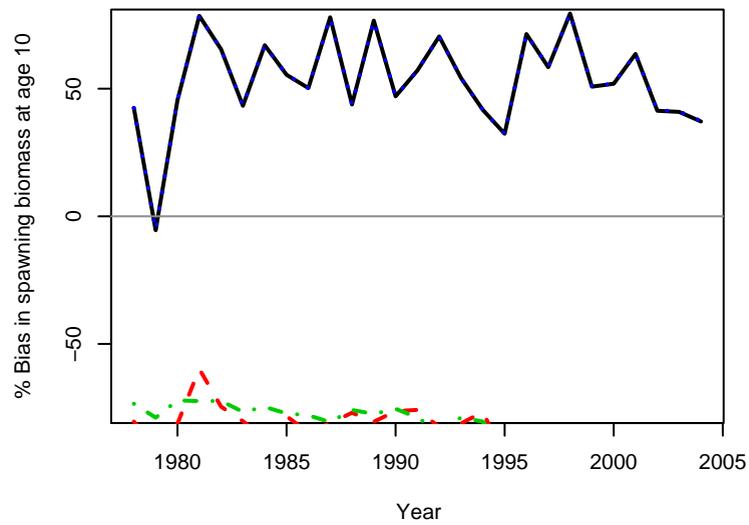
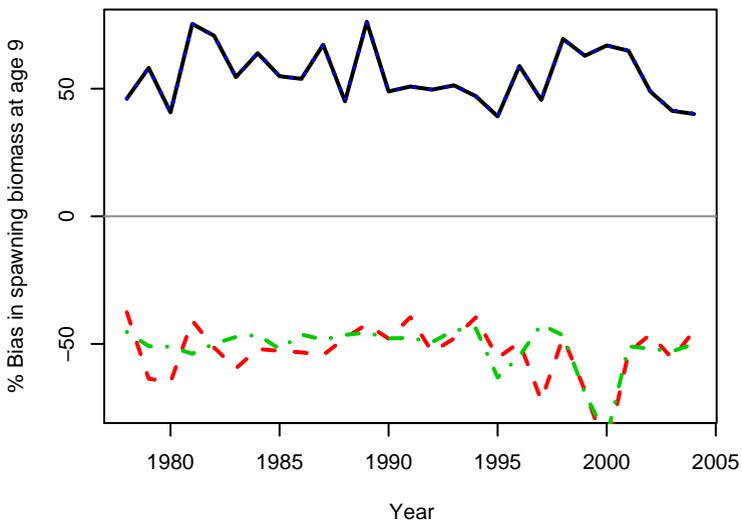
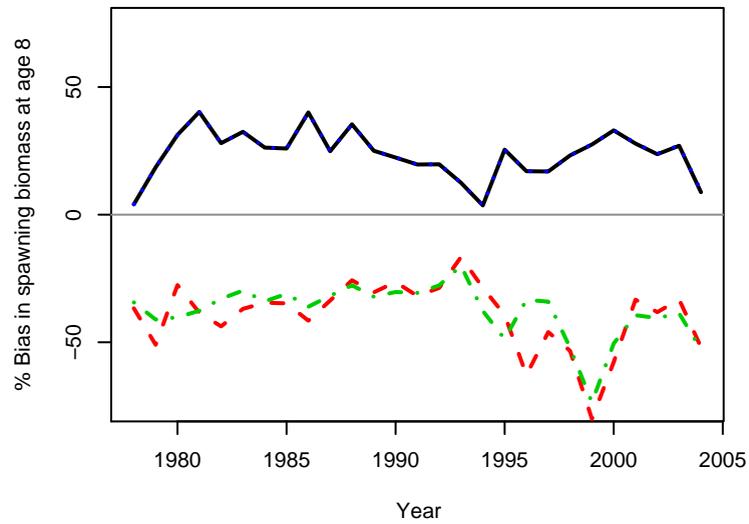
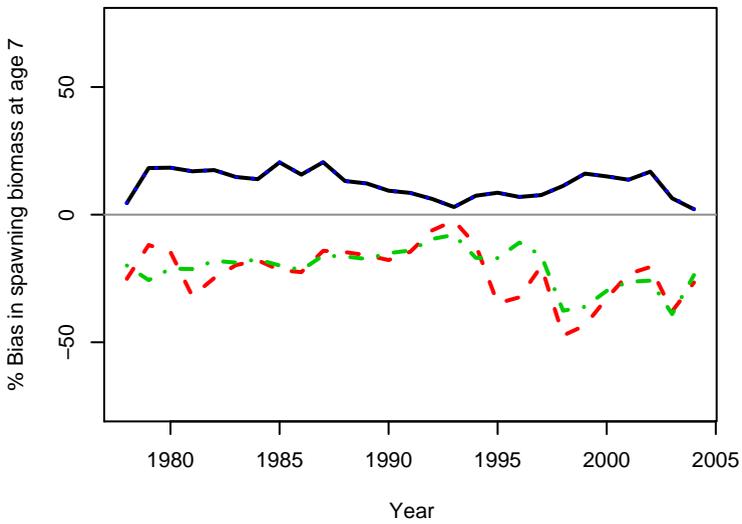


Figure GBCOD.9 (cont.)

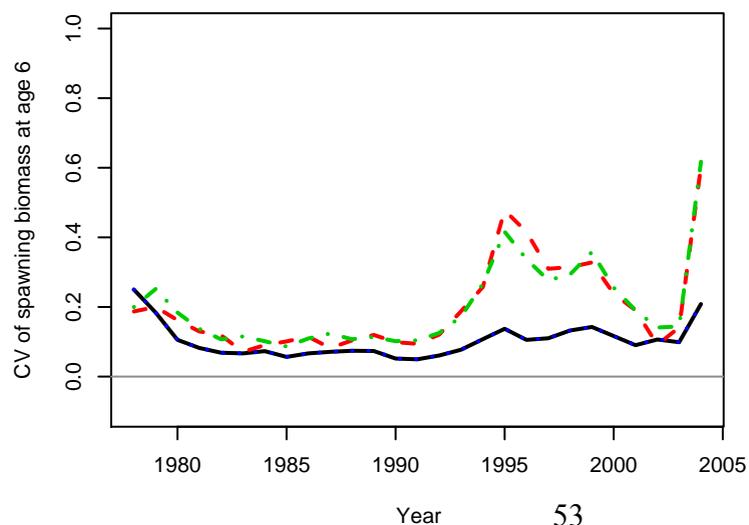
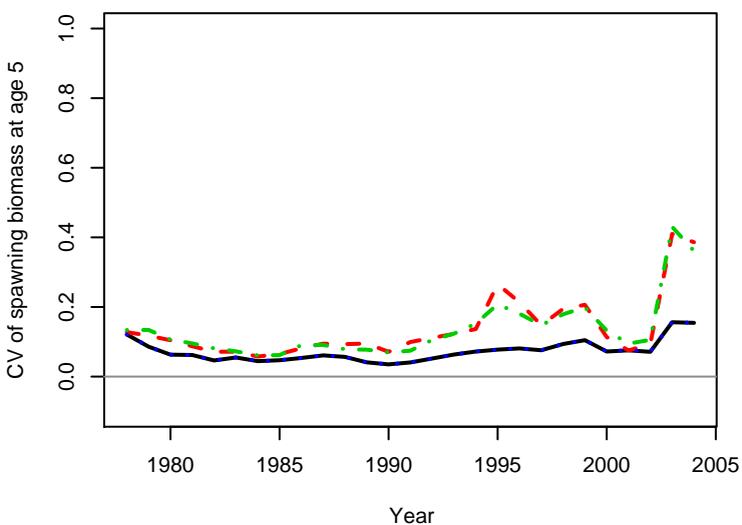
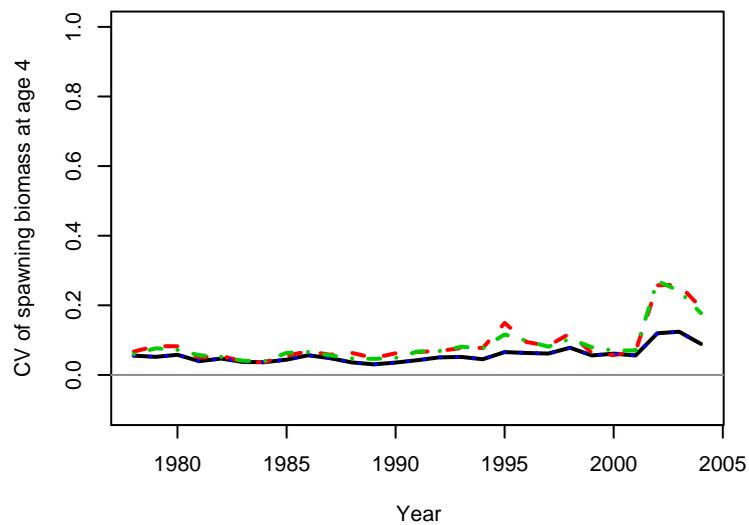
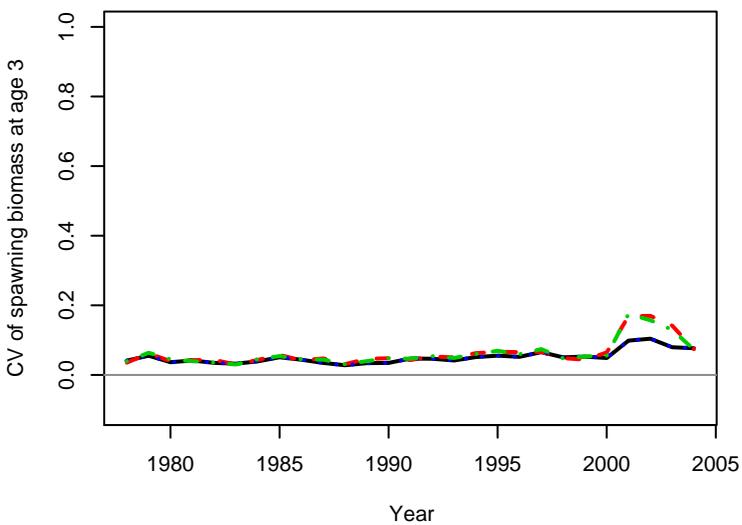
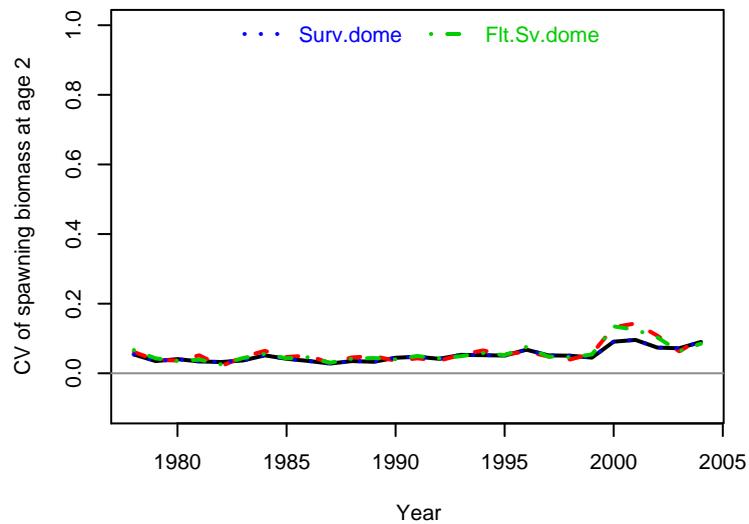
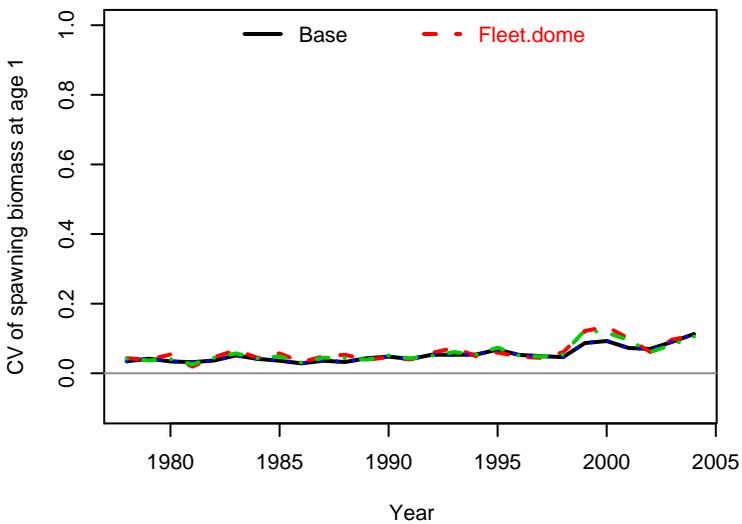


Figure GBCOD.10 CV of spawning biomass at age for all simulated data sets using VPA.

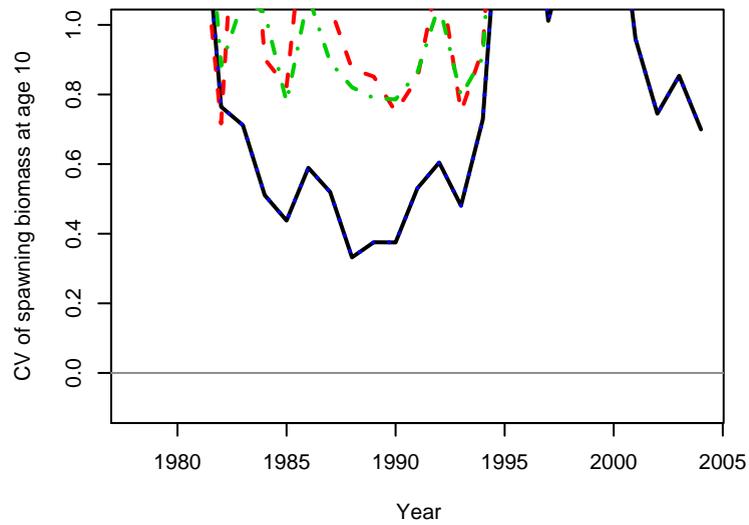
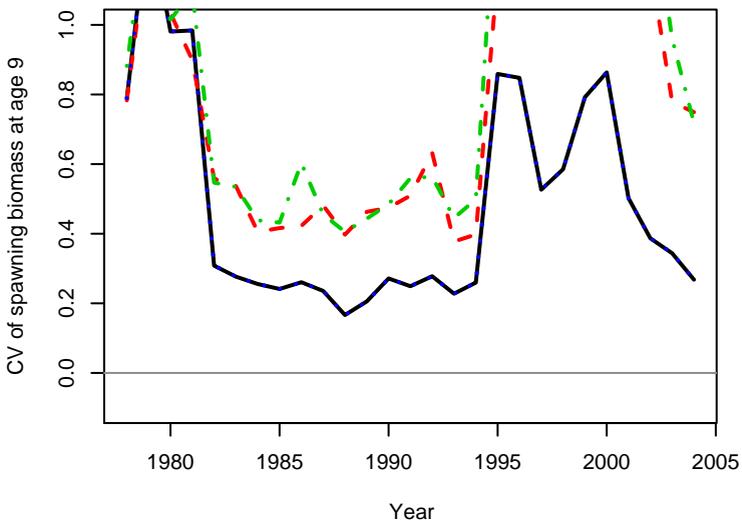
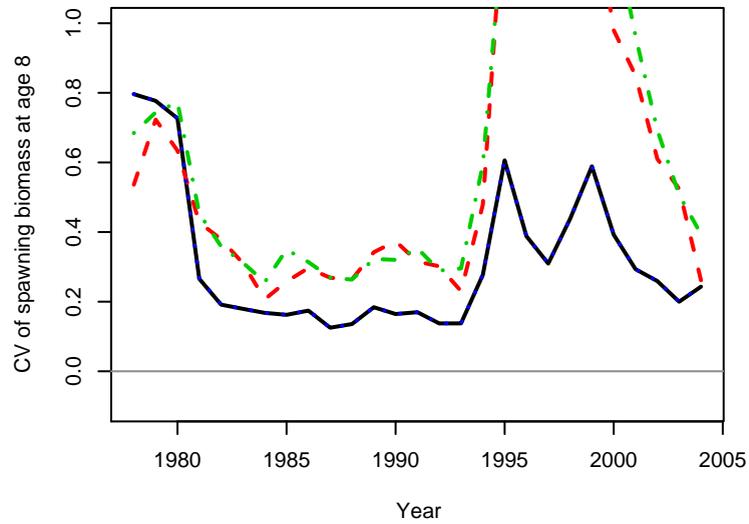
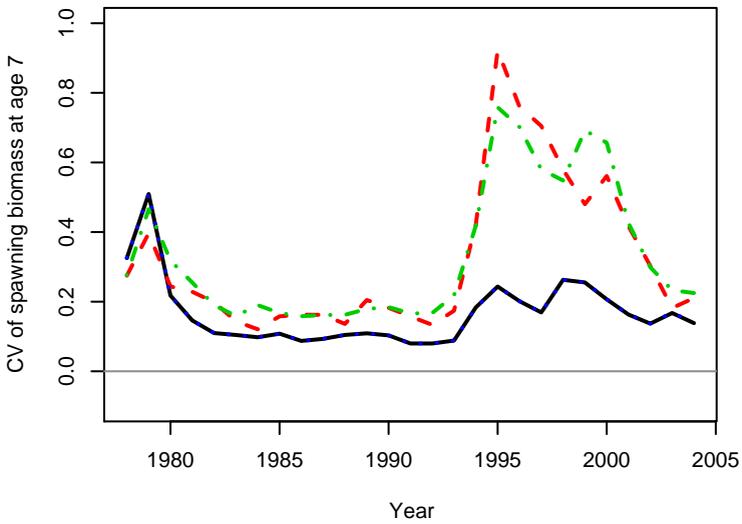


Figure GBCOD.10 (cont.)

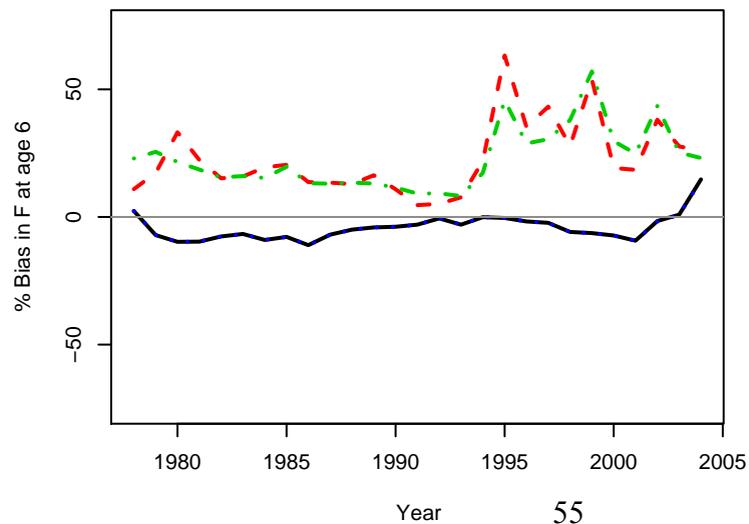
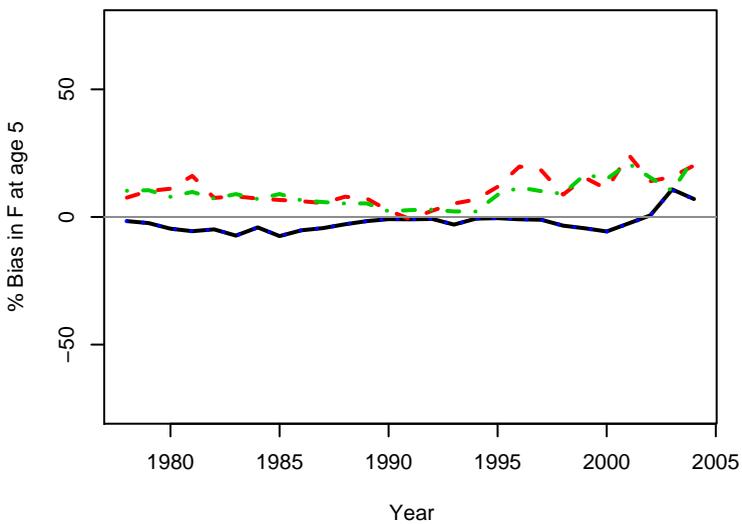
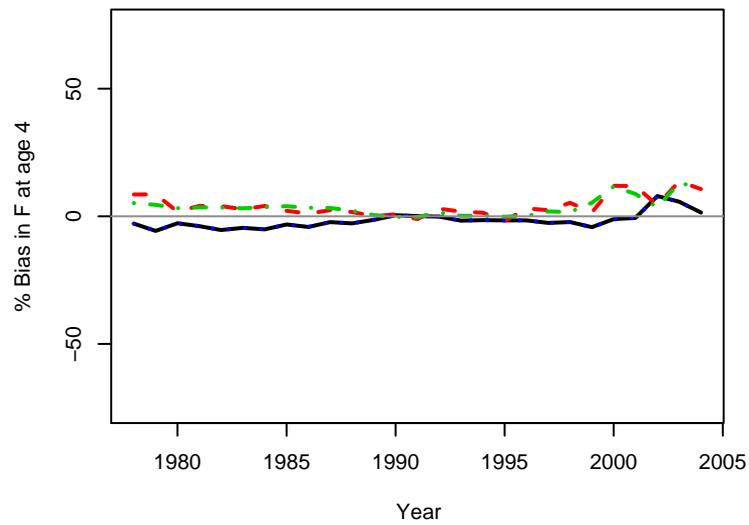
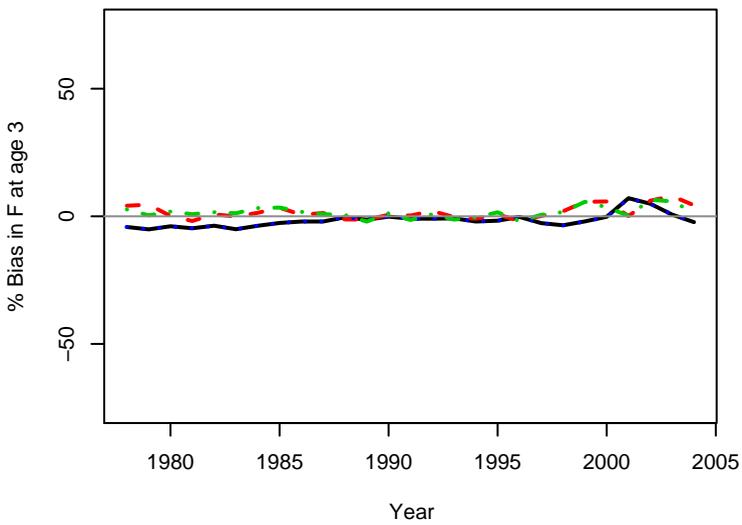
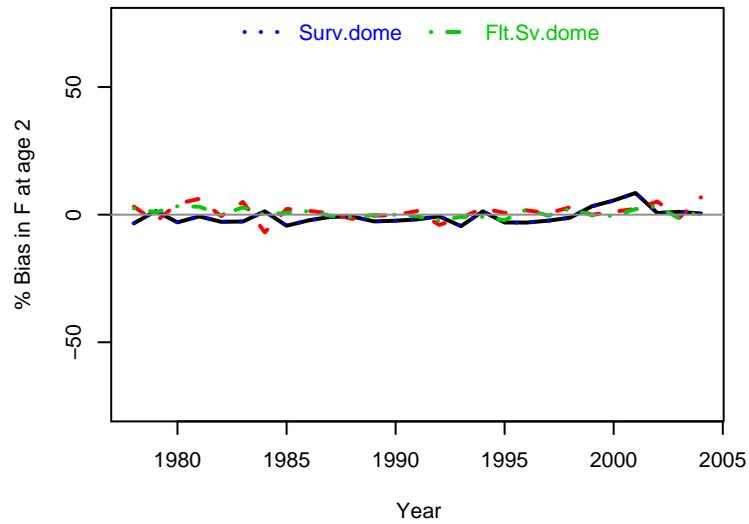
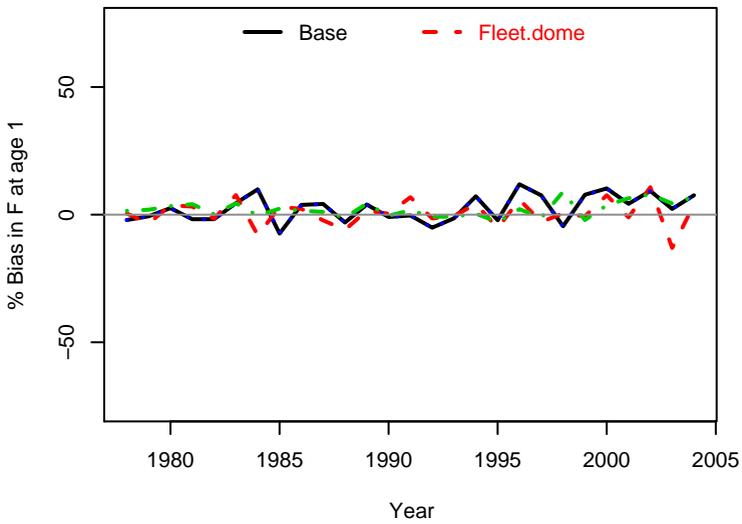
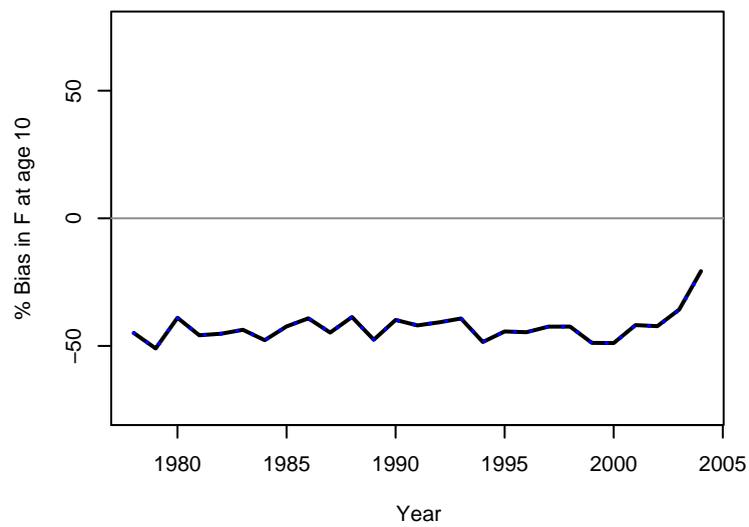
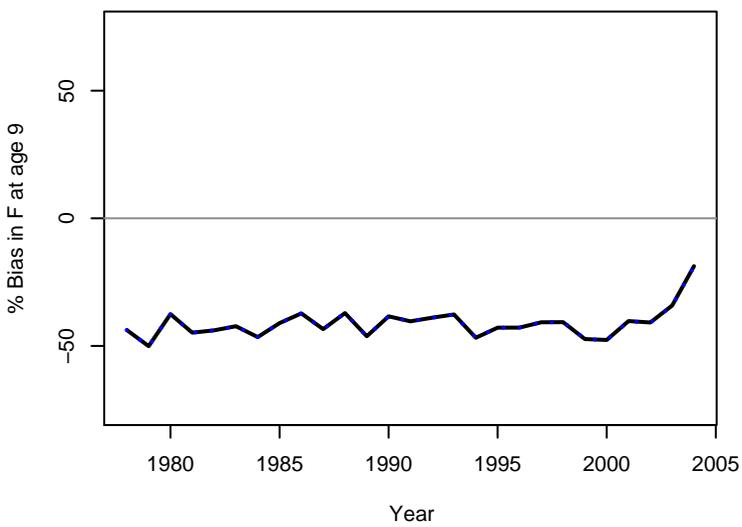
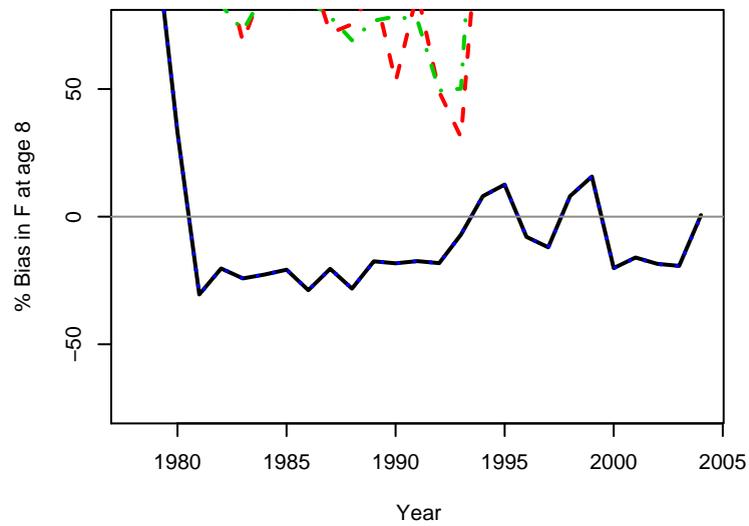
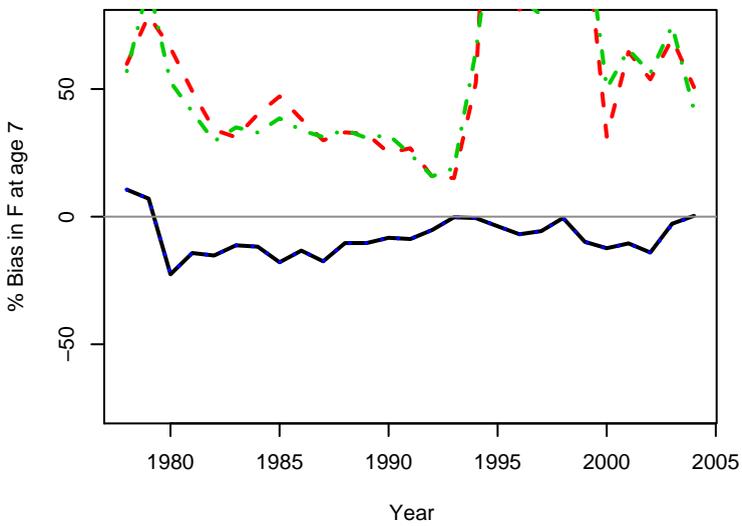


Figure GBCOD.11 Bias in F at age for all simulated data sets using VPA.



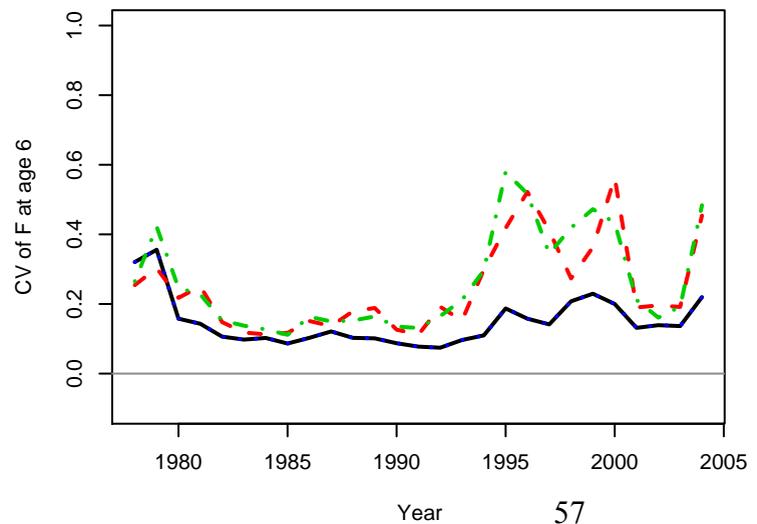
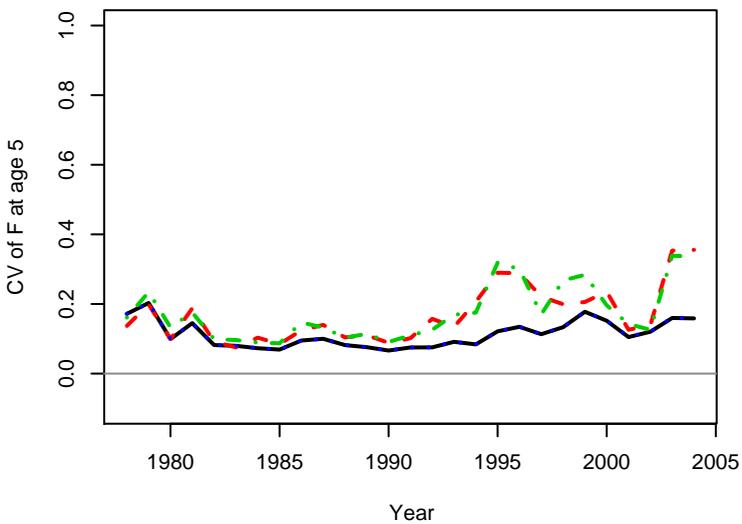
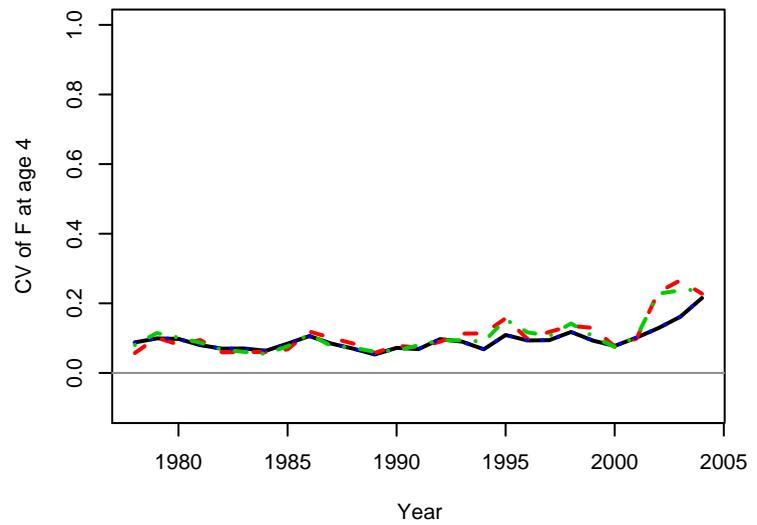
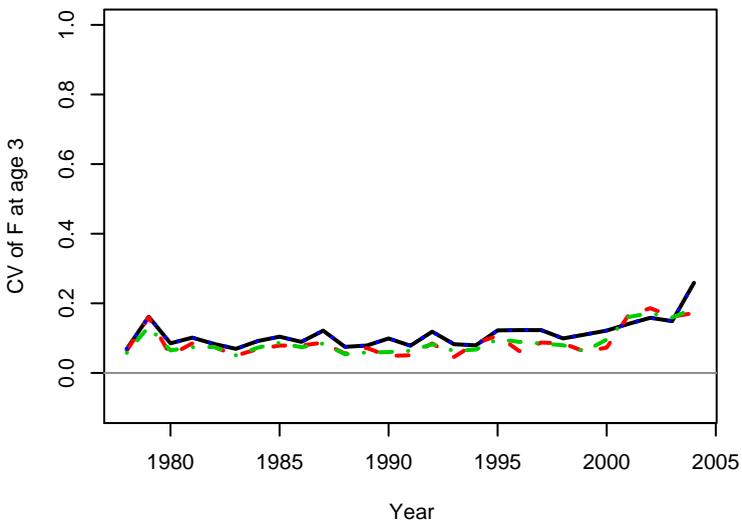
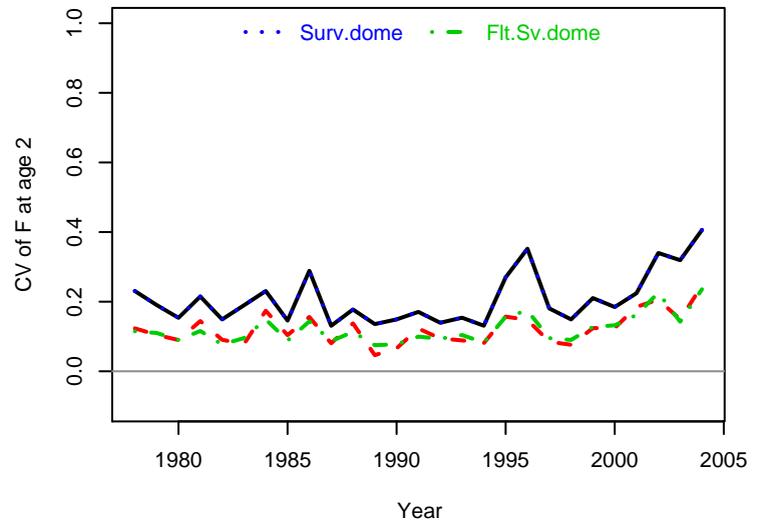
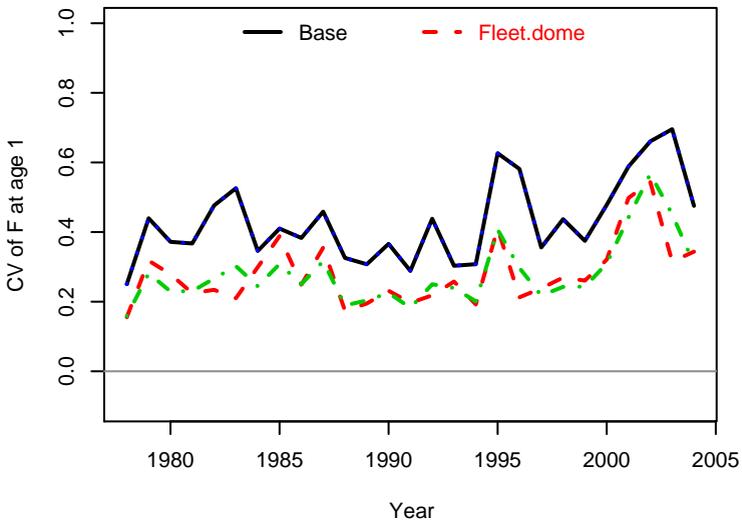
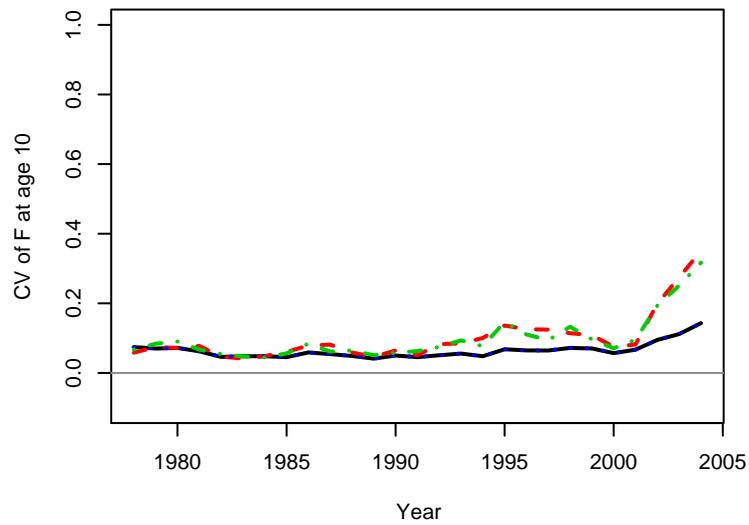
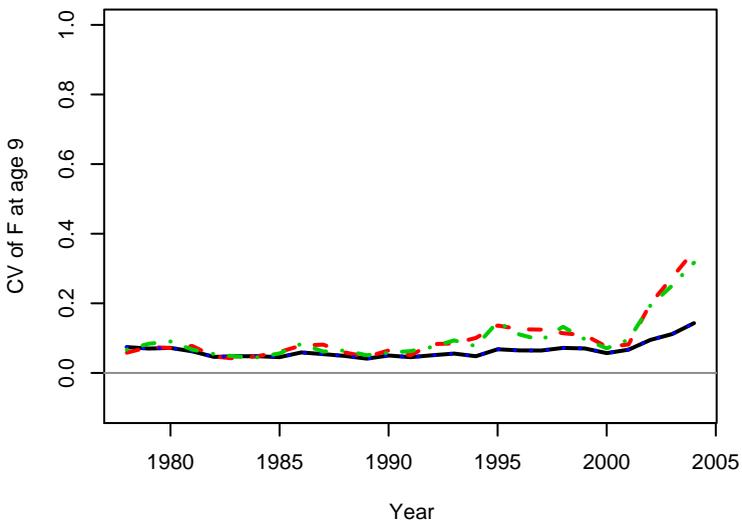
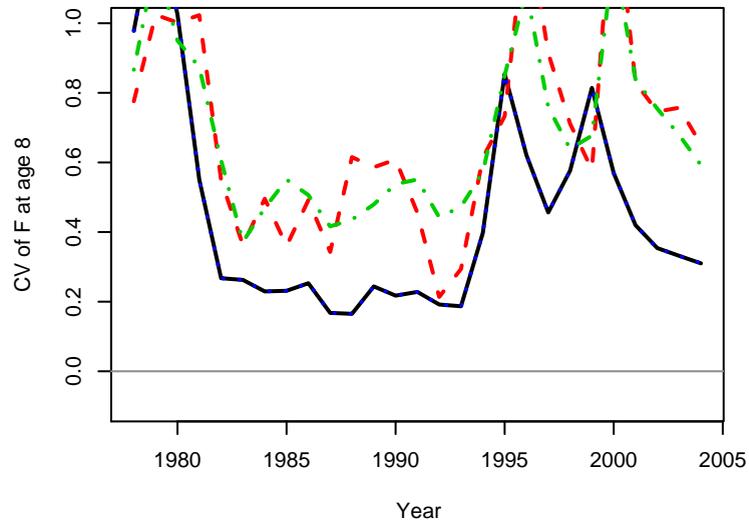
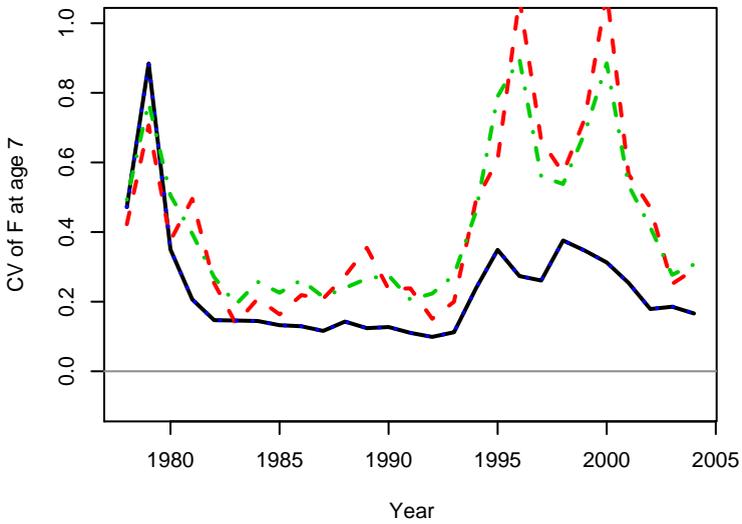


Figure GBCOD.12 CV of F at age for all simulated data sets using VPA.



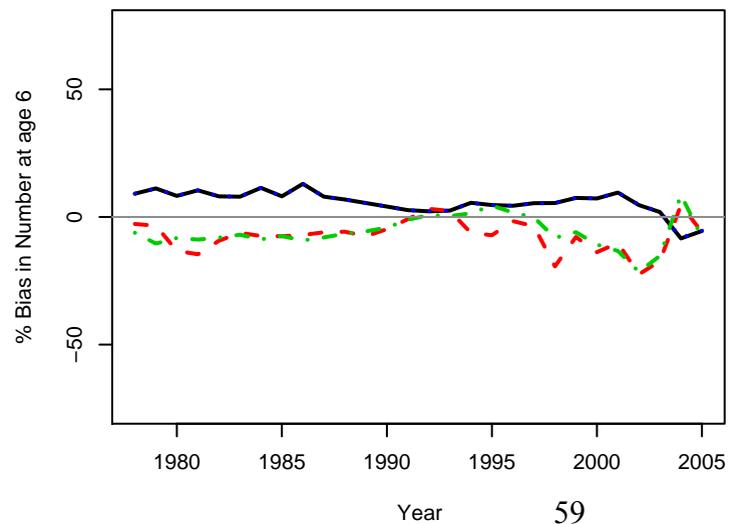
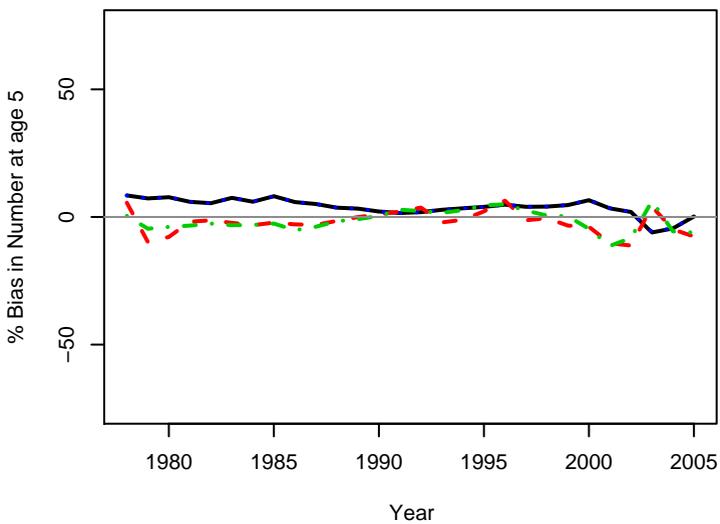
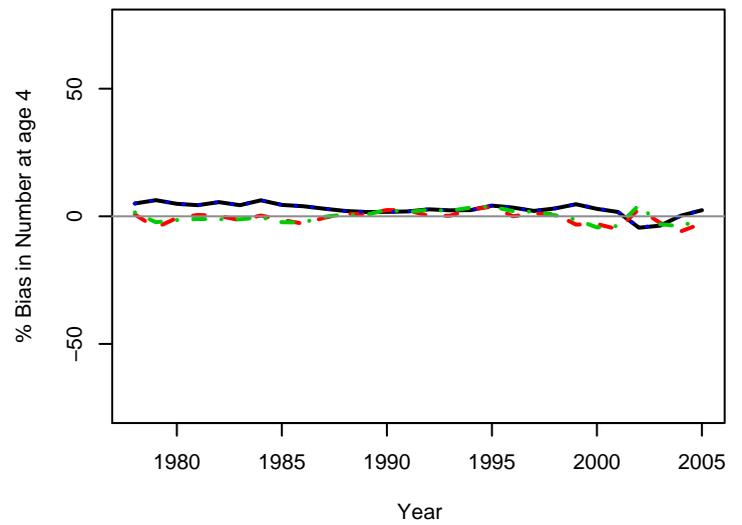
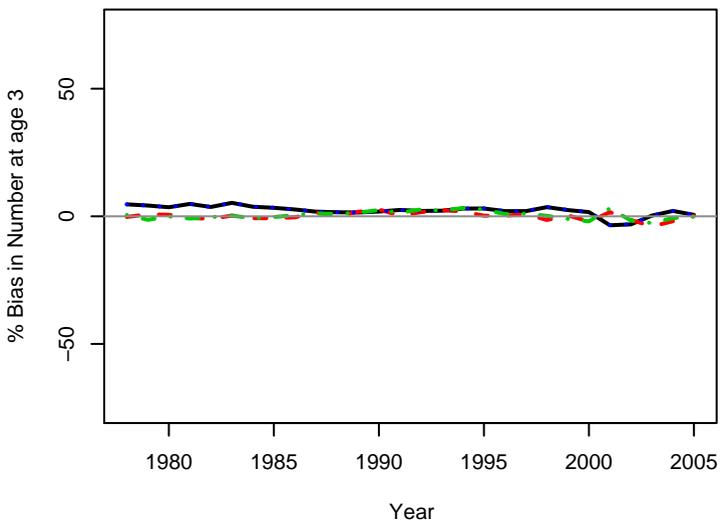
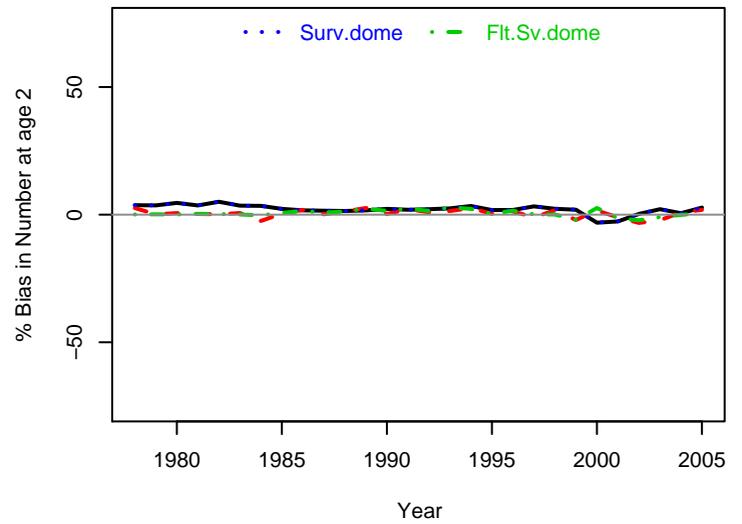
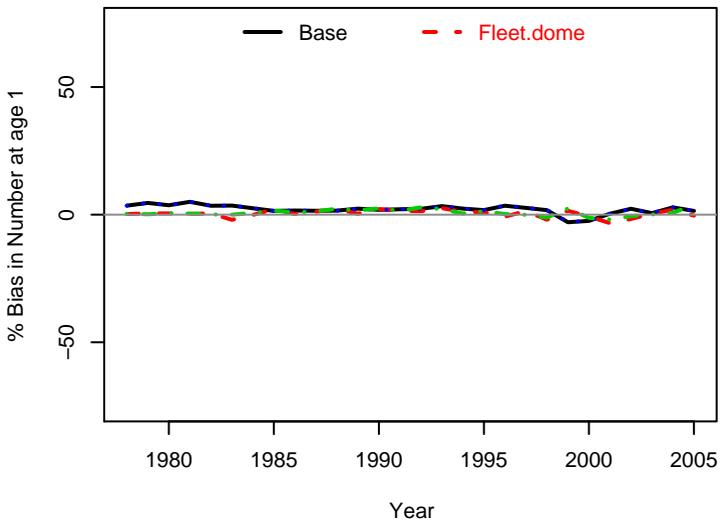


Figure GBCOD.13 Bias in number at age for all simulated data sets using VPA.

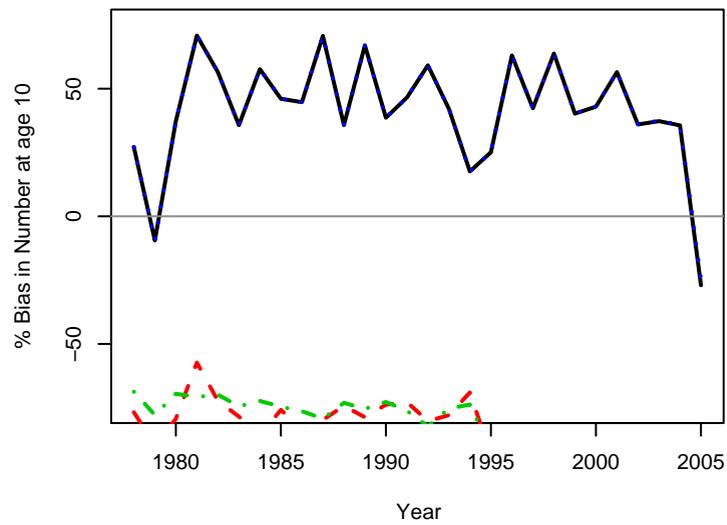
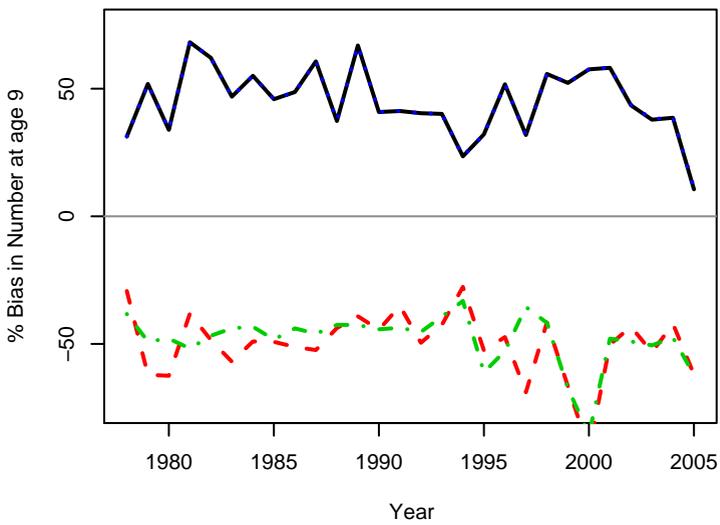
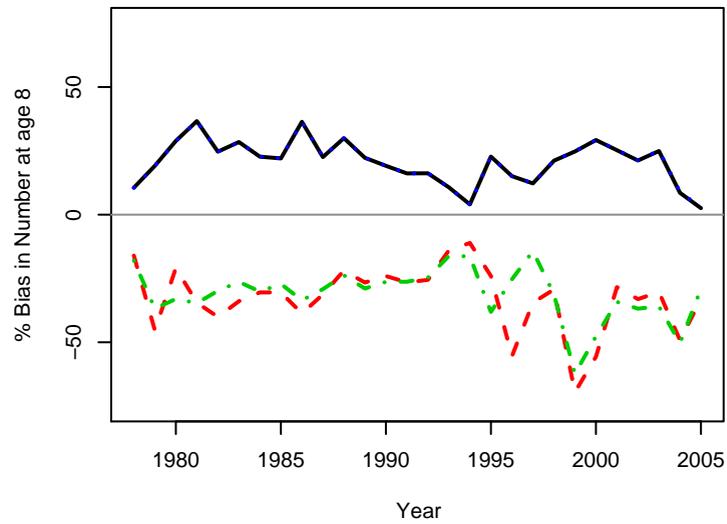
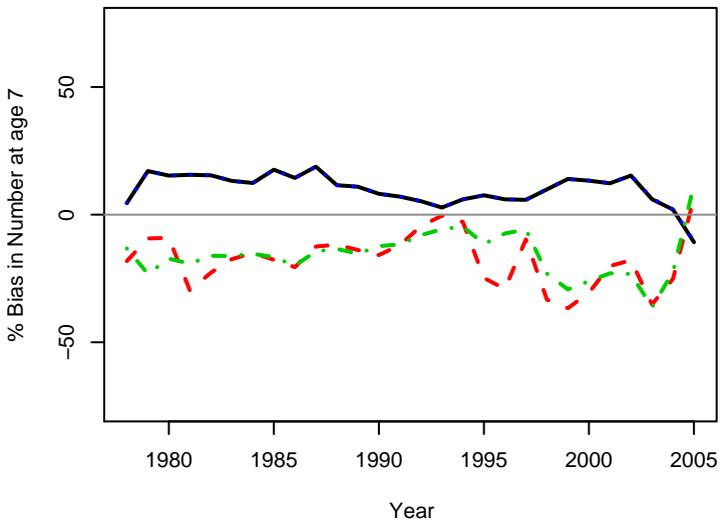


Figure GBCOD.13 (cont.)

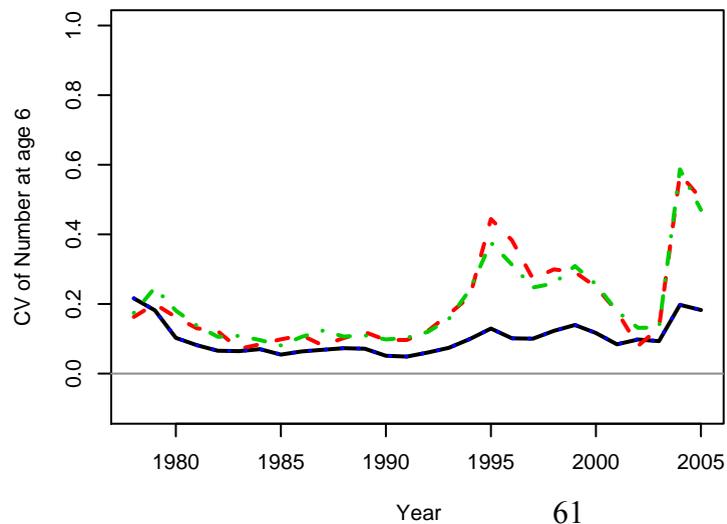
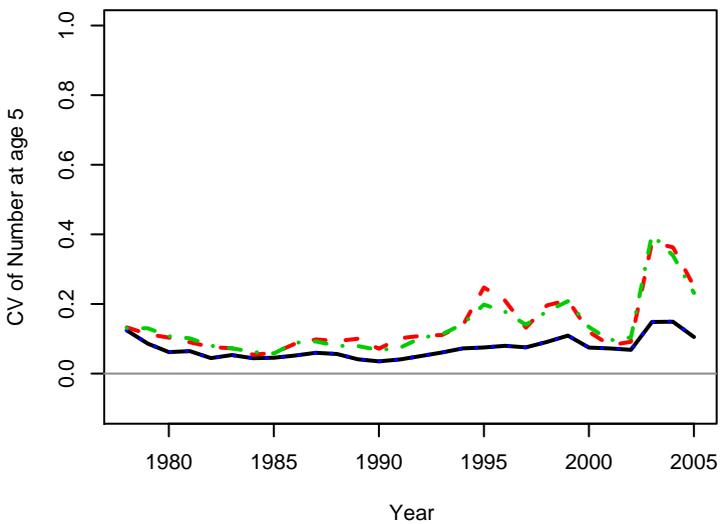
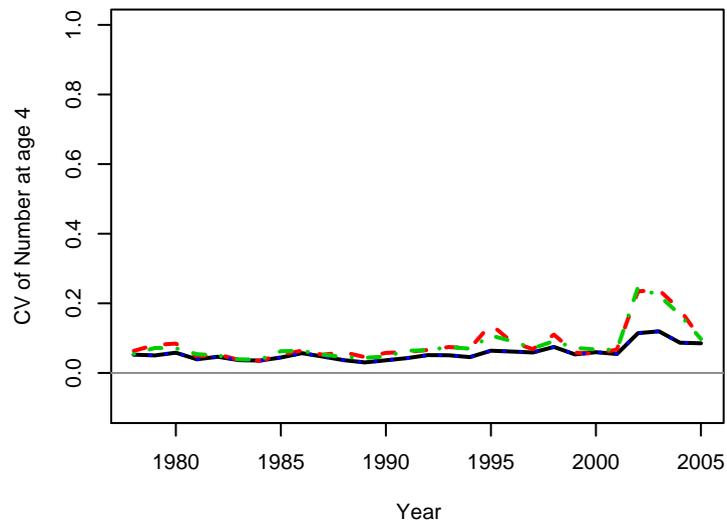
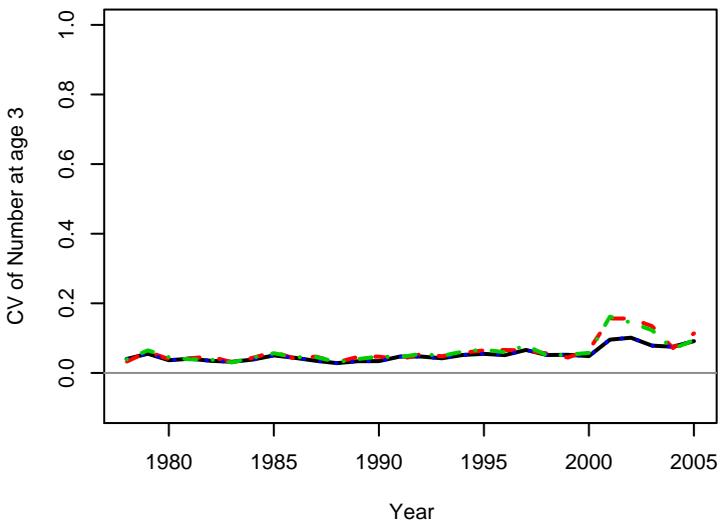
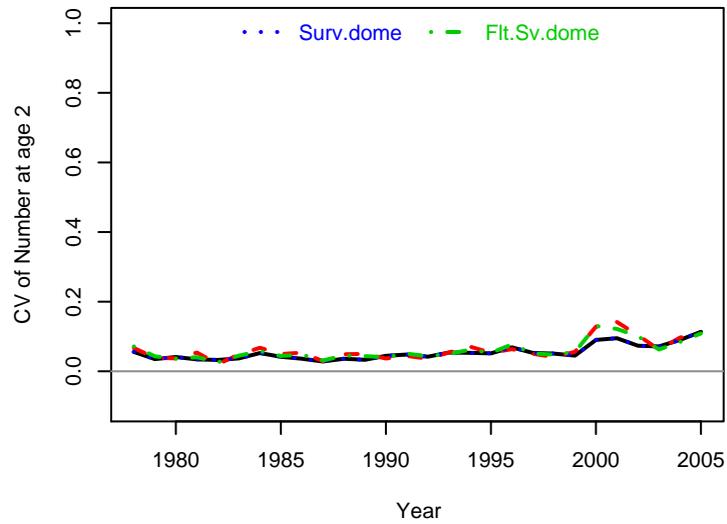
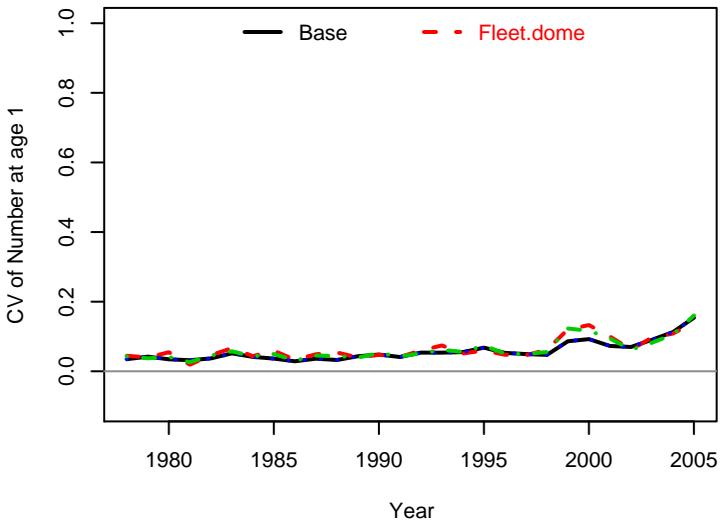


Figure GBCOD.14 CV of number at age for all simulated data sets using VPA.

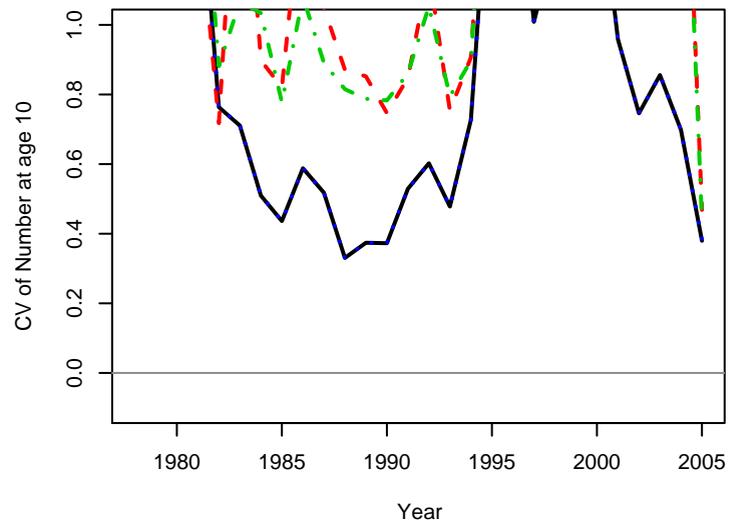
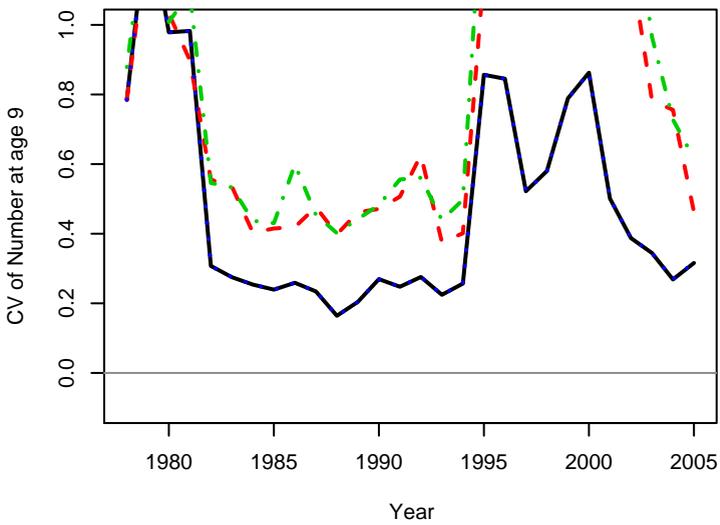
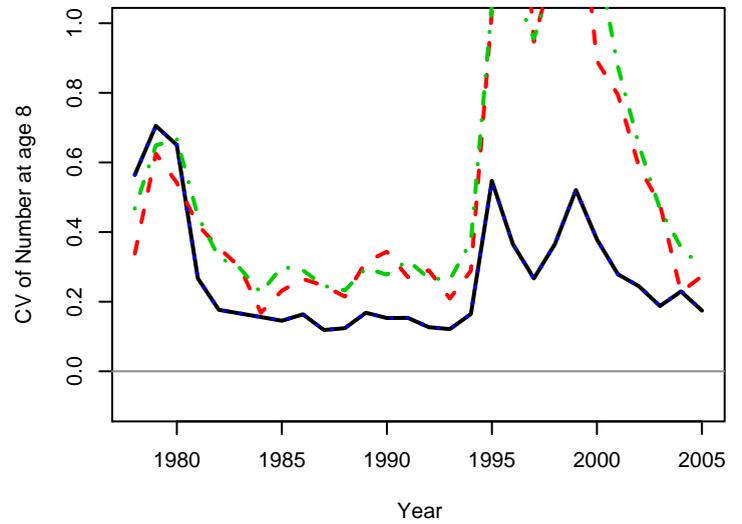
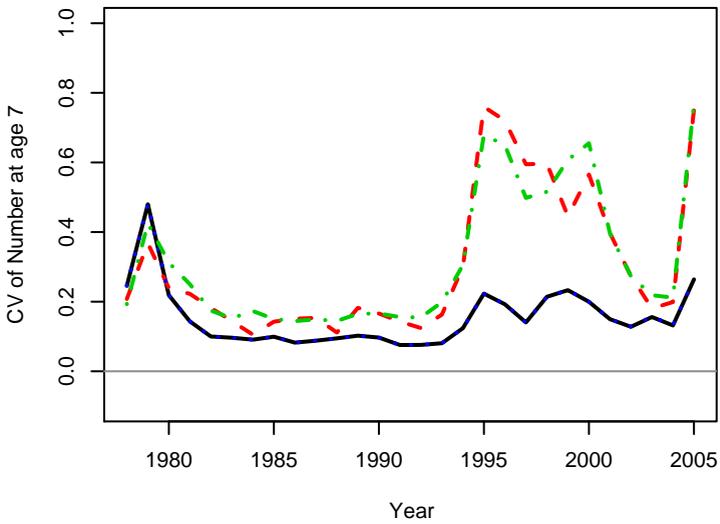


Figure GBCOD.14 (cont.)

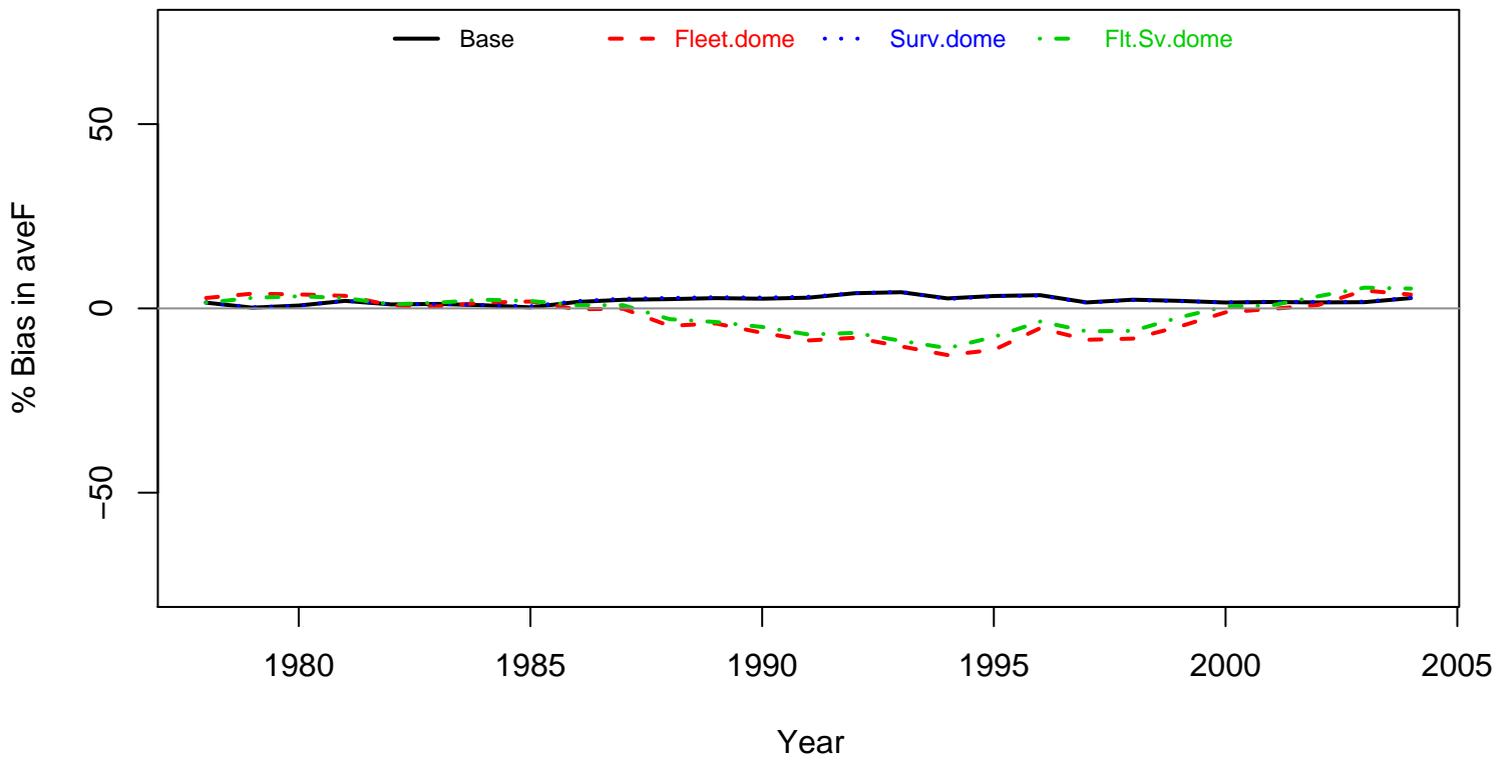
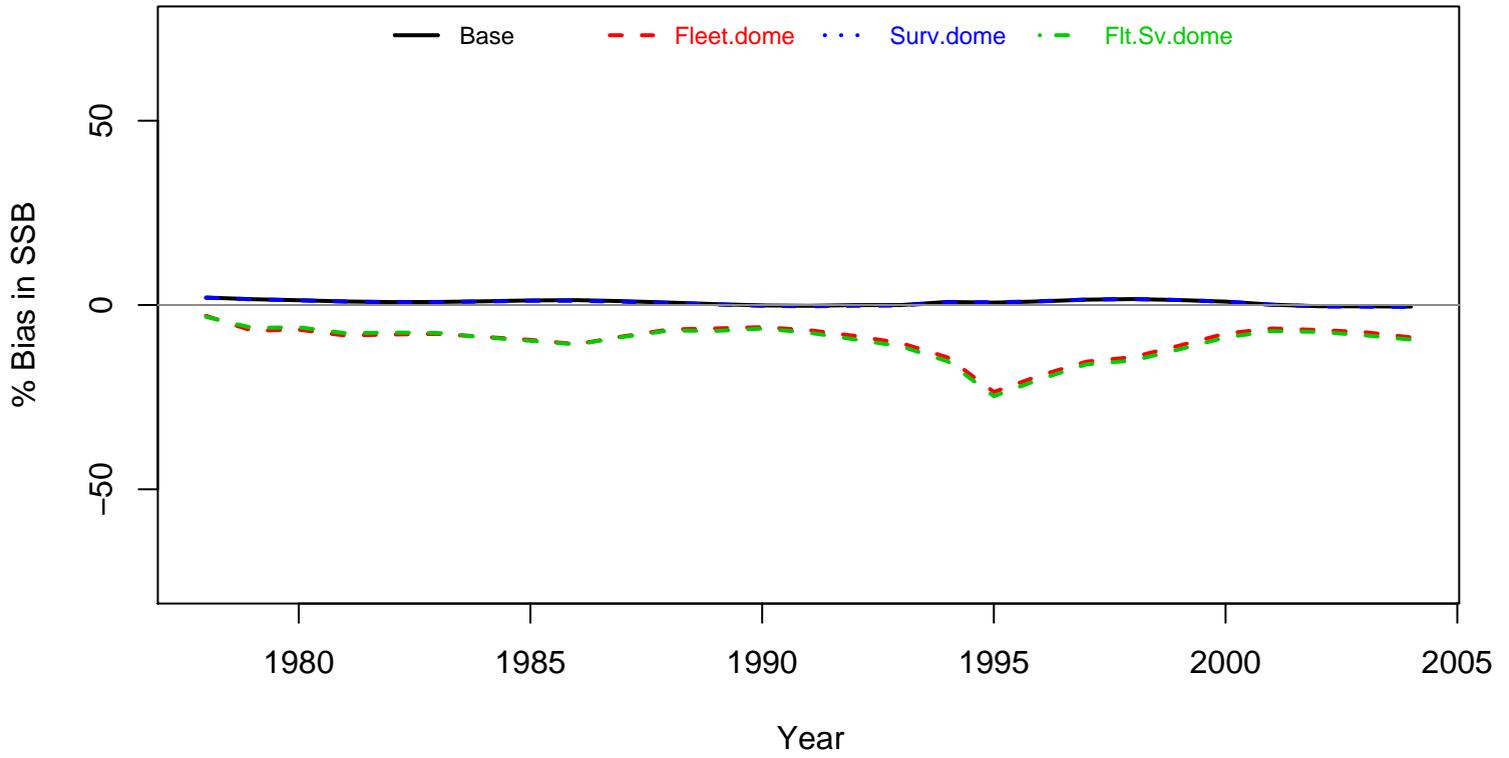


Figure GBCOD.15 Bias in spawning biomass and average F for all simulated data sets using ASAP.

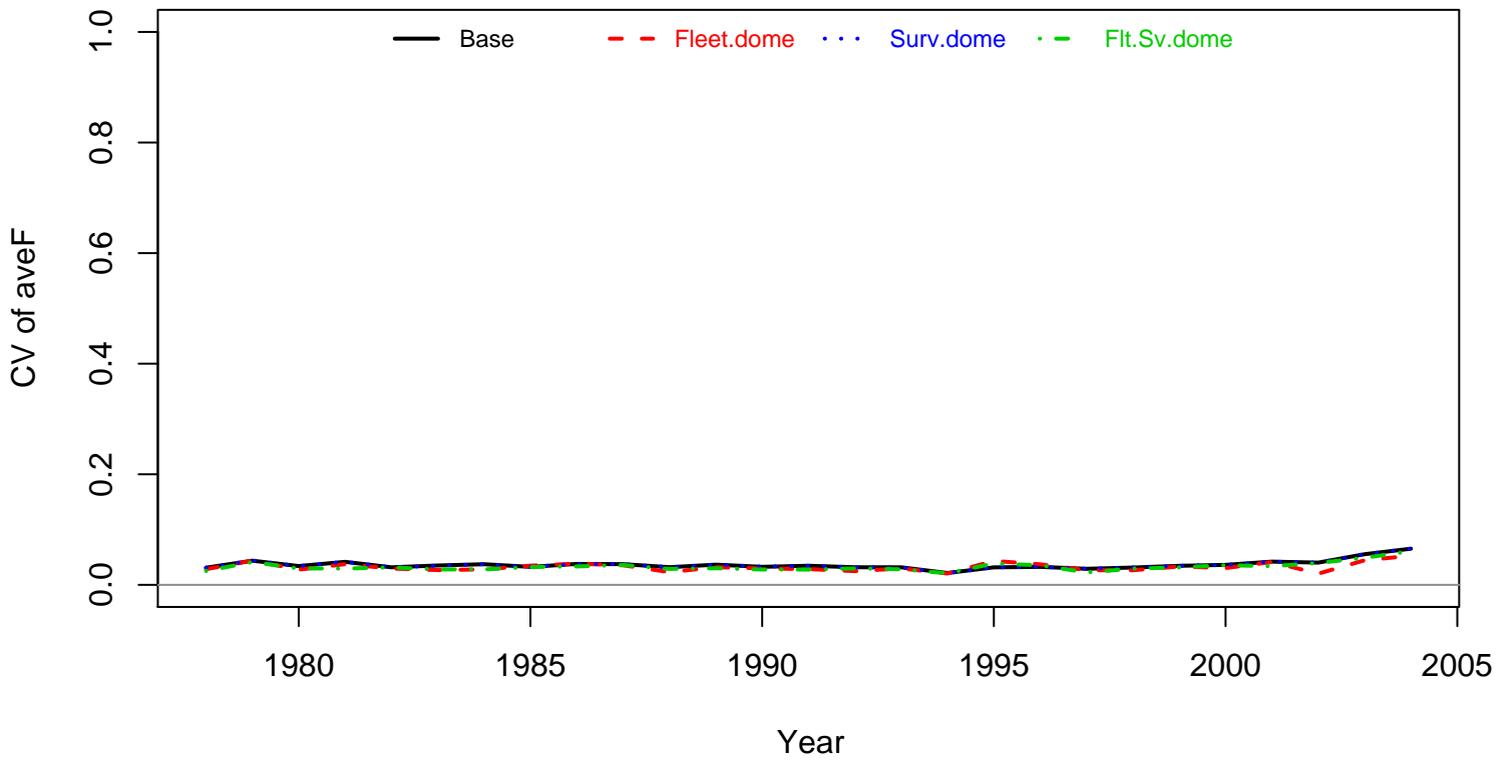
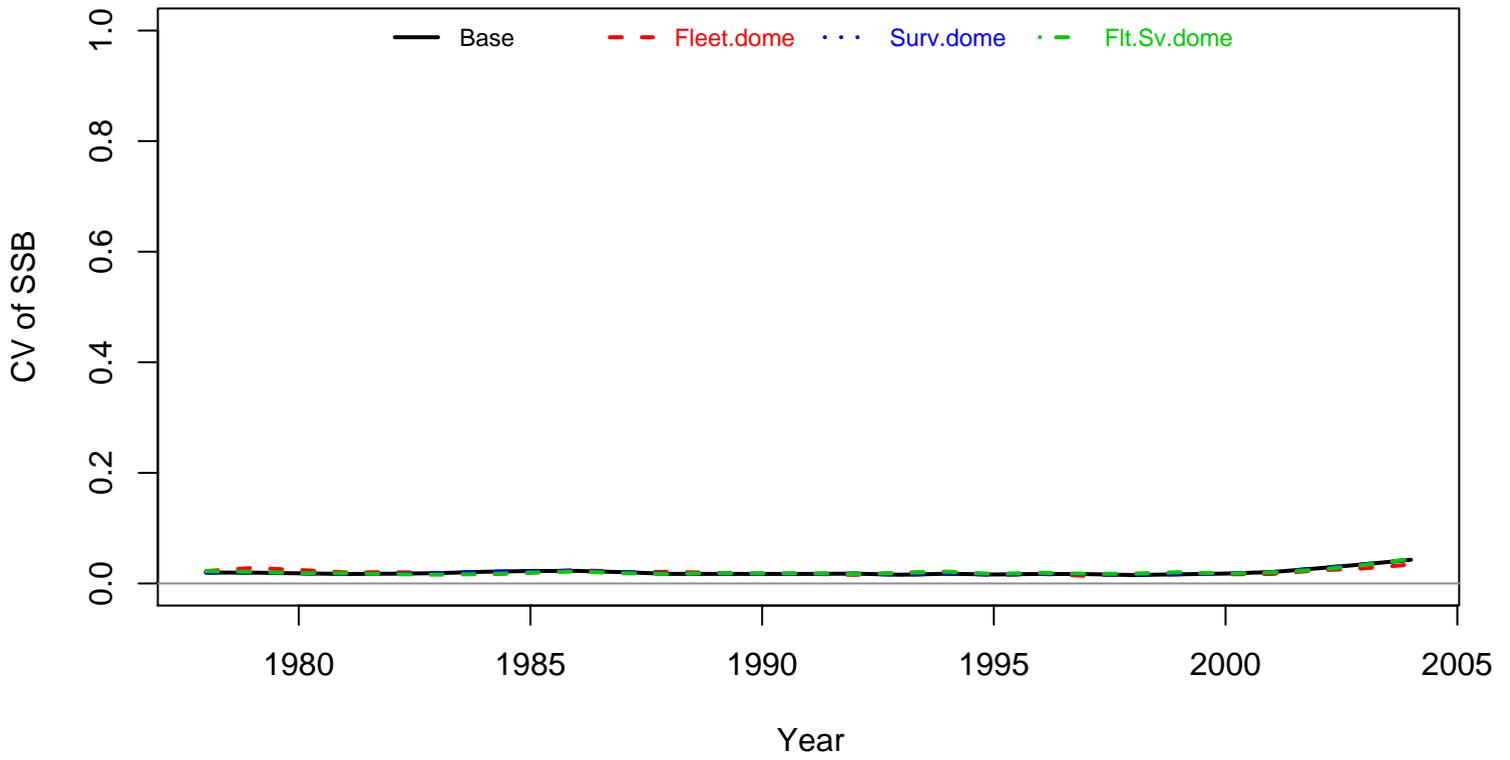


Figure GBCOD.16 CV of spawning biomass and average F for all simulated data sets using ASAP.

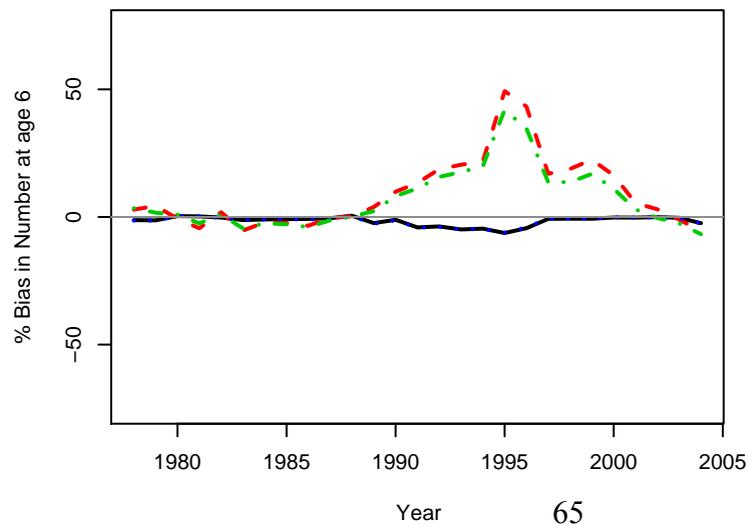
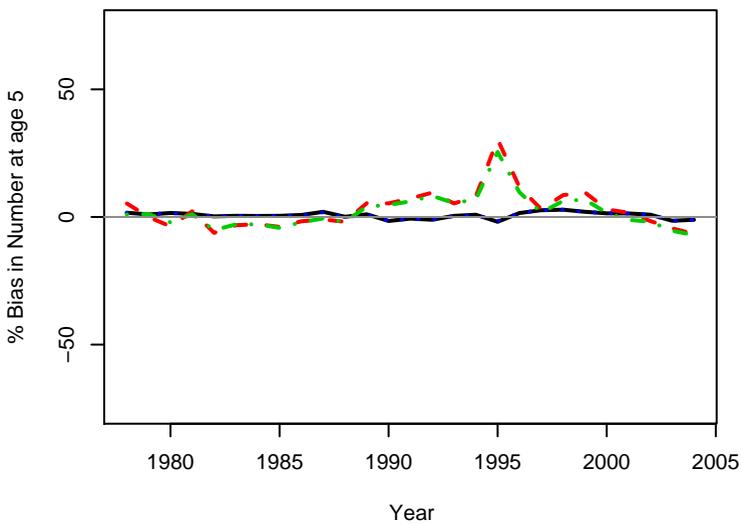
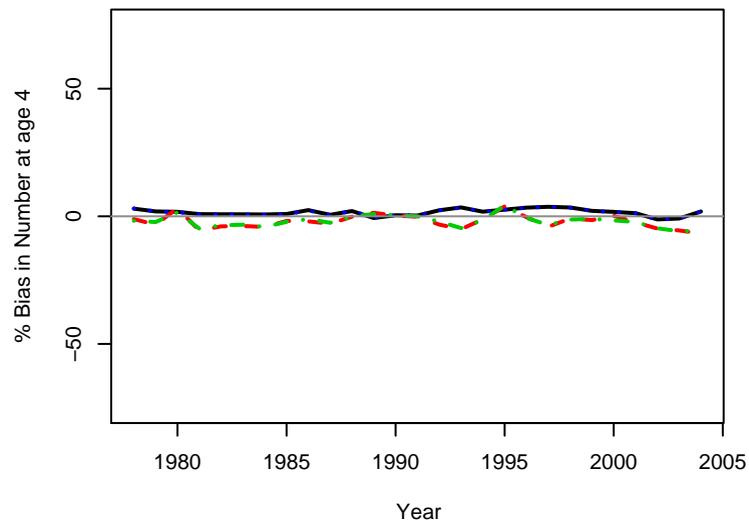
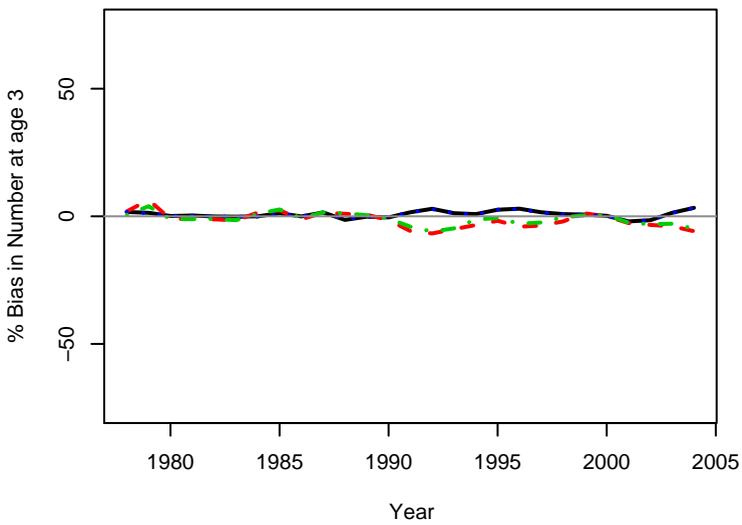
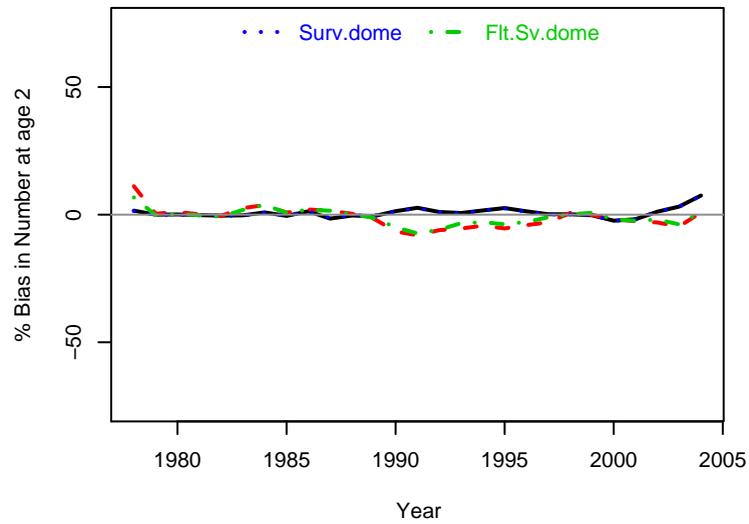
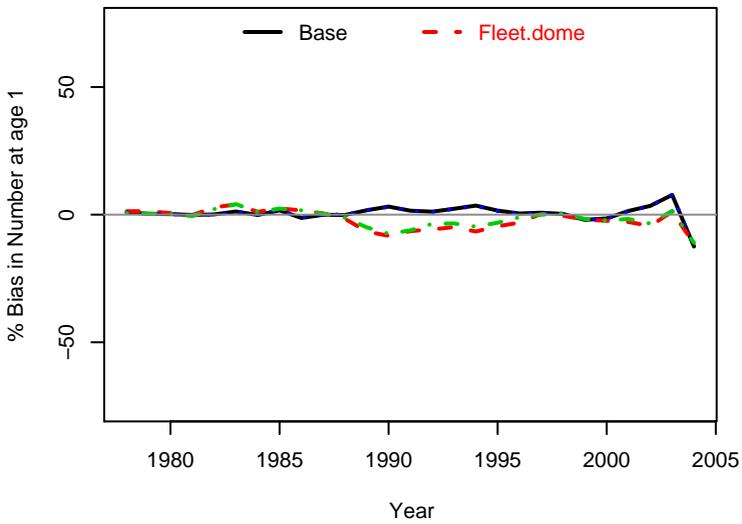


Figure GBCOD.17 Bias in number at age for all simulated data sets using ASAP.

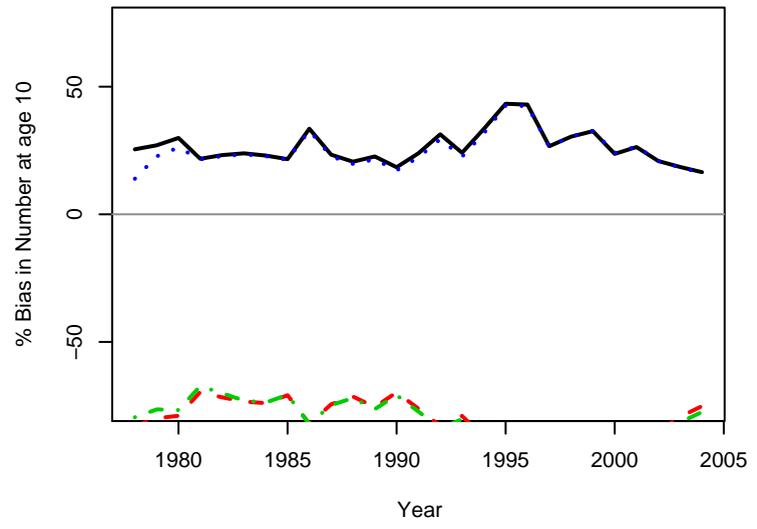
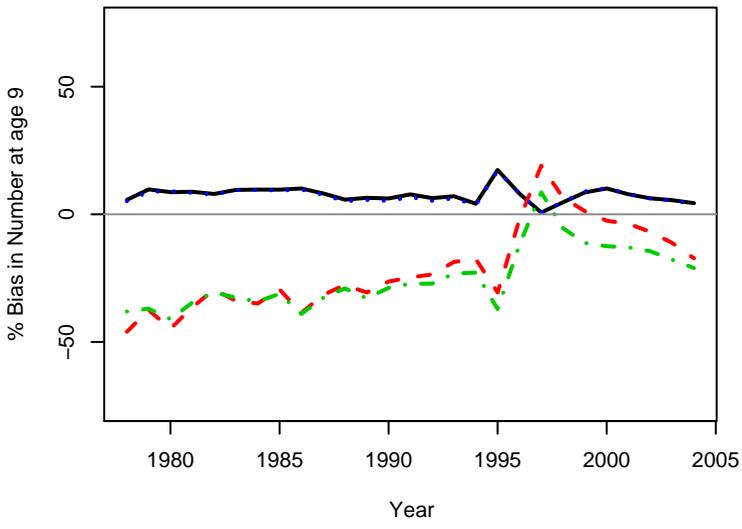
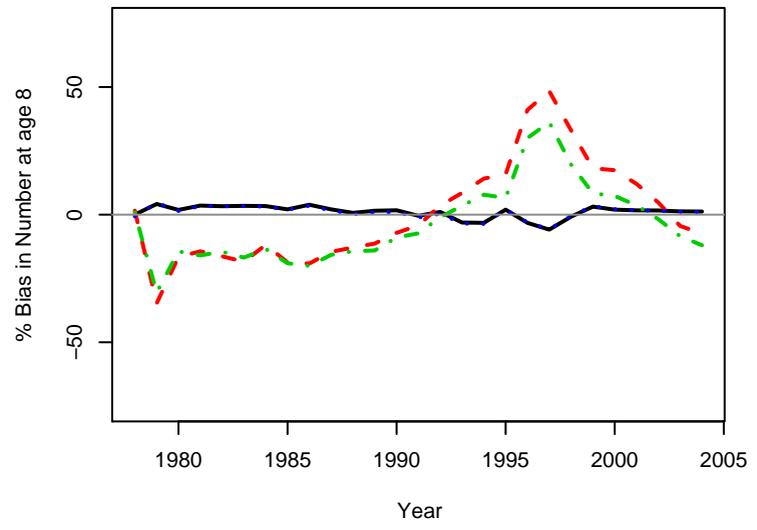
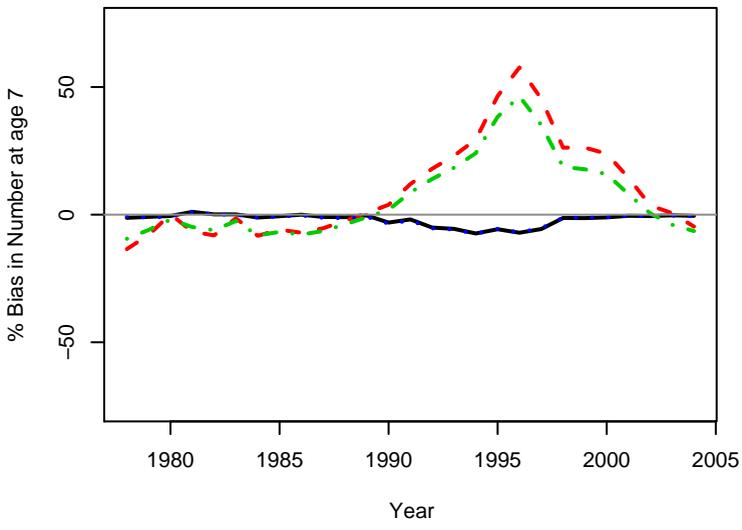


Figure GBCOD.17 (cont.)

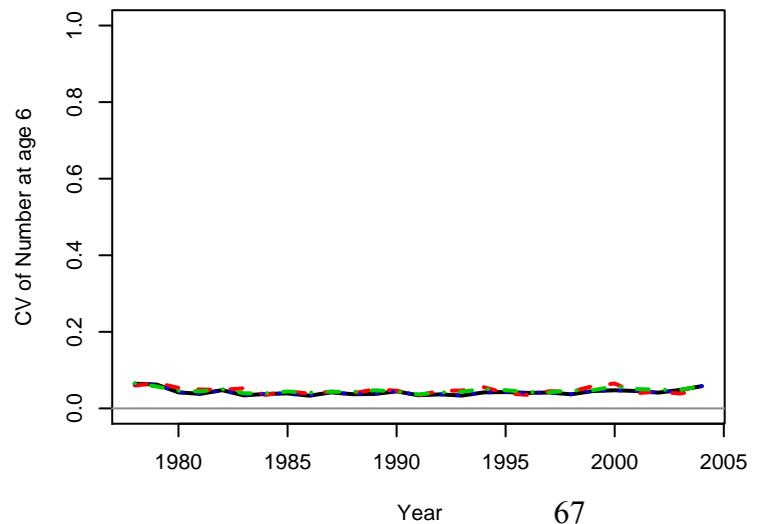
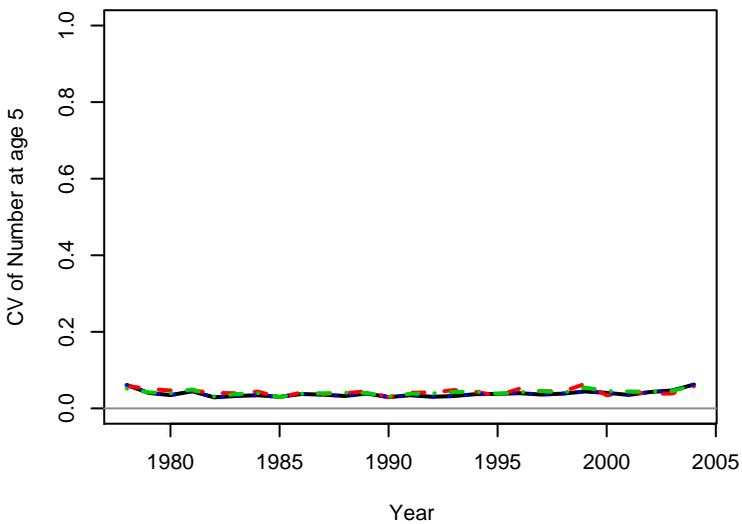
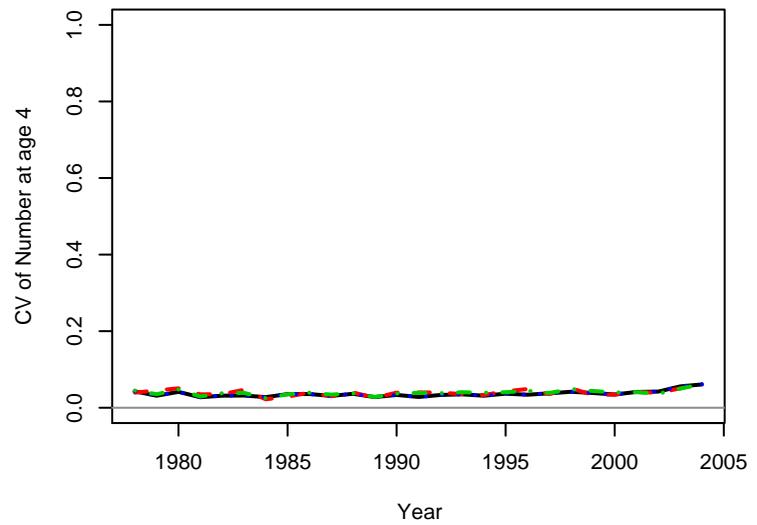
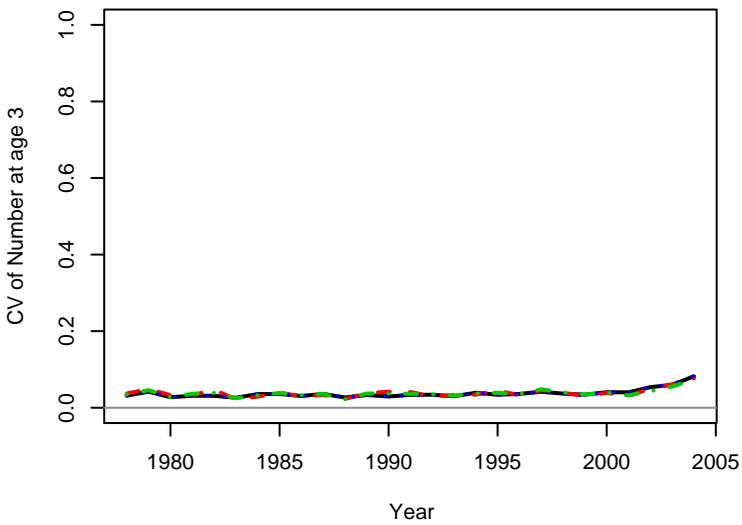
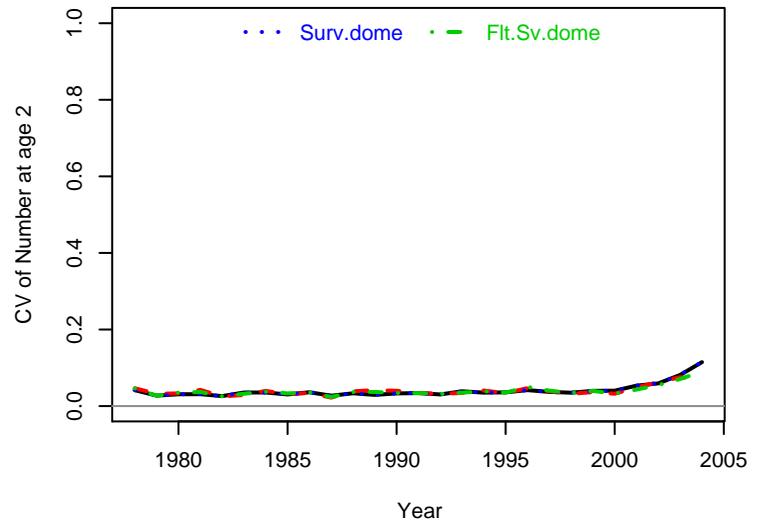
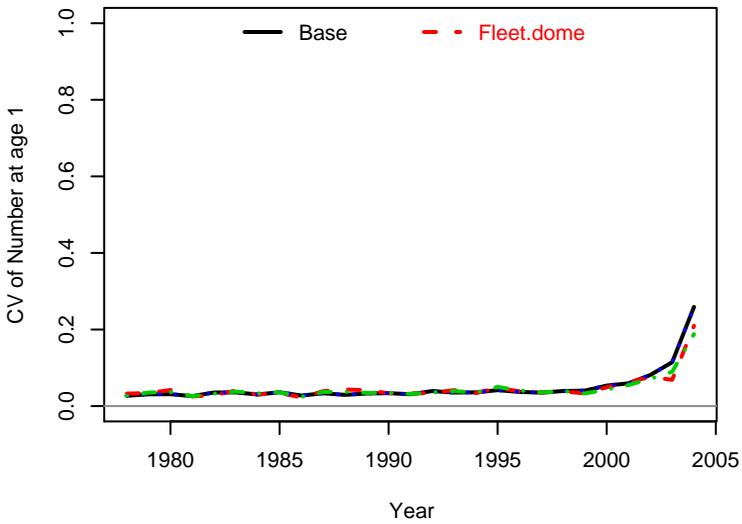


Figure GBCOD.18 CV of number at age for all simulated data sets using ASAP.

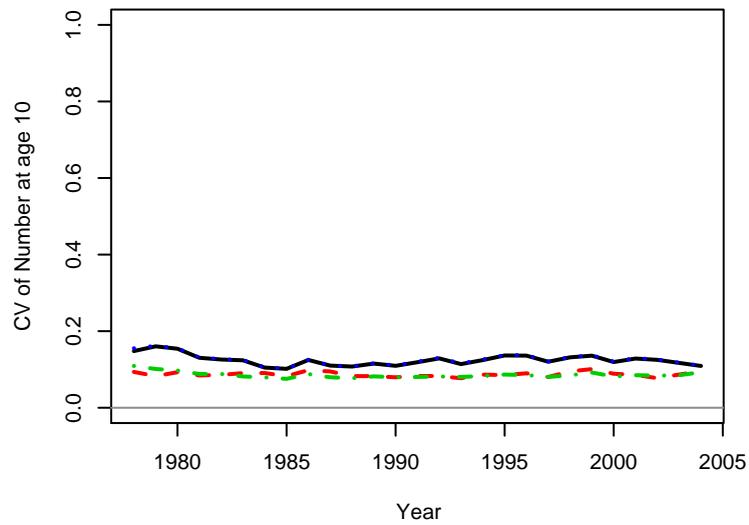
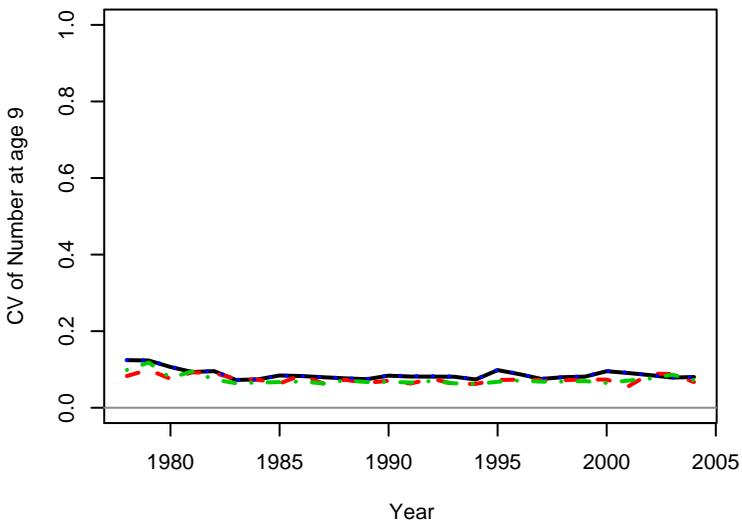
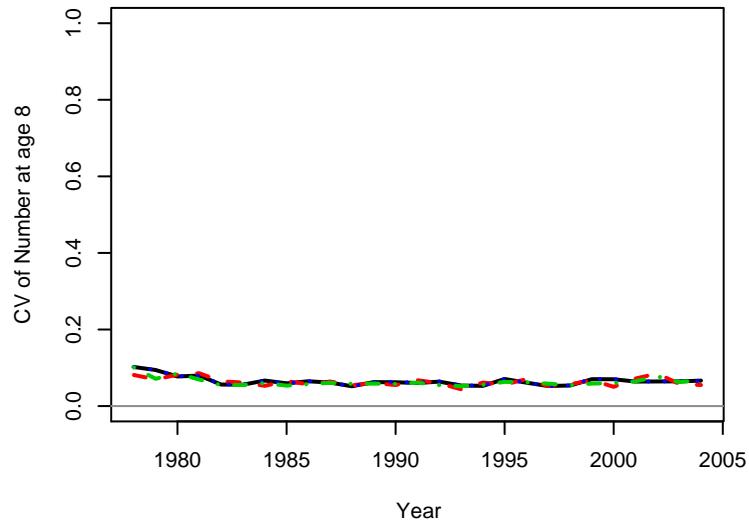
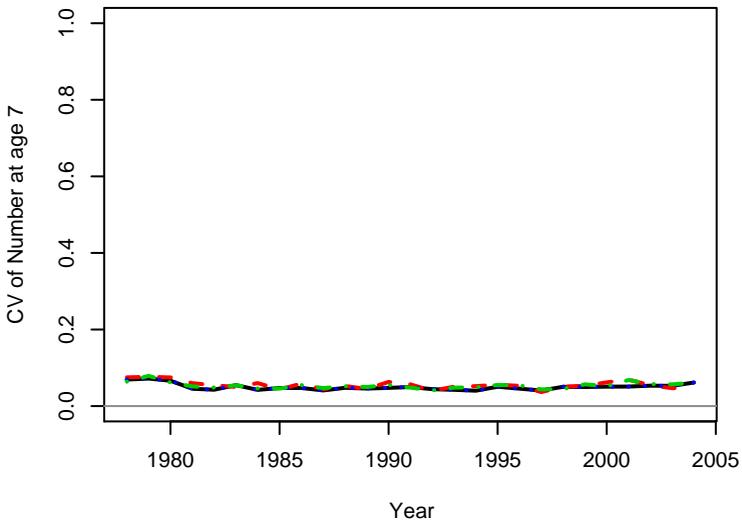


Figure GBCOD.18 (cont.)

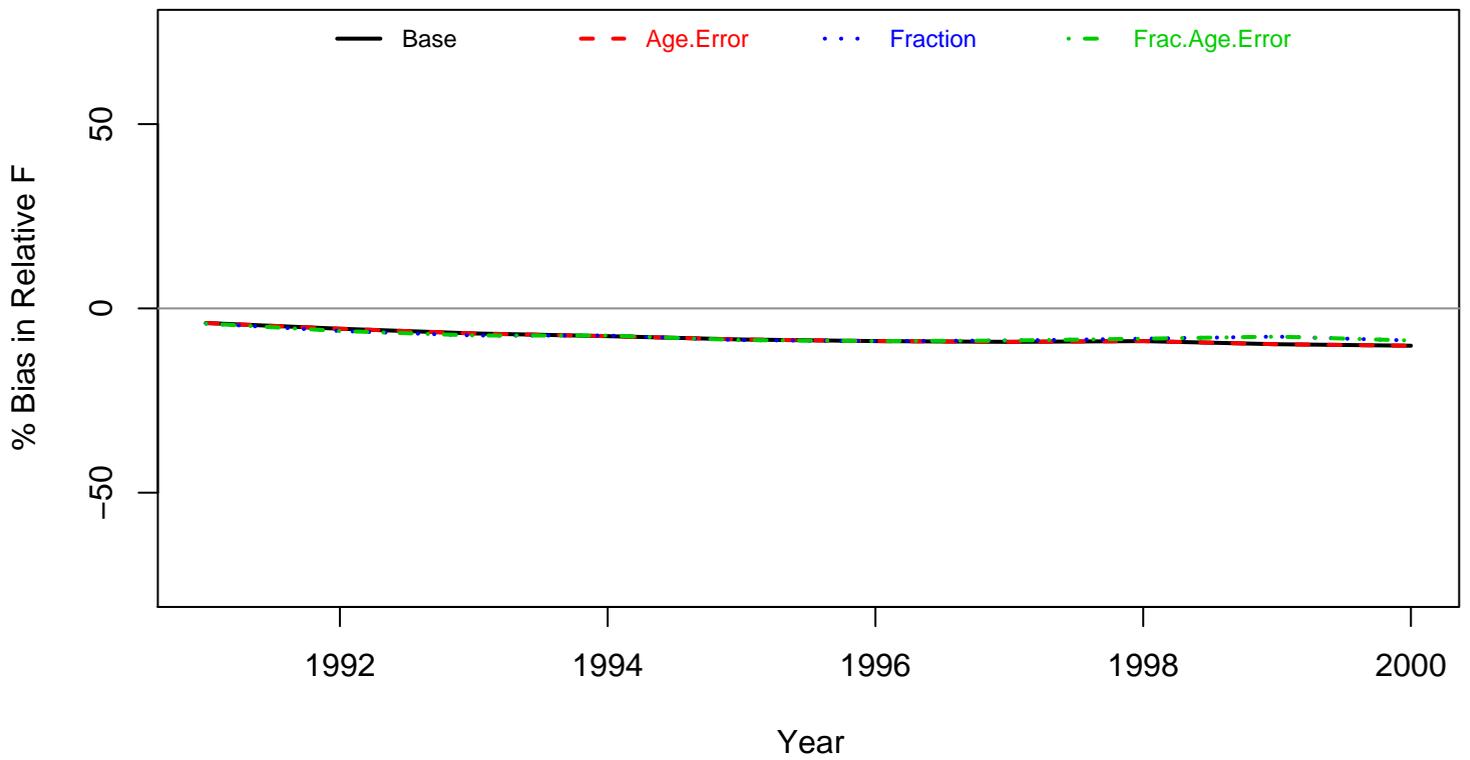
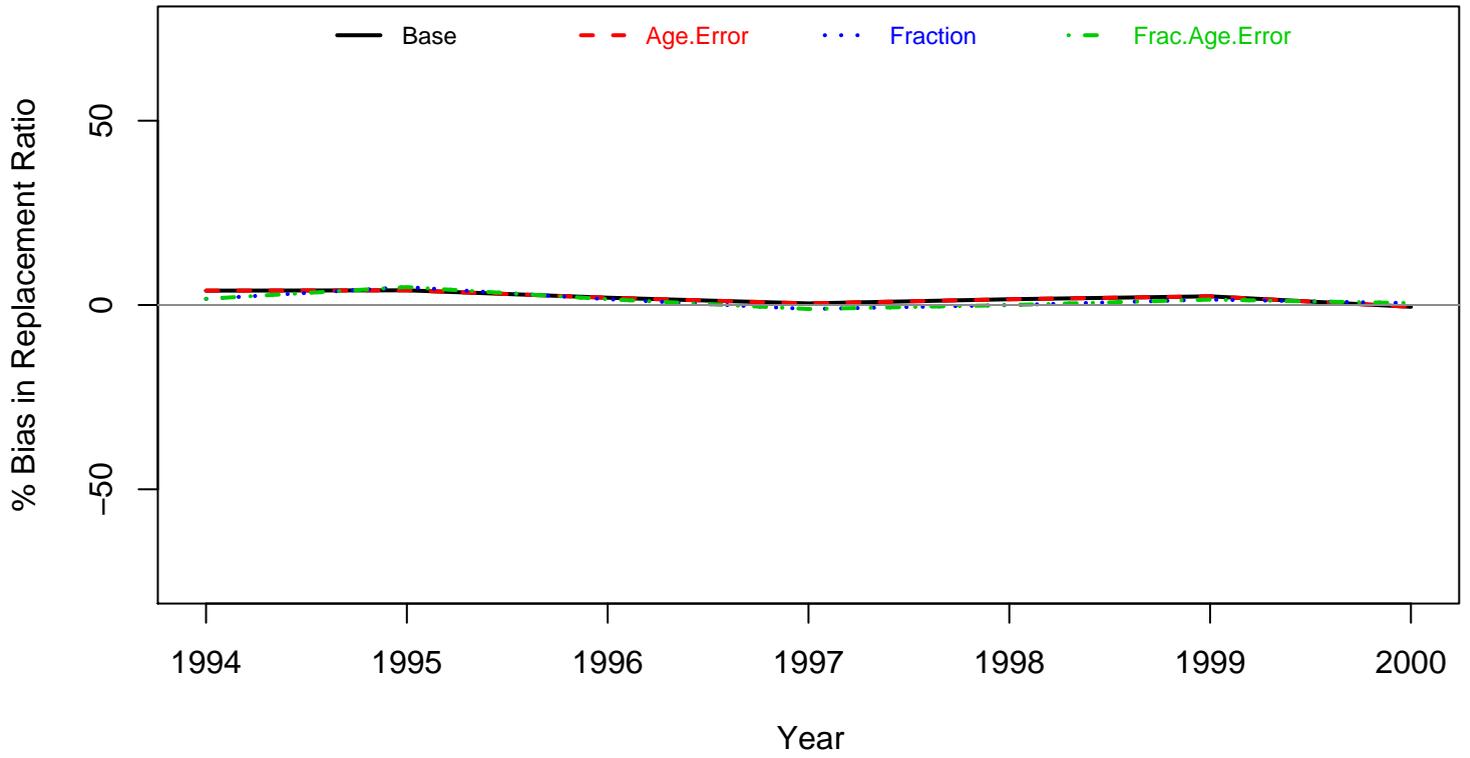


Figure WHAKE.1 Bias in replacement ratio and relative F for all simulated data sets using AIM.⁶⁹

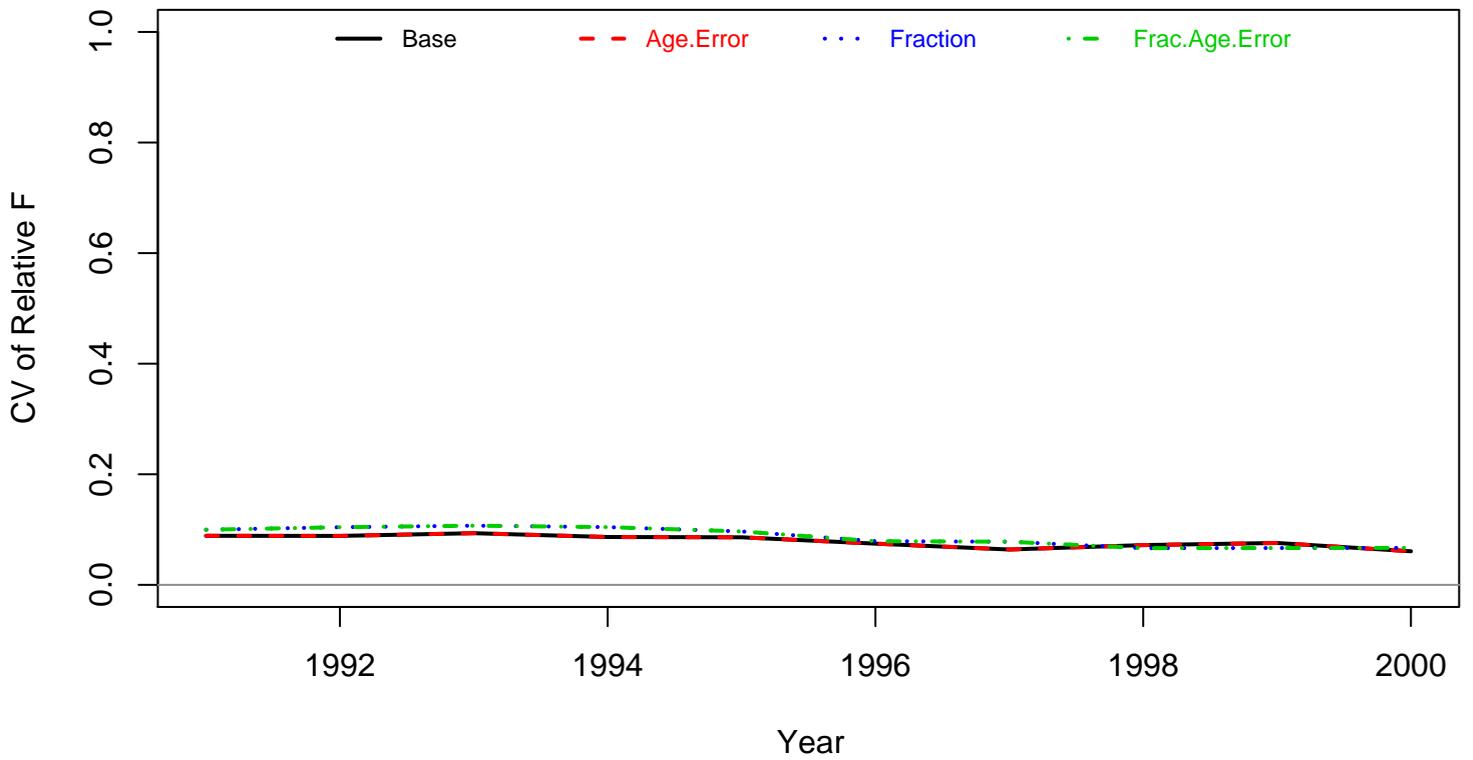
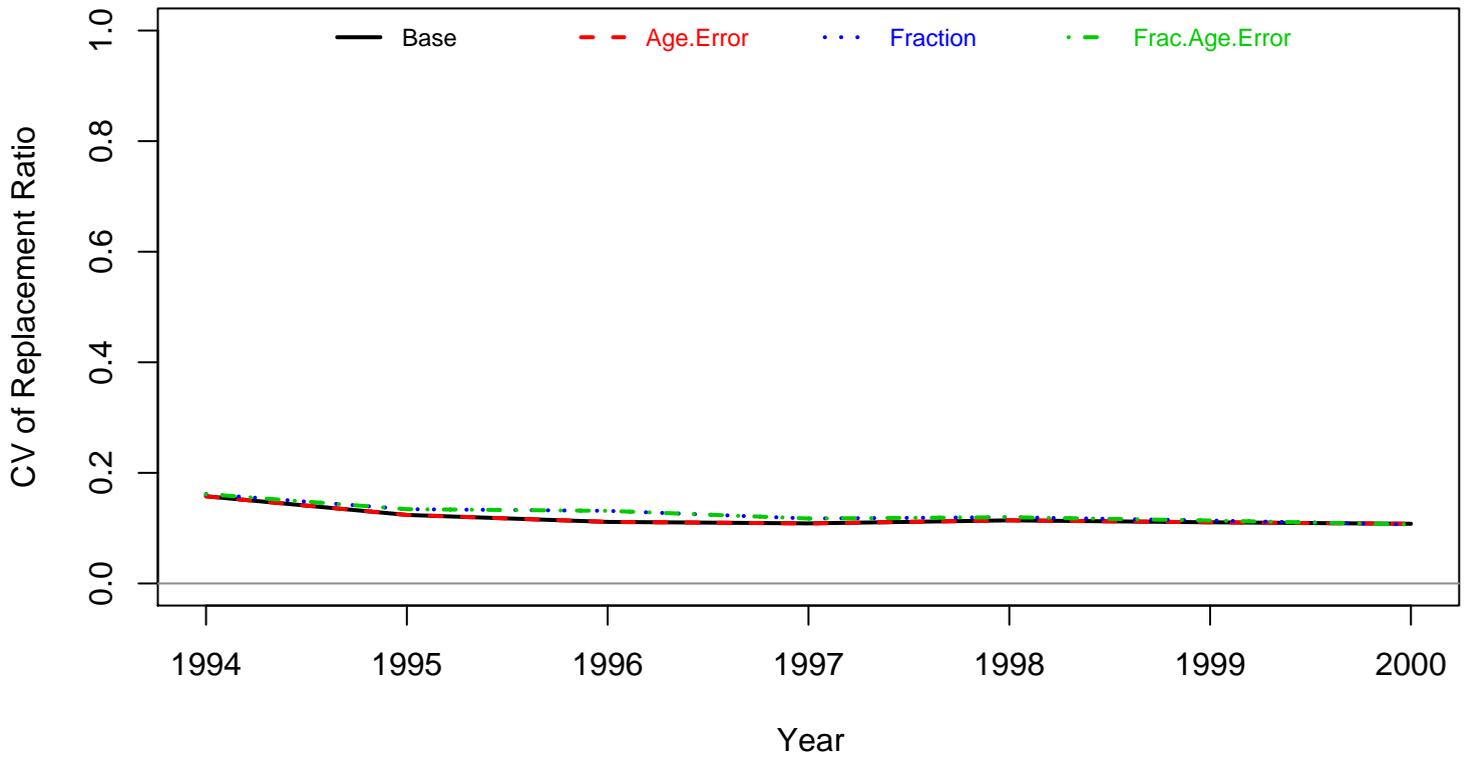


Figure WHAKE.2 CV in replacement ratio and relative F for all simulated data sets using AIM. ⁷⁰

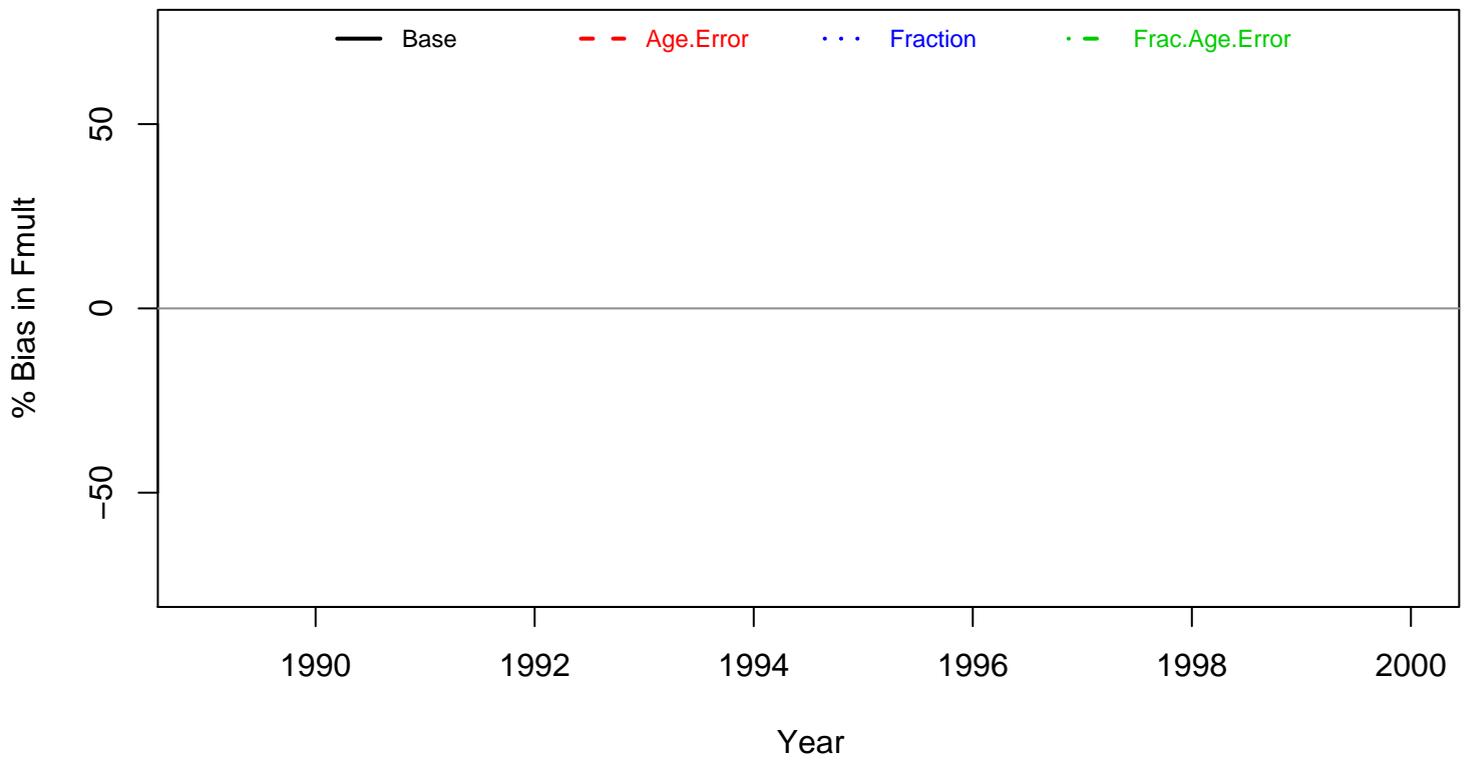


Figure WHAKE.3 Bias in total biomass and F for all simulated data sets using ASPIC. 71

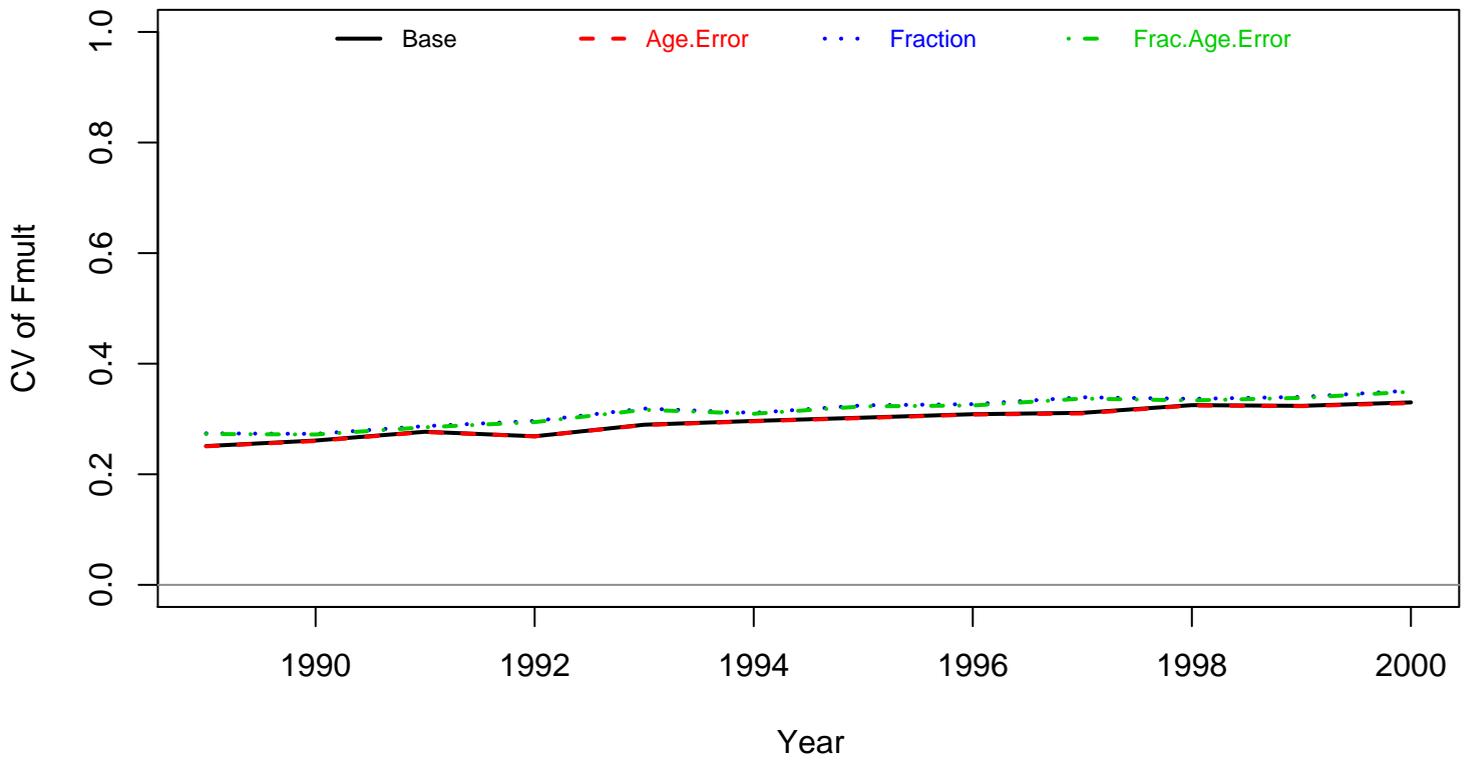
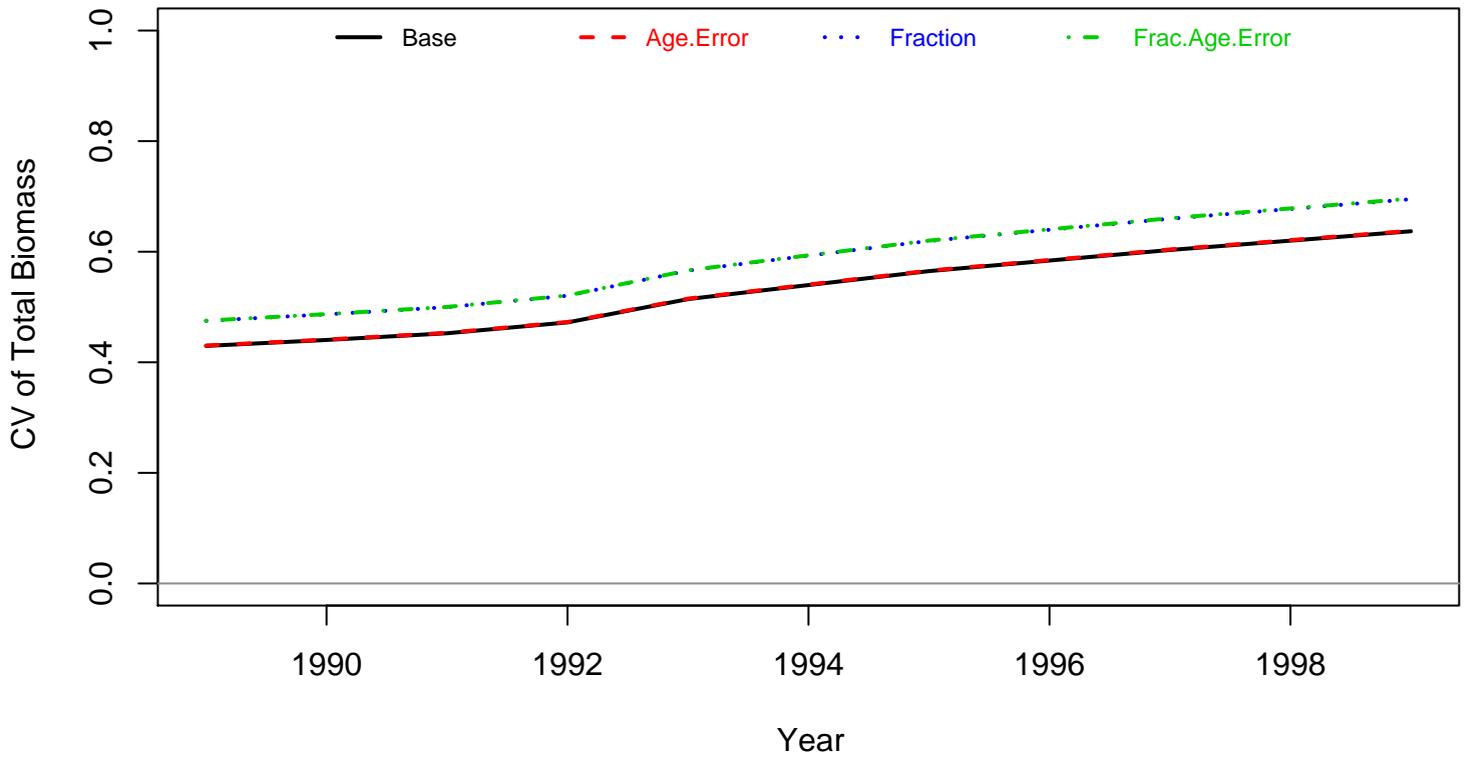


Figure WHAKE.4 CV in total biomass and F for all simulated data sets using ASPIC.

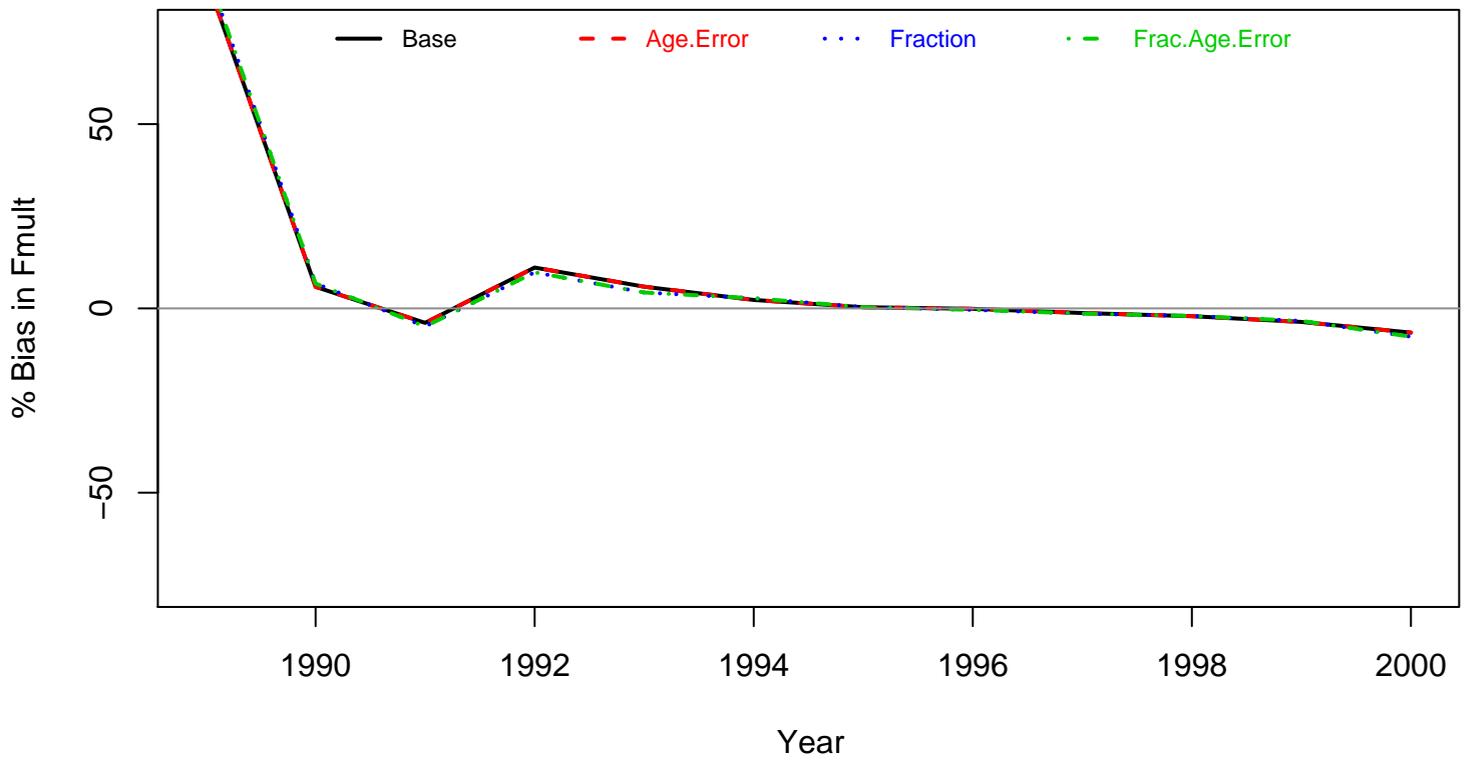
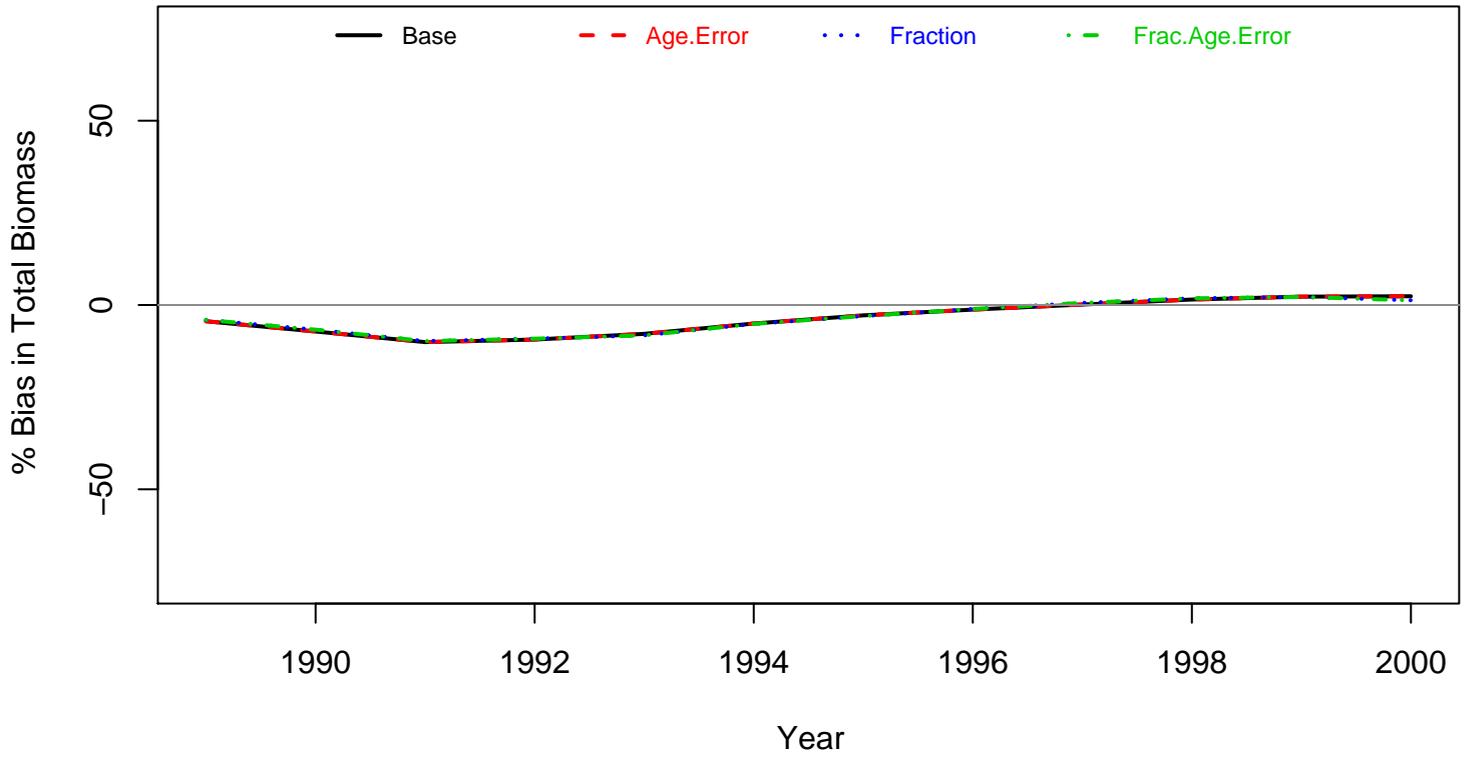


Figure WHAKE.5 Bias in total biomass and F multiplier for all simulated data sets using Scale. ⁷³

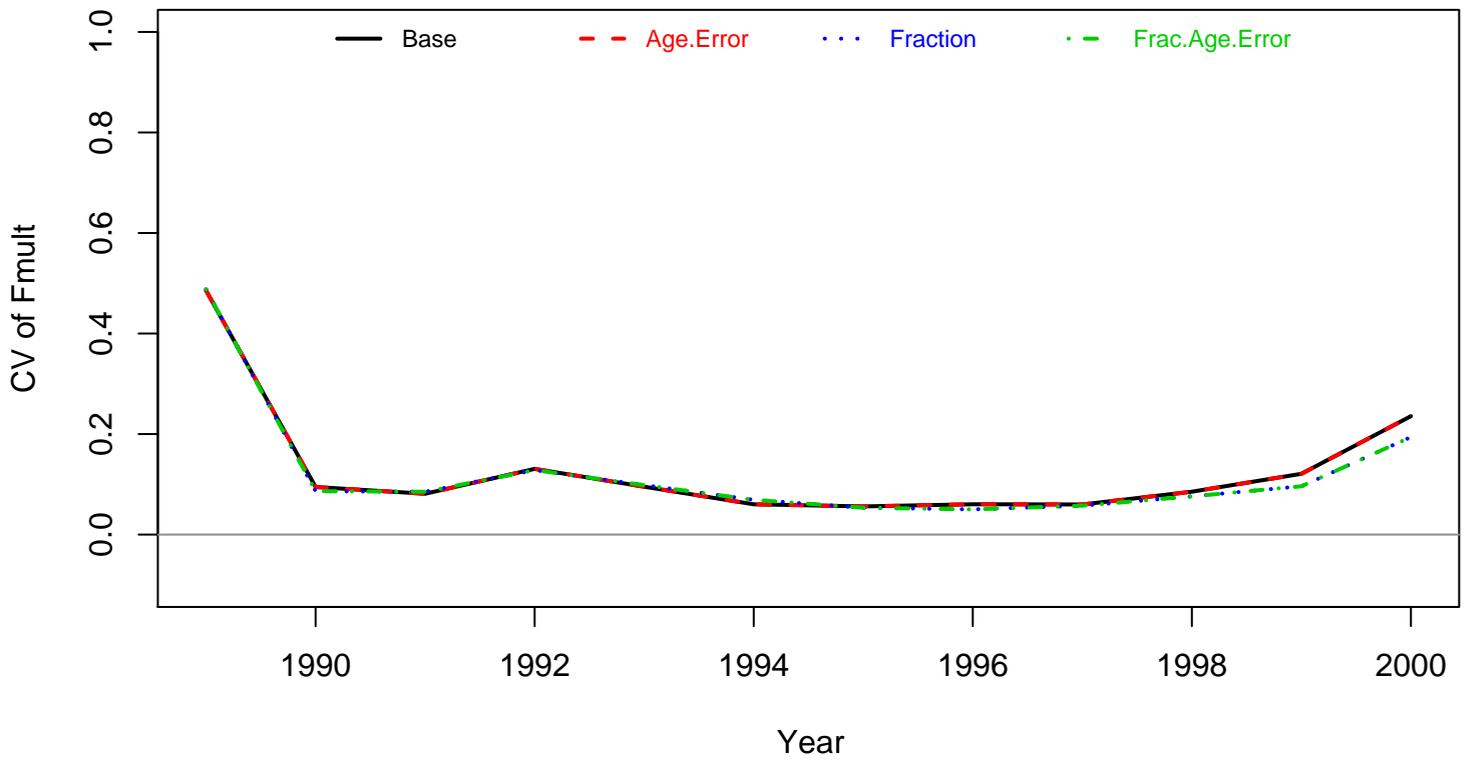
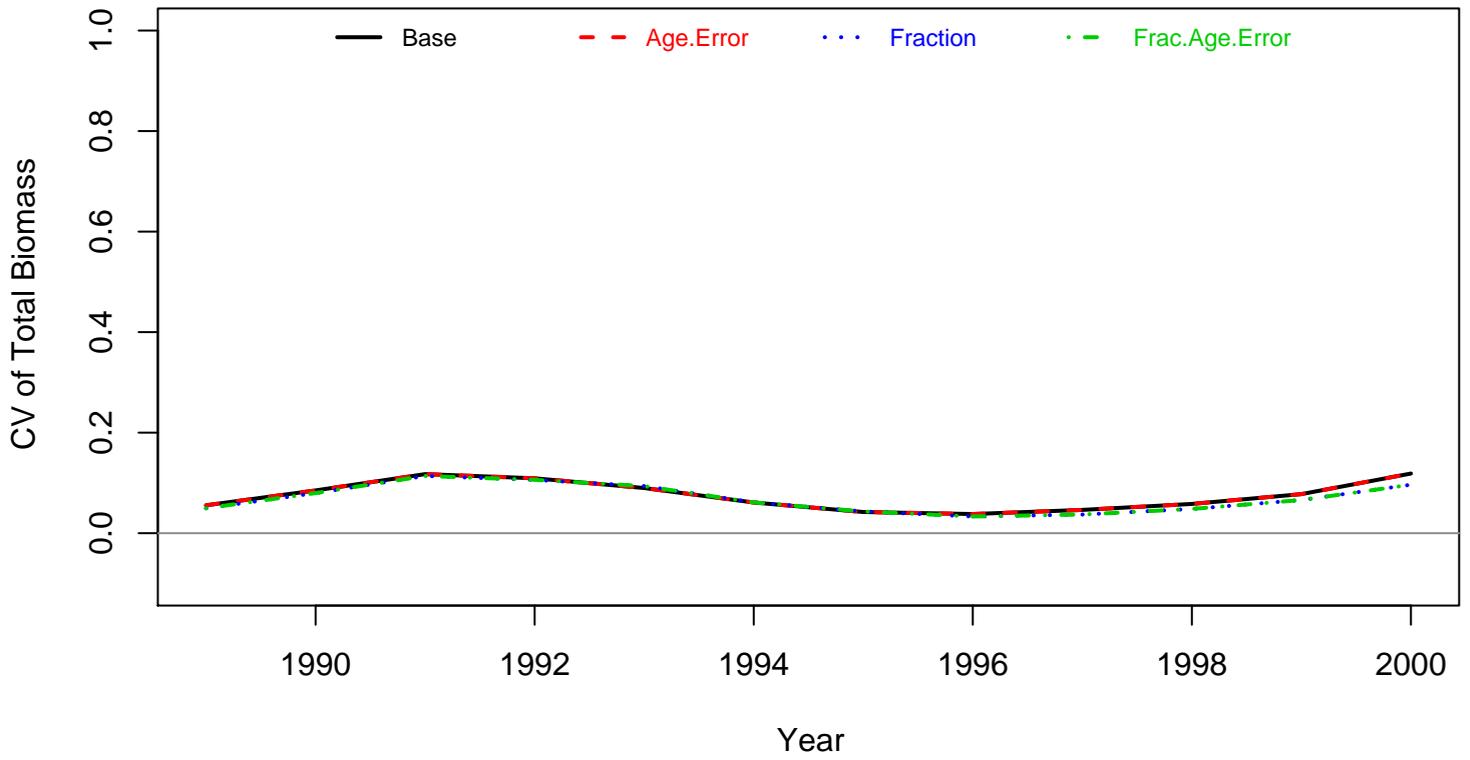


Figure WHAKE.6 CV in total biomass and F multiplier for all simulated data sets using Scale. ⁷⁴

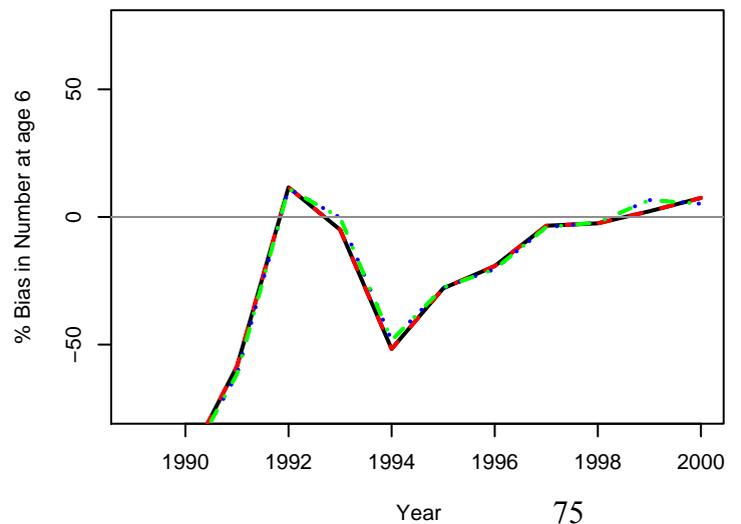
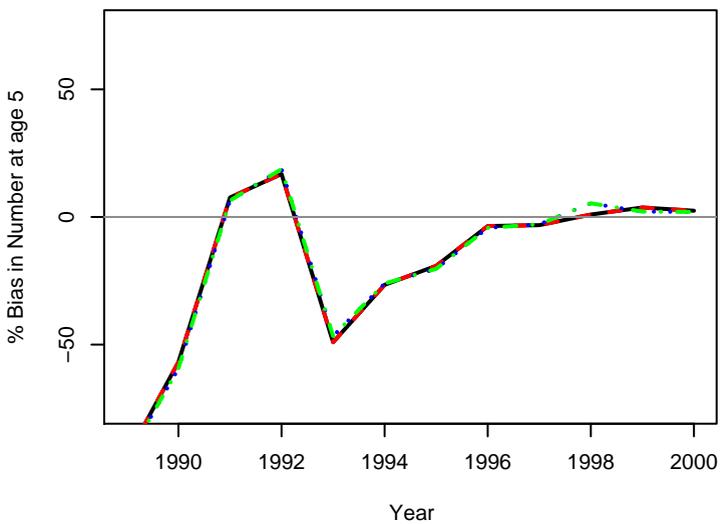
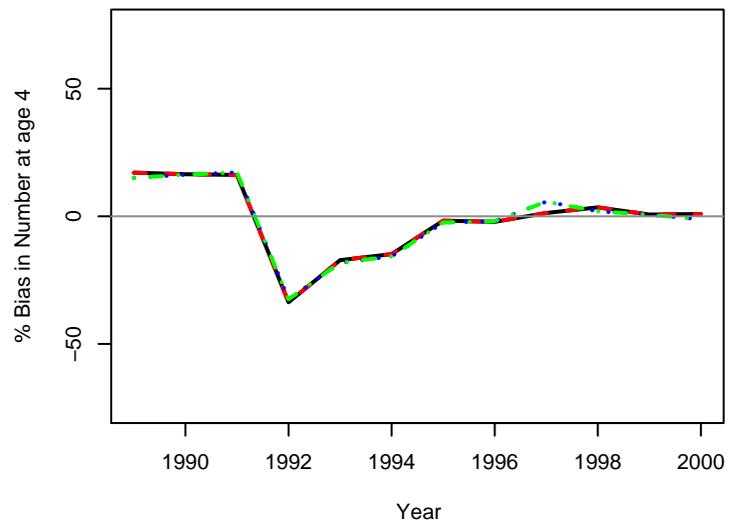
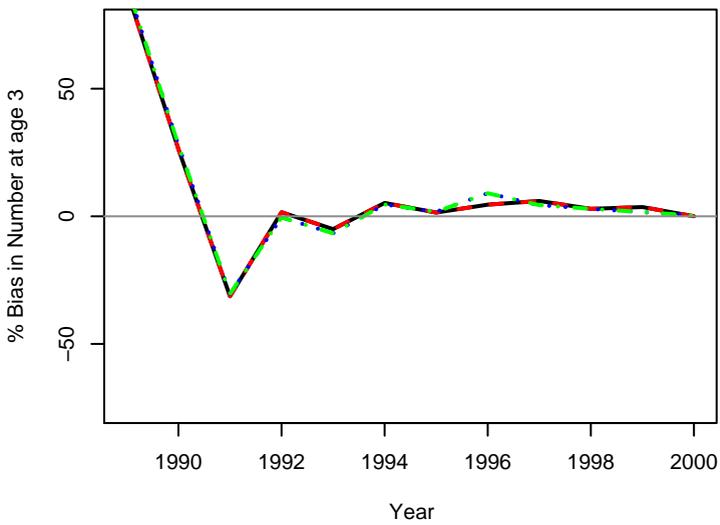
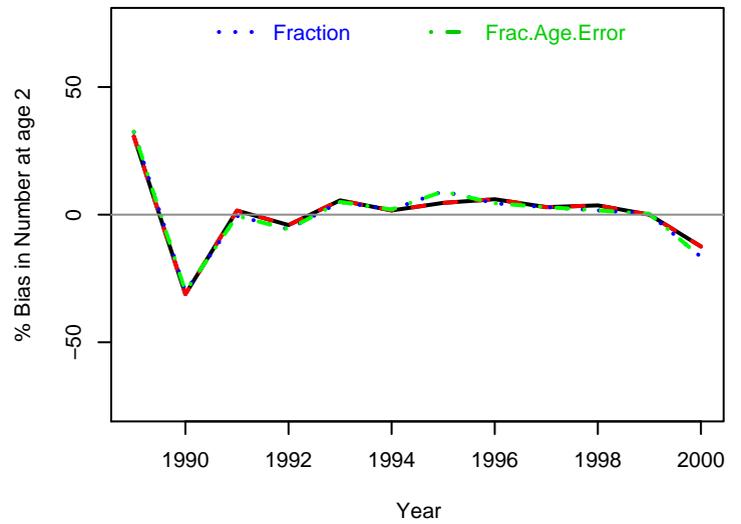
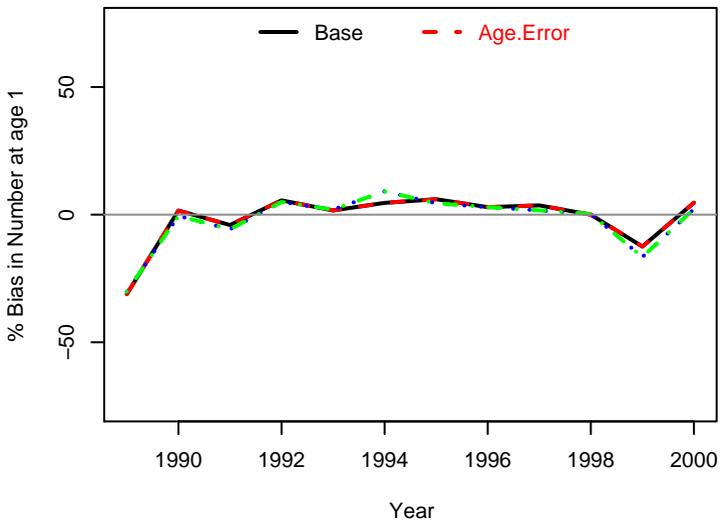


Figure WHAKE.7 Bias in number at age for all simulated data sets using Scale.

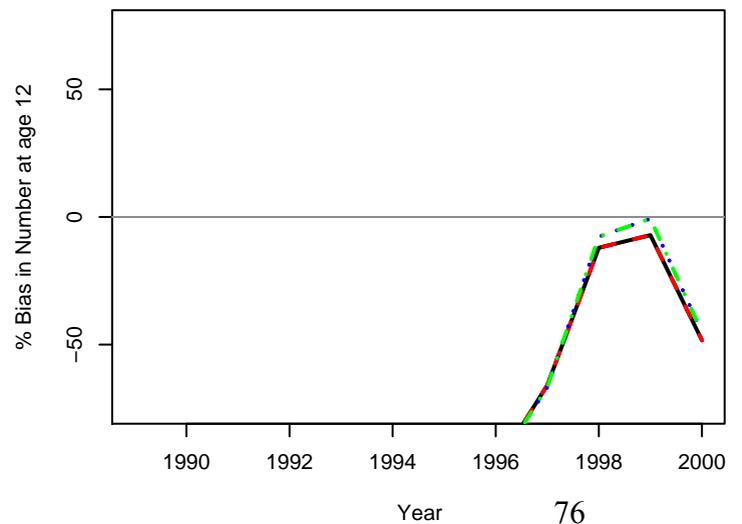
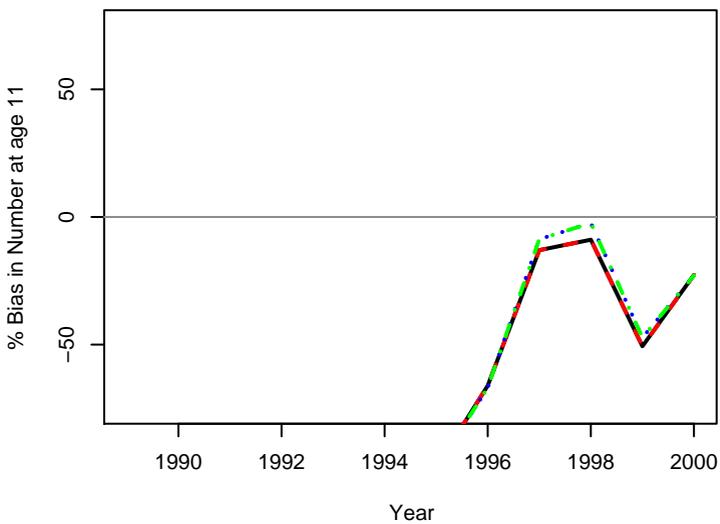
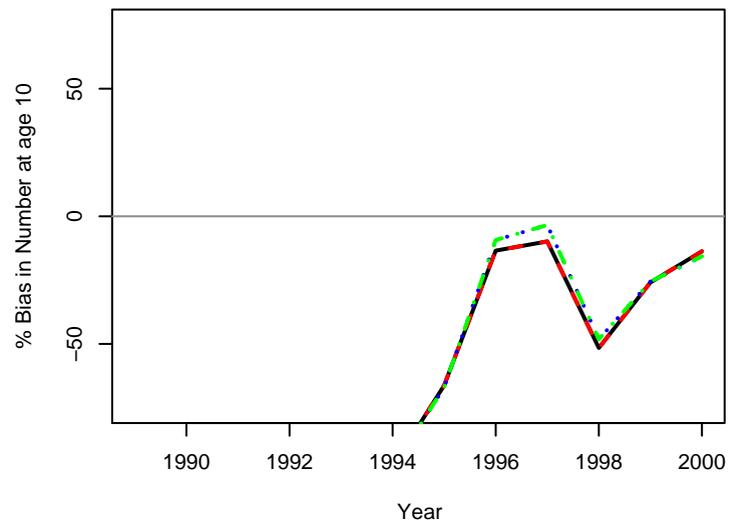
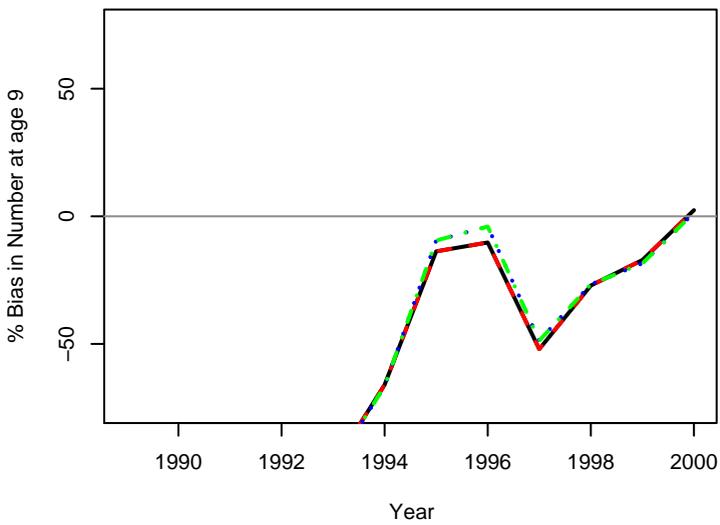
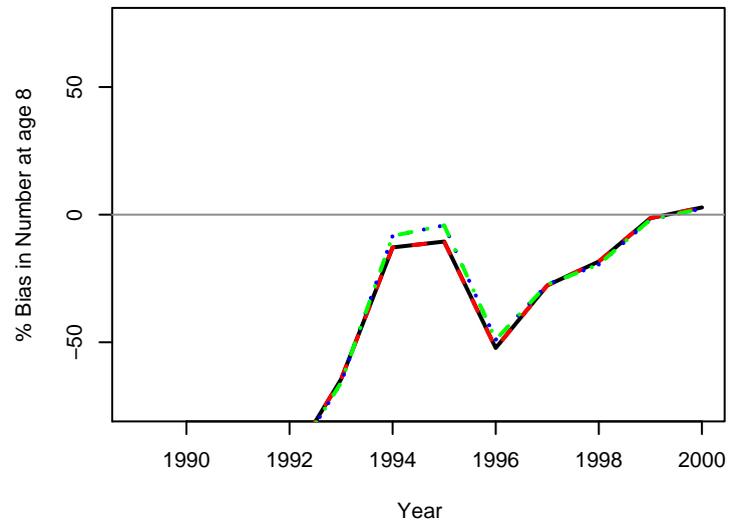
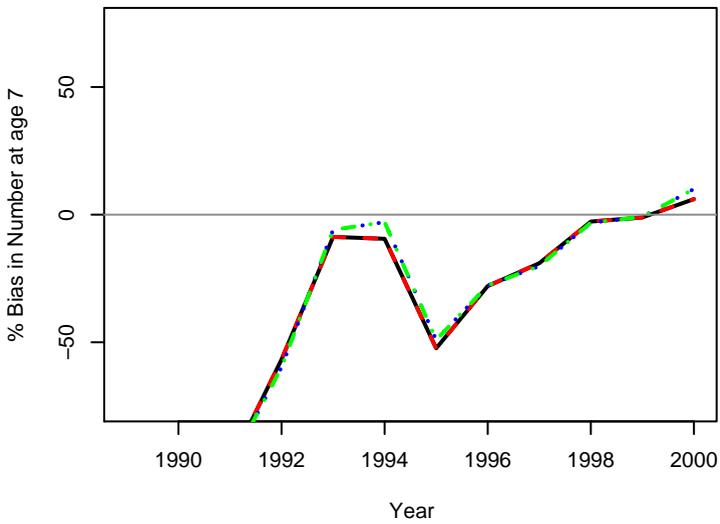


Figure WHAKE.7 (cont.)

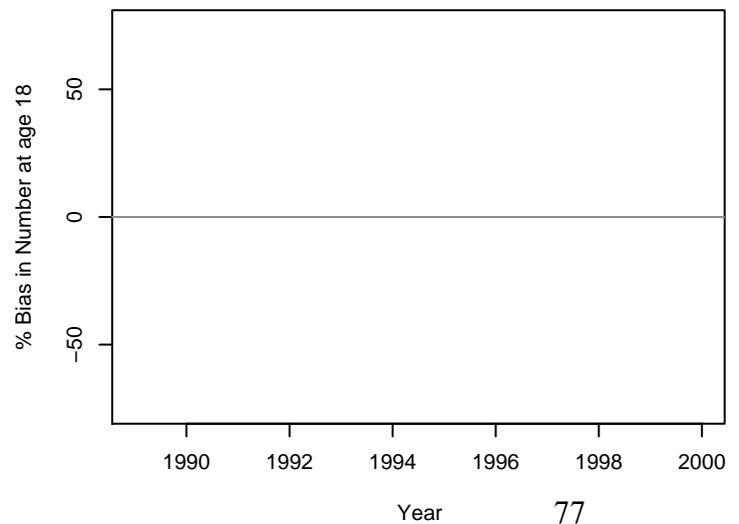
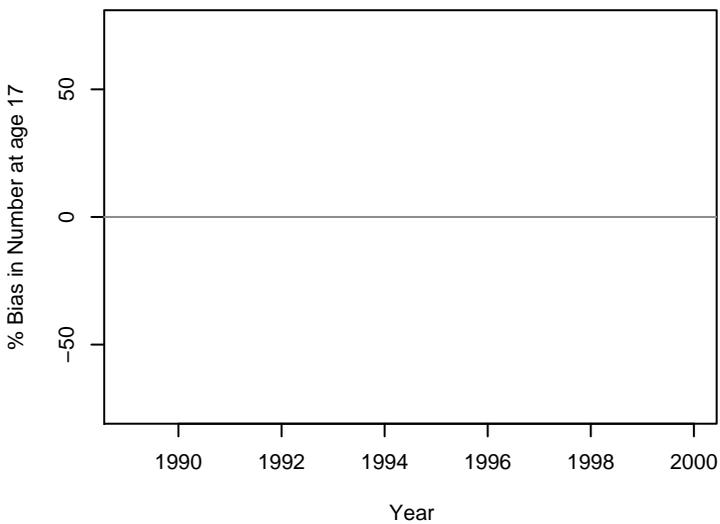
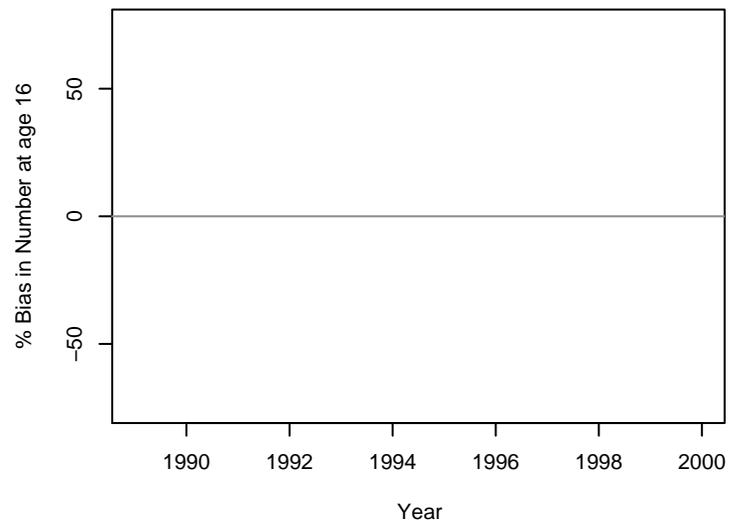
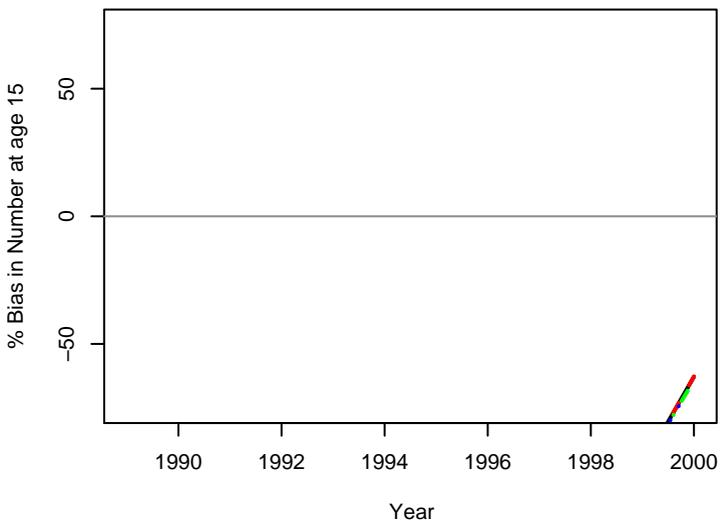
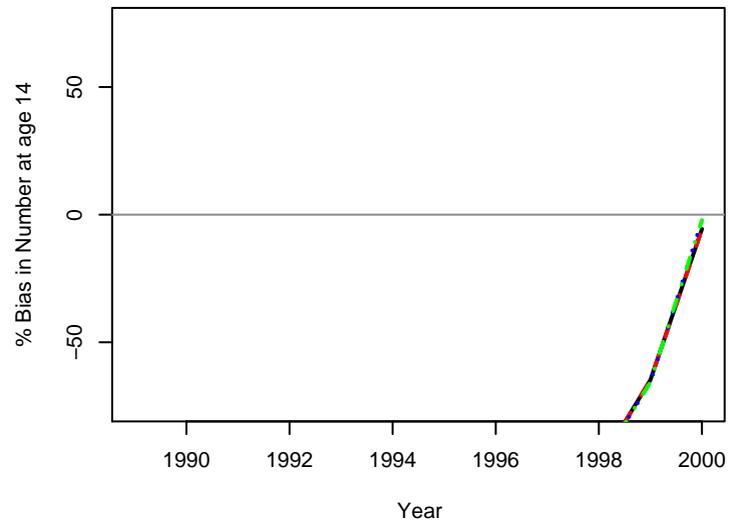
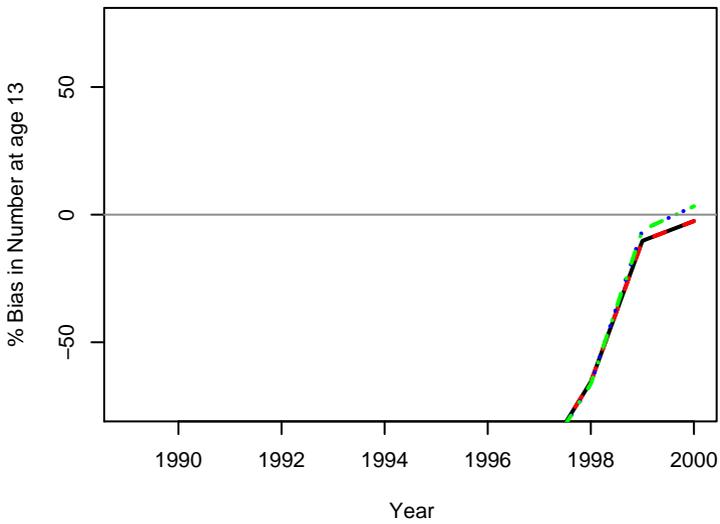


Figure WHAKE.7 (cont.)

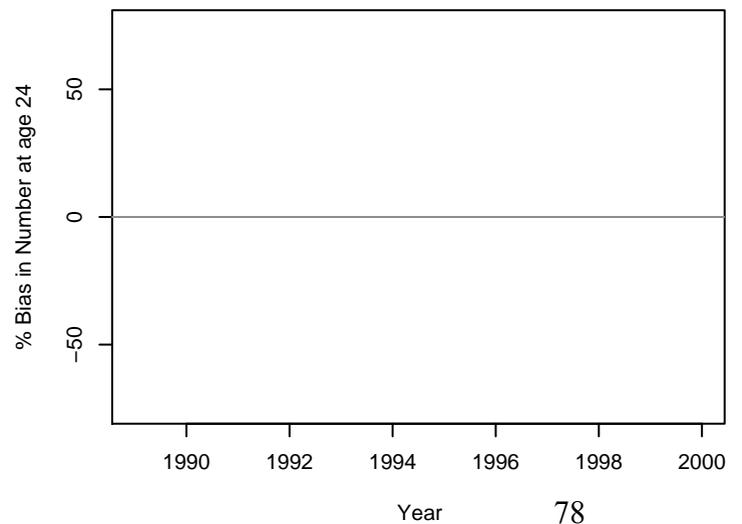
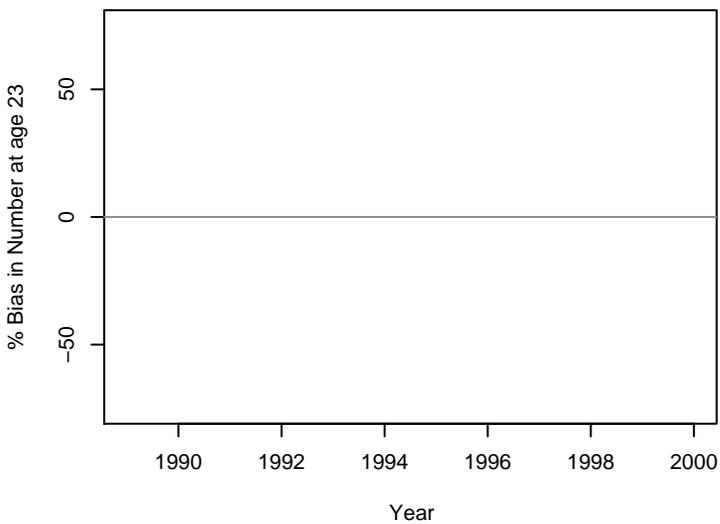
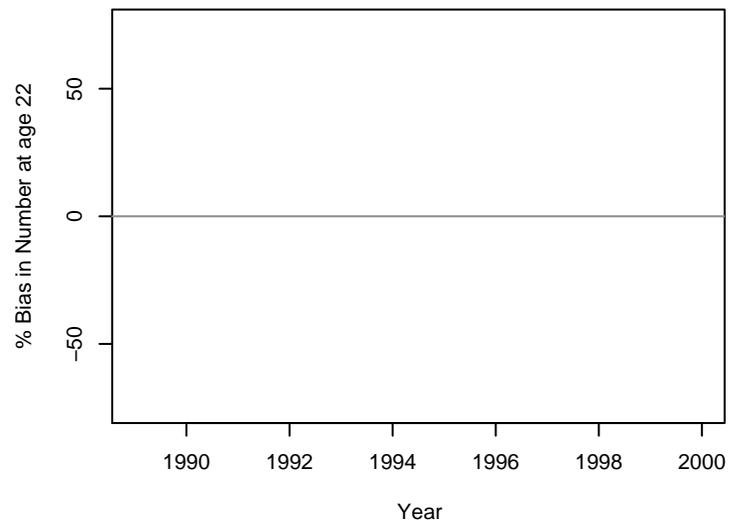
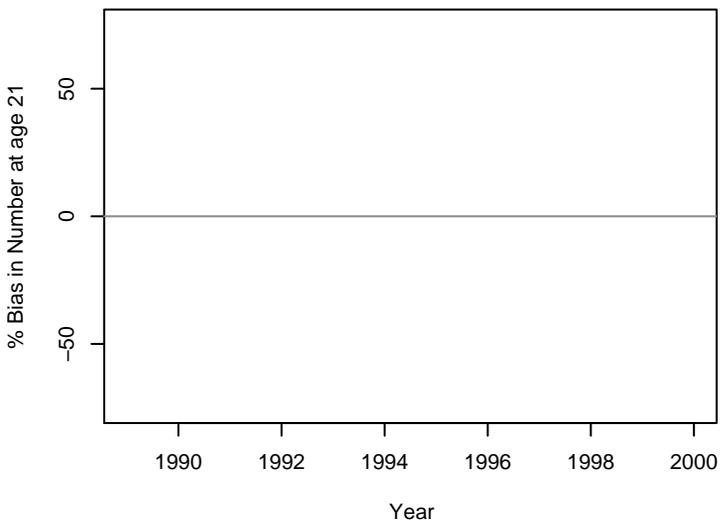
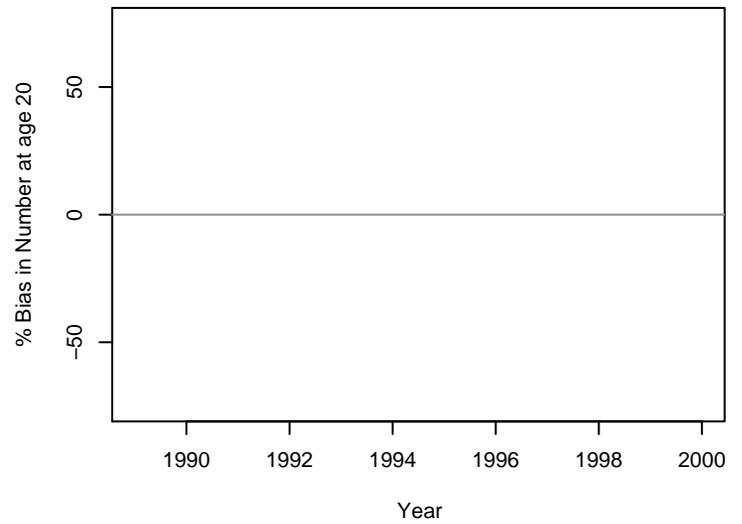
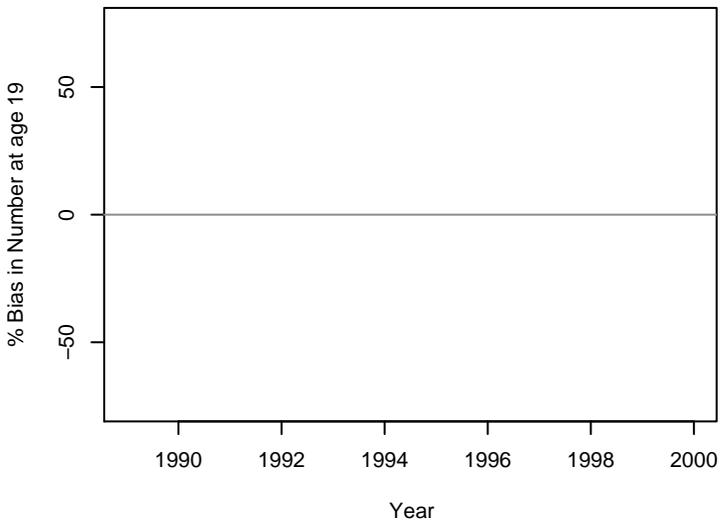


Figure WHAKE.7 (cont.)

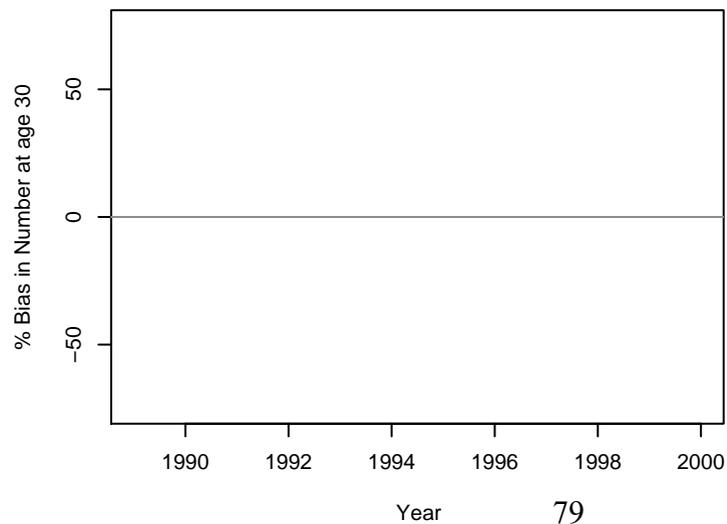
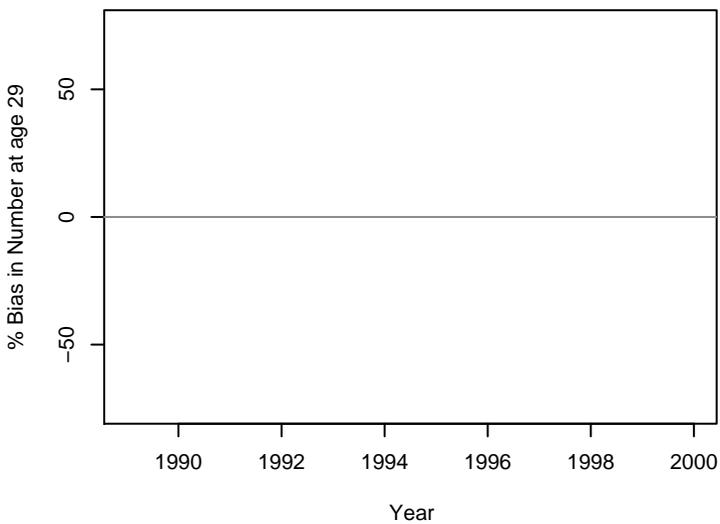
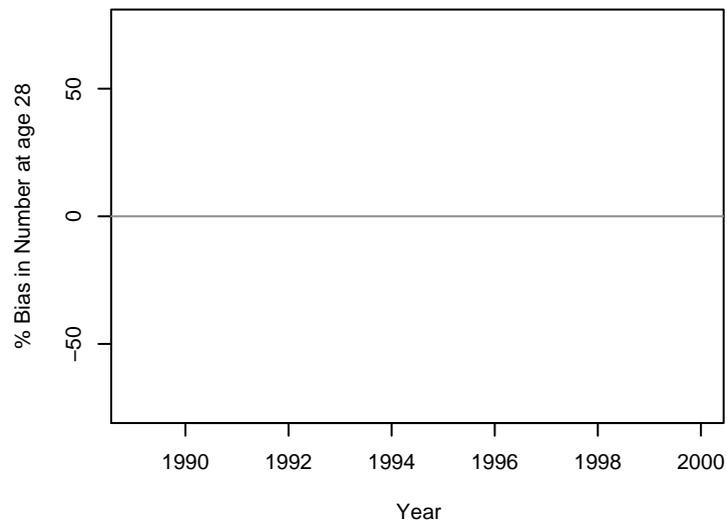
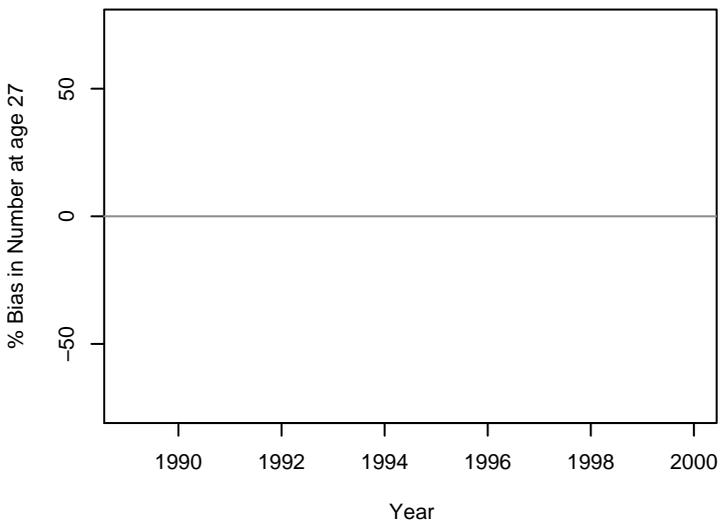
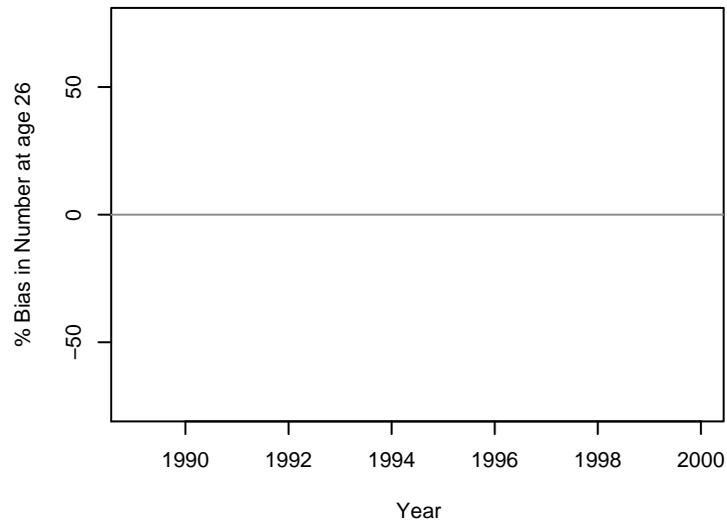
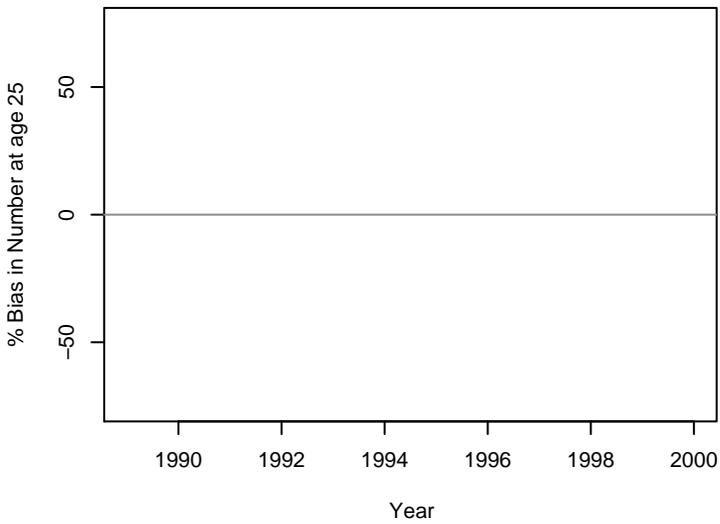


Figure WHAKE.7 (cont.)

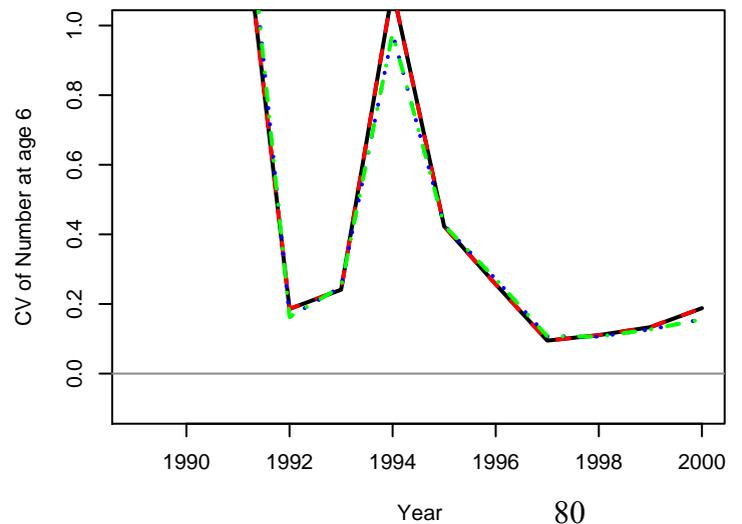
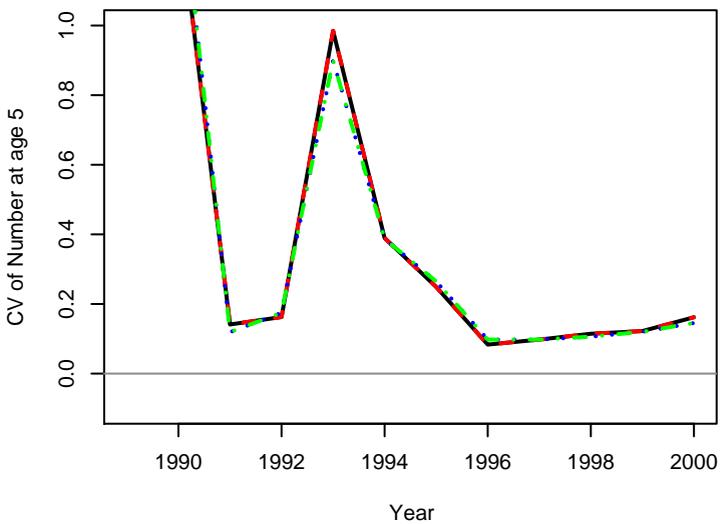
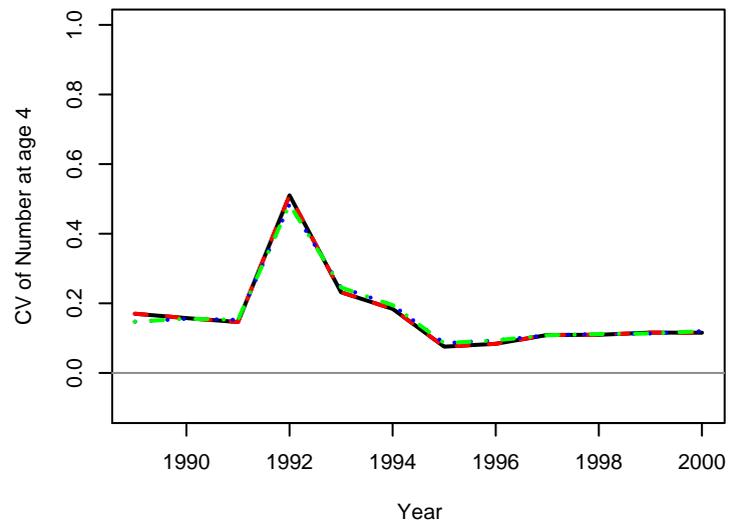
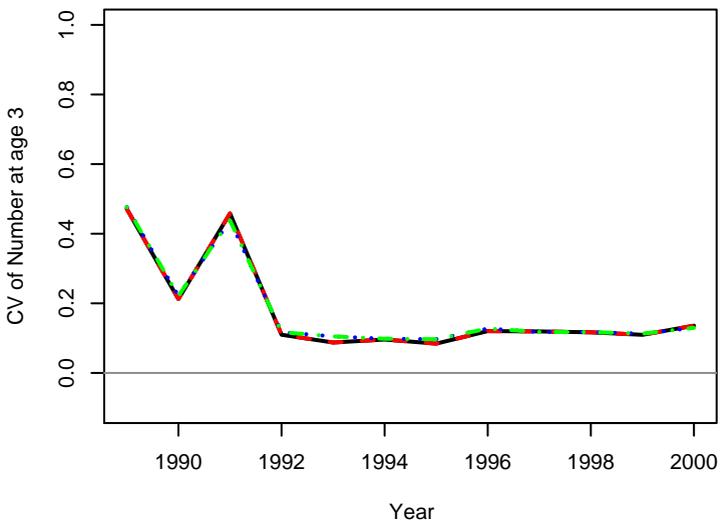
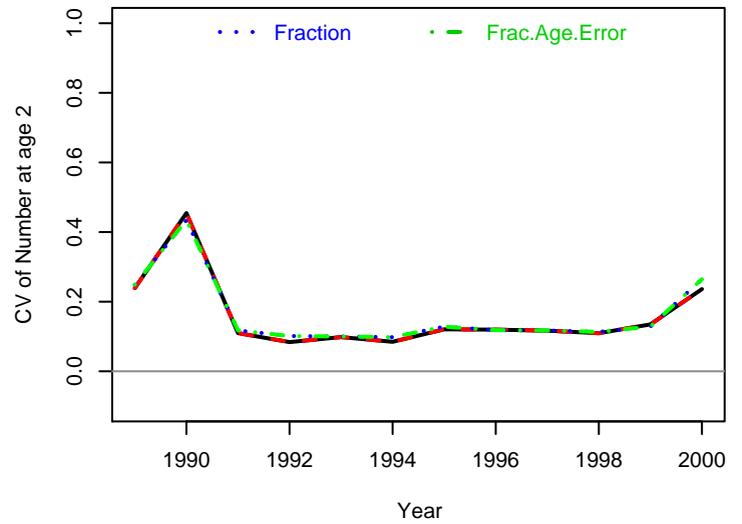
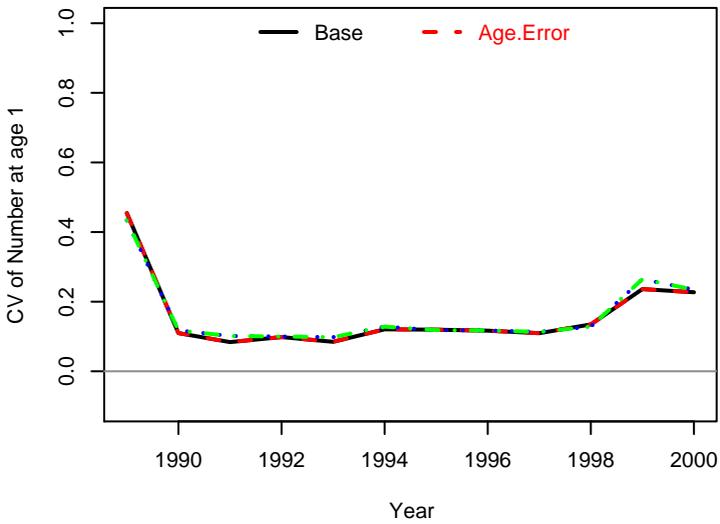


Figure WHAKE.8 CV in number at age for all simulated data sets using Scale.

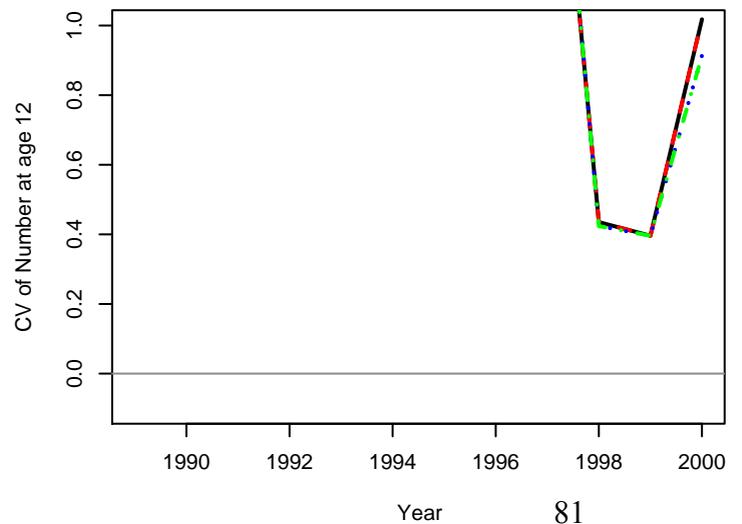
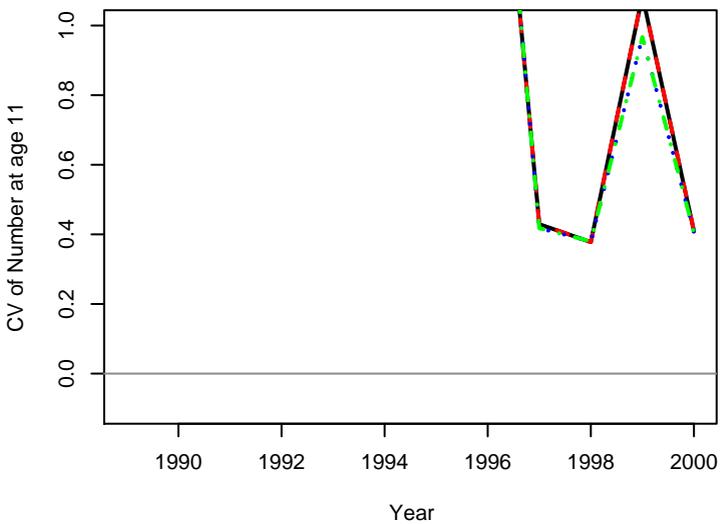
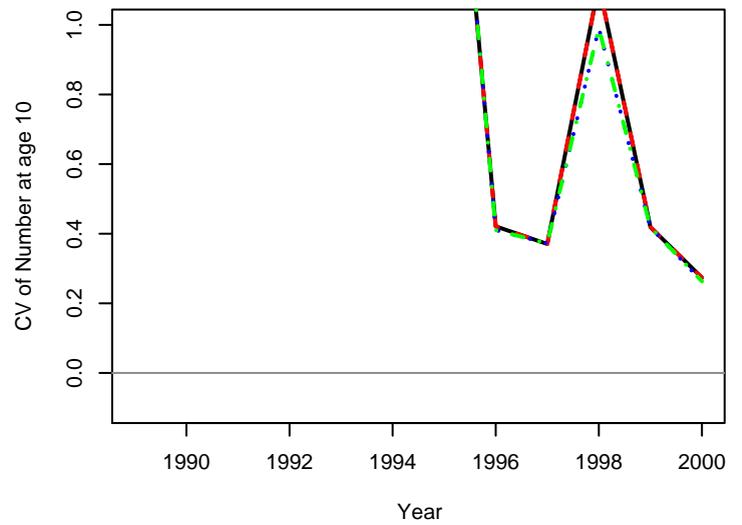
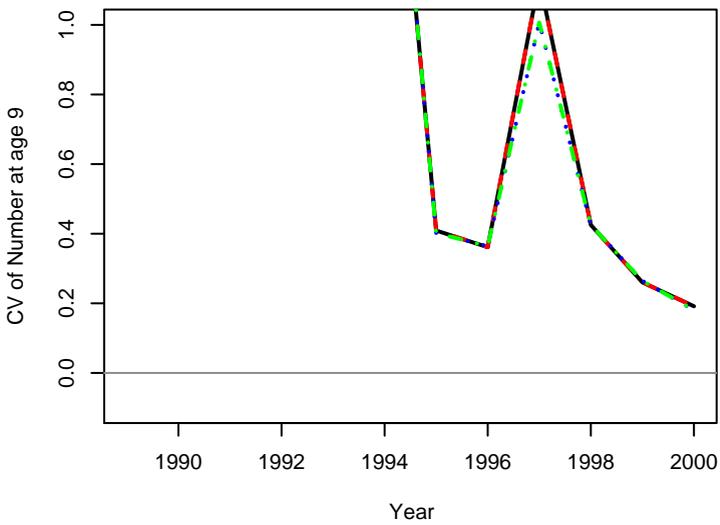
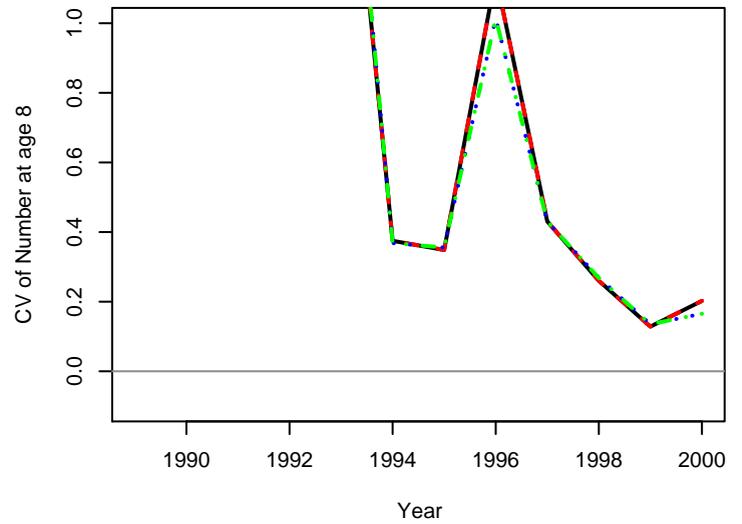
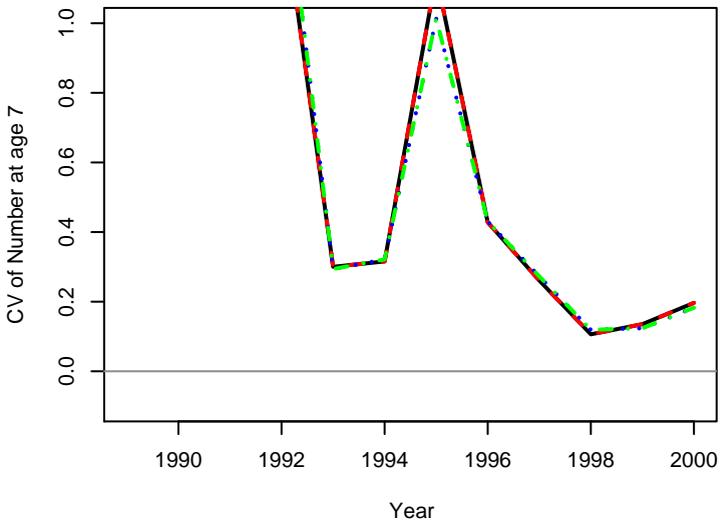


Figure WHAKE.8 (cont.)

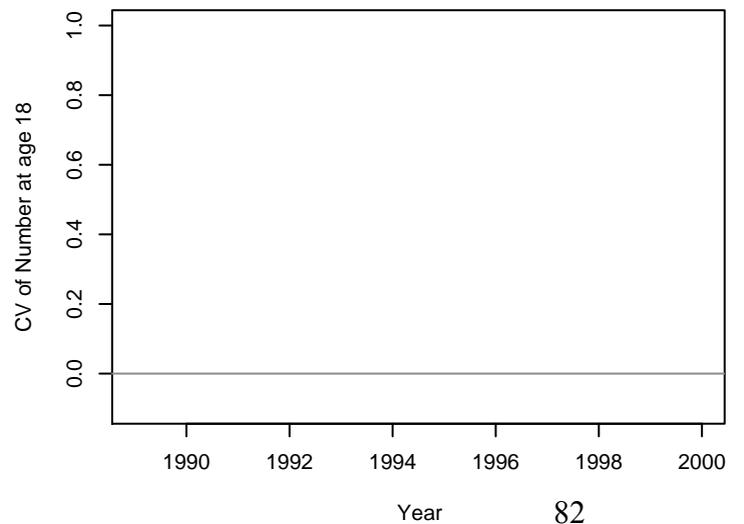
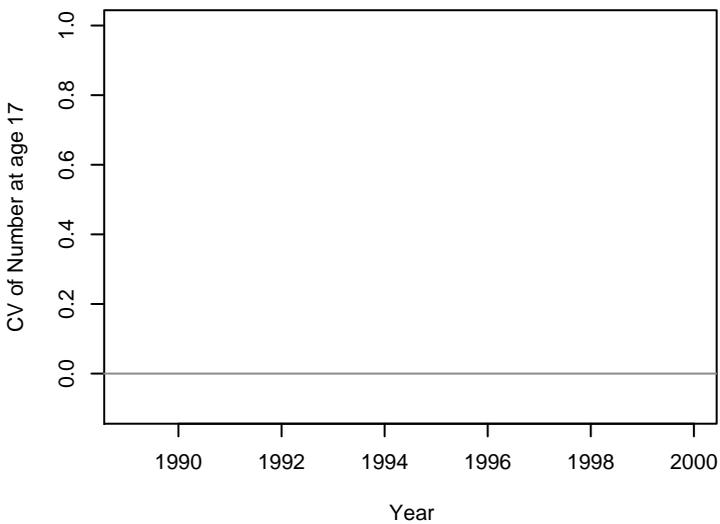
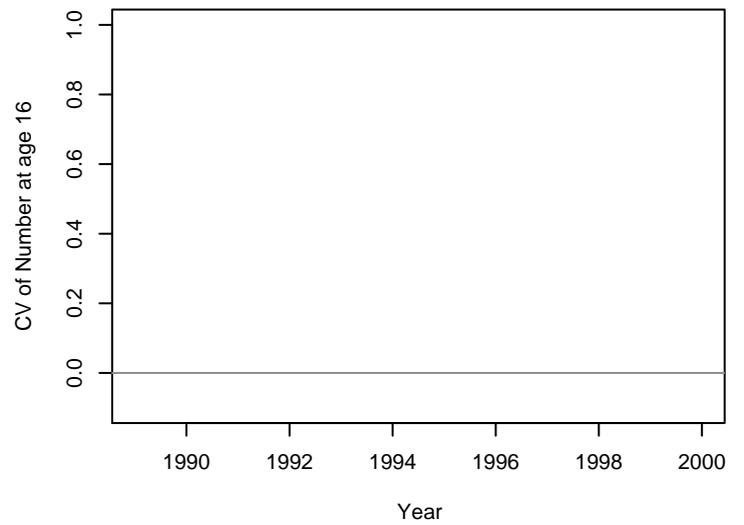
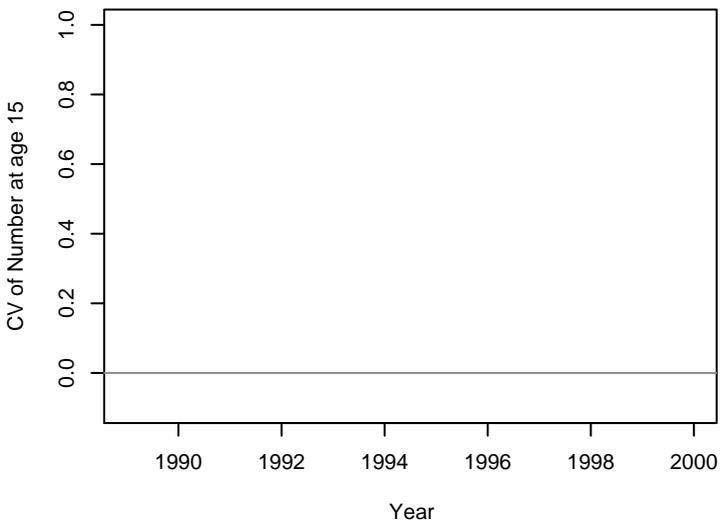
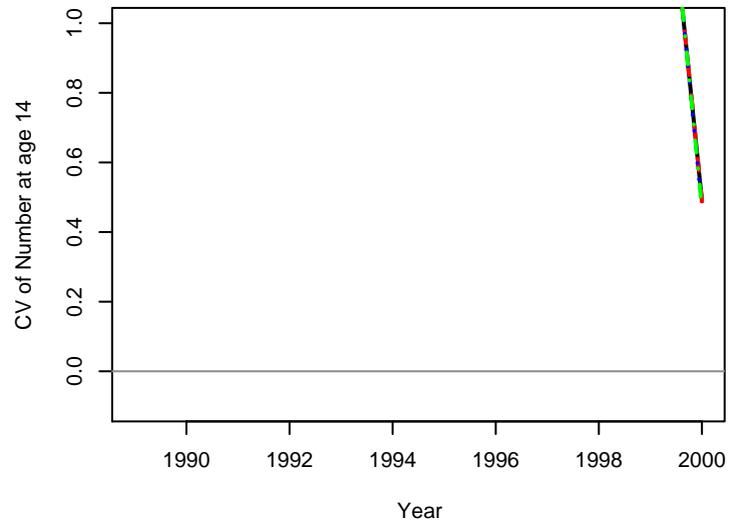
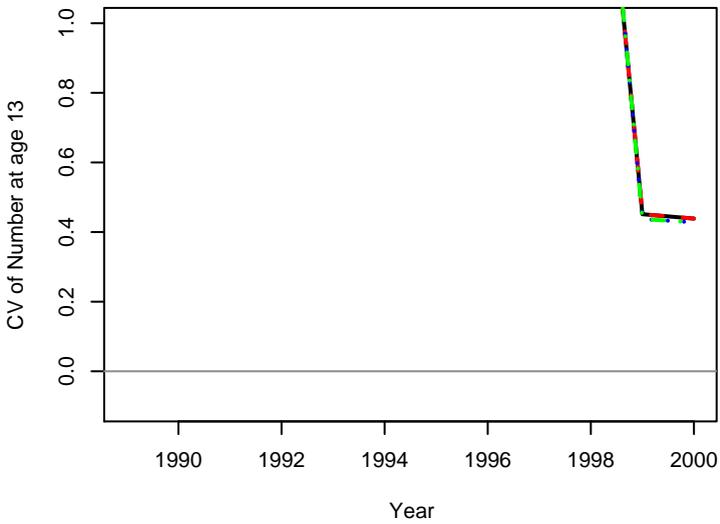


Figure WHAKE.8 (cont.)

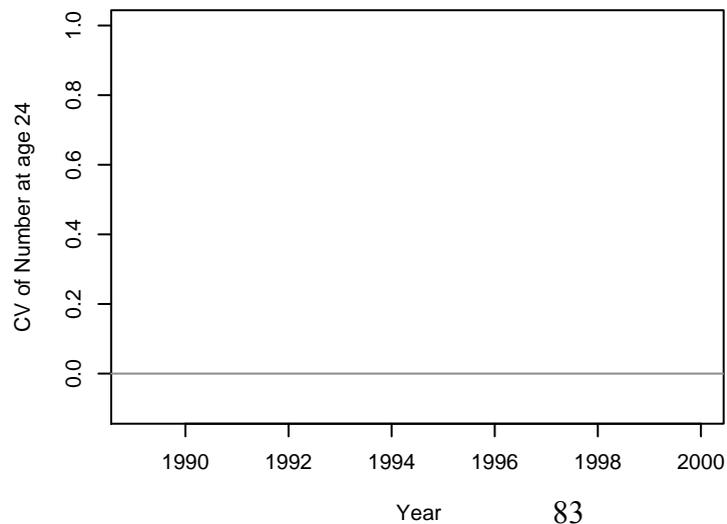
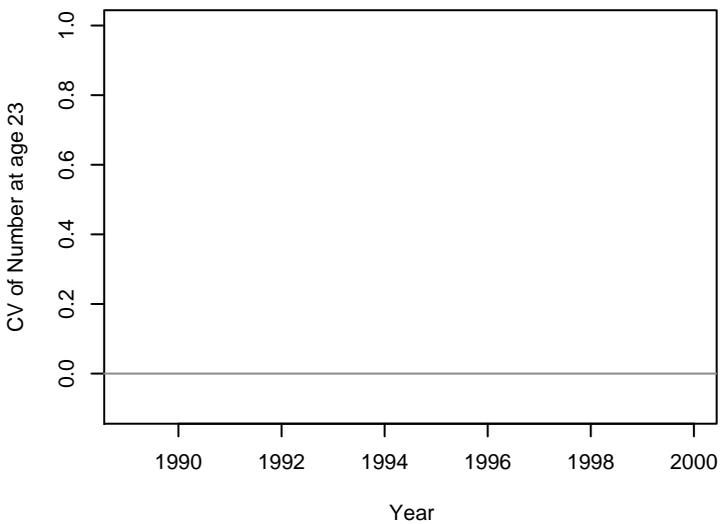
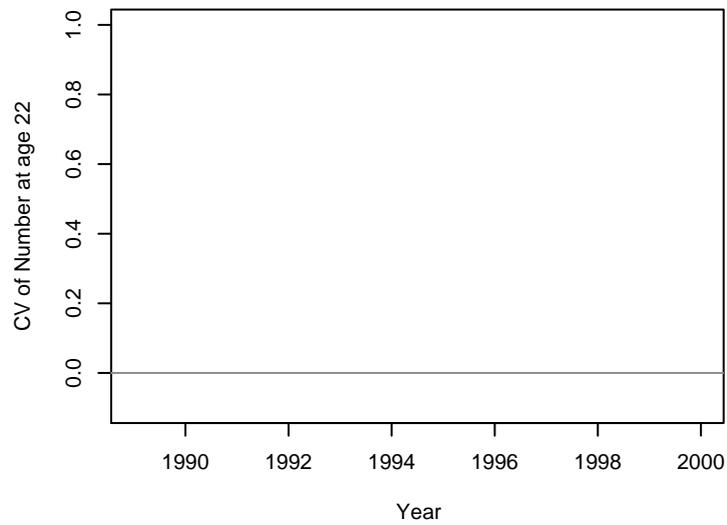
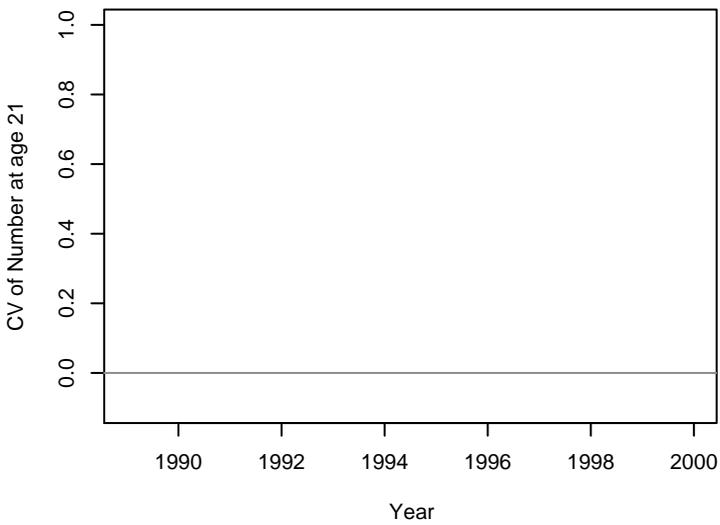
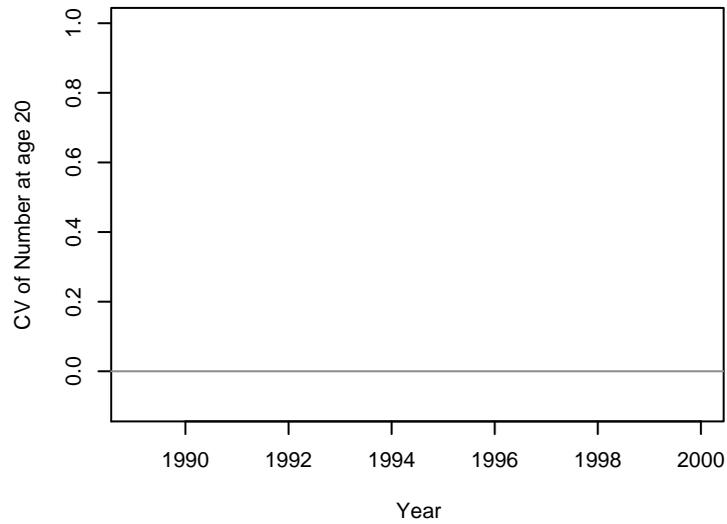
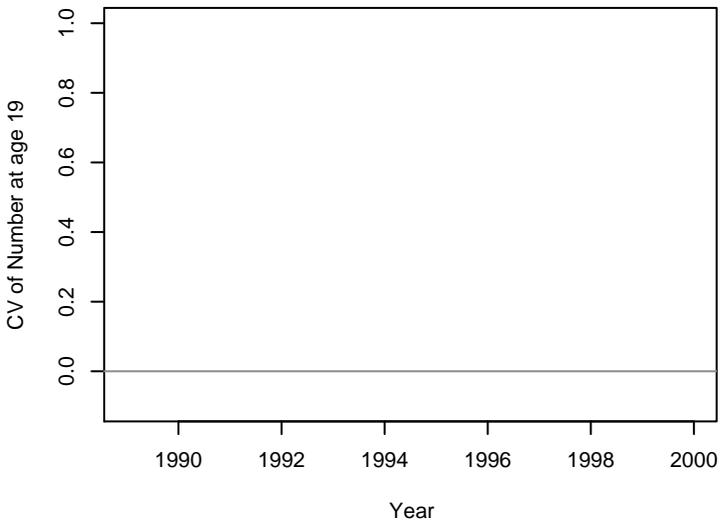


Figure WHAKE.8 (cont.)

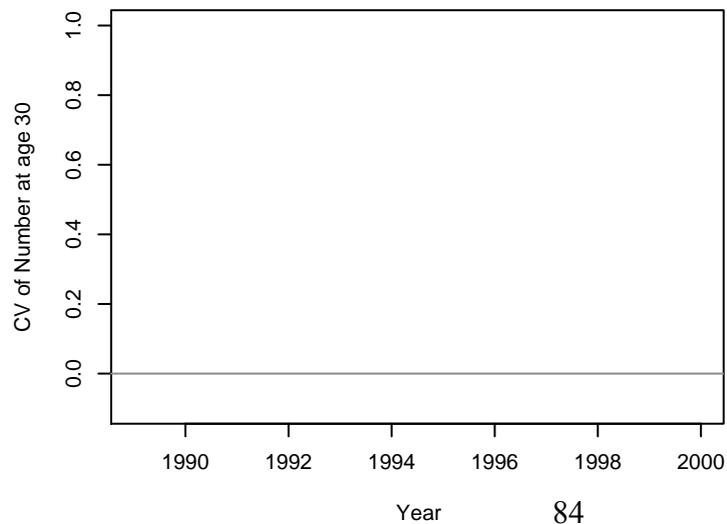
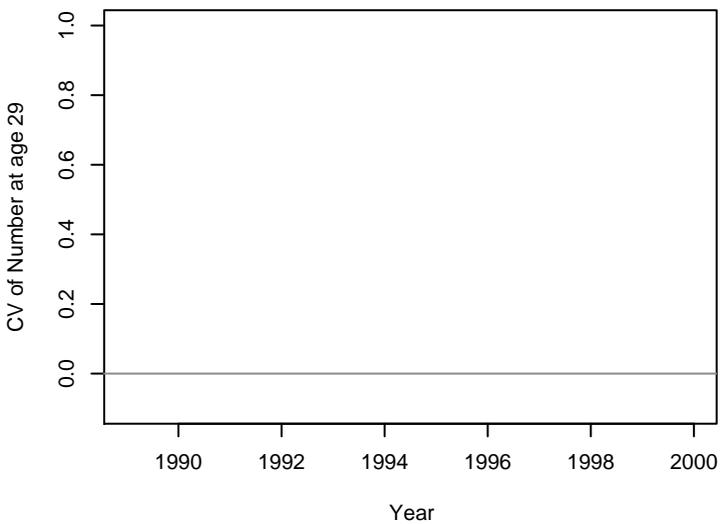
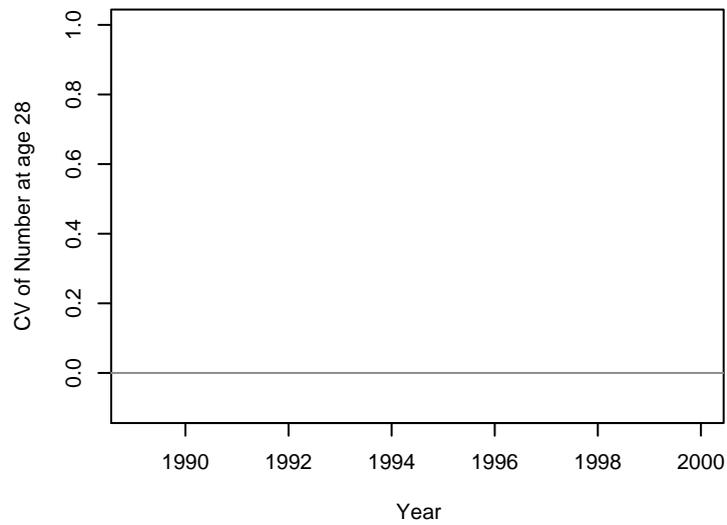
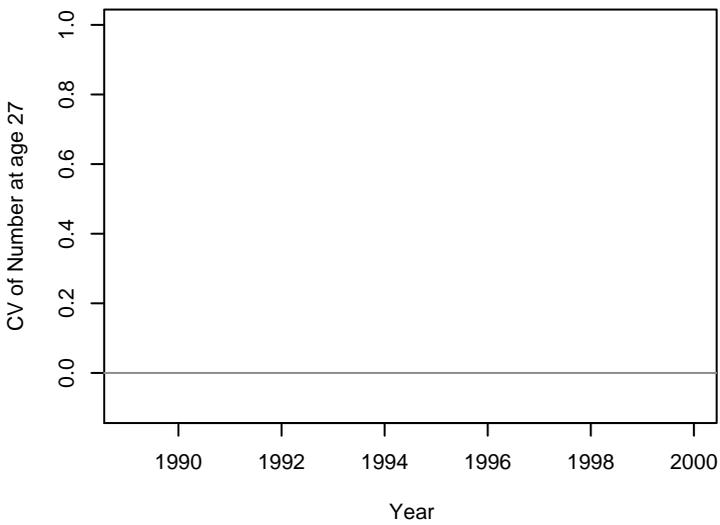
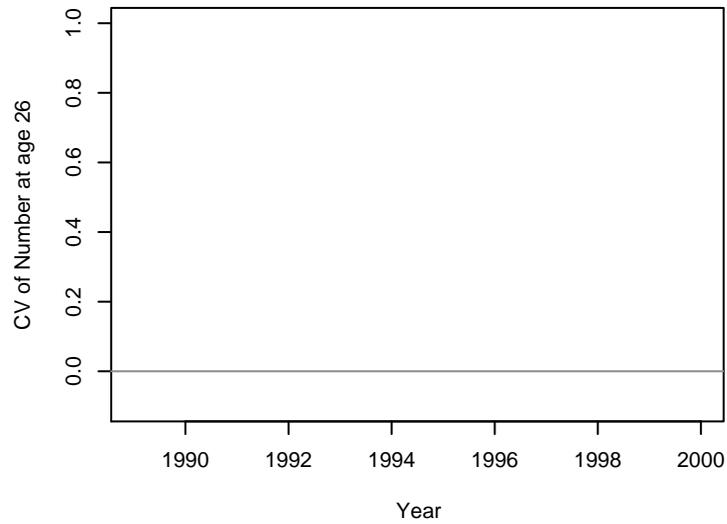
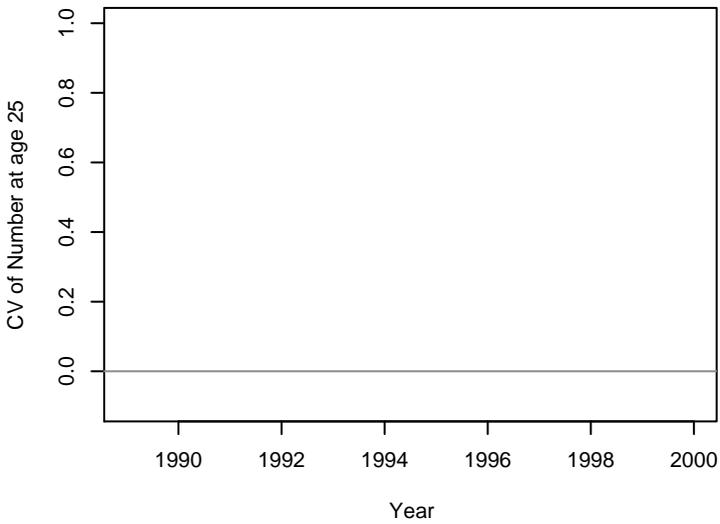


Figure WHAKE.8 (cont.)

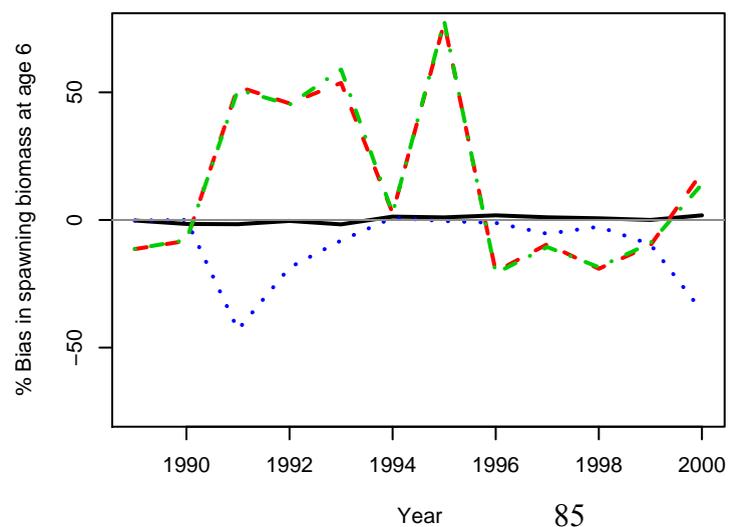
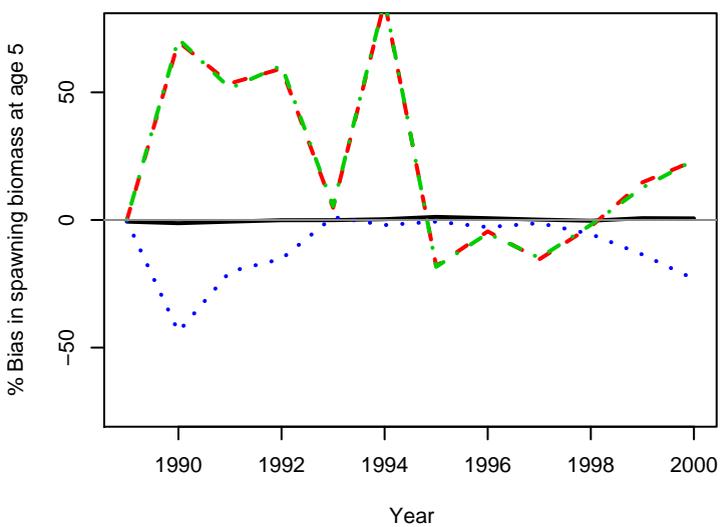
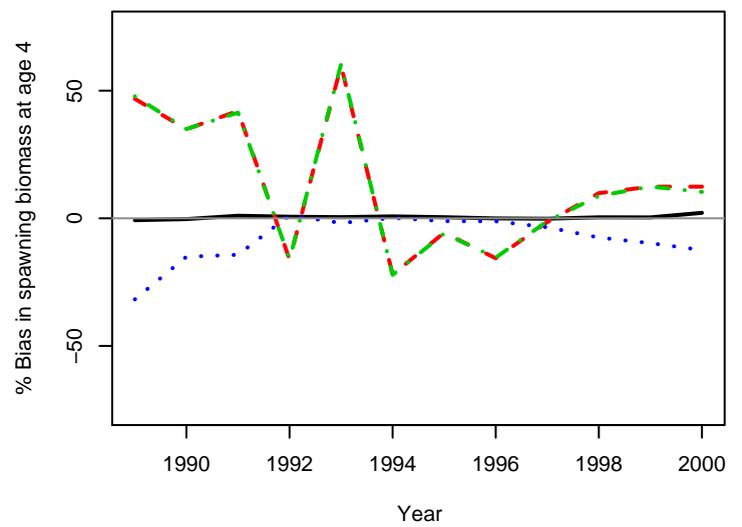
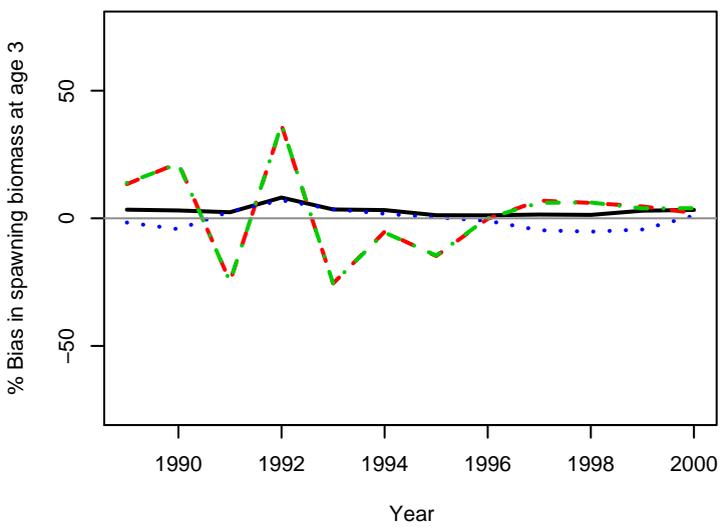
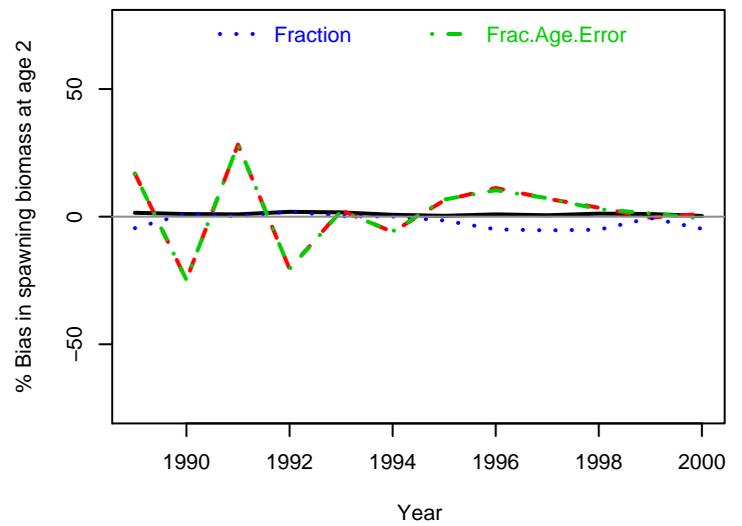
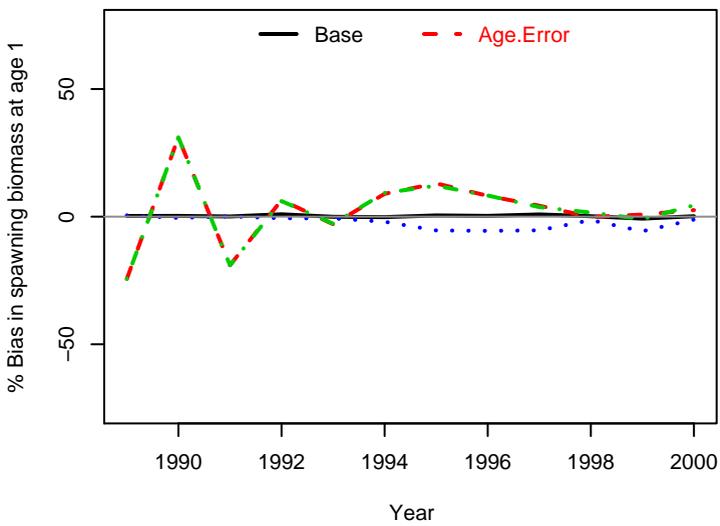


Figure WHAKE.9 Bias in spawning biomass at age for all simulated data sets using VPA.

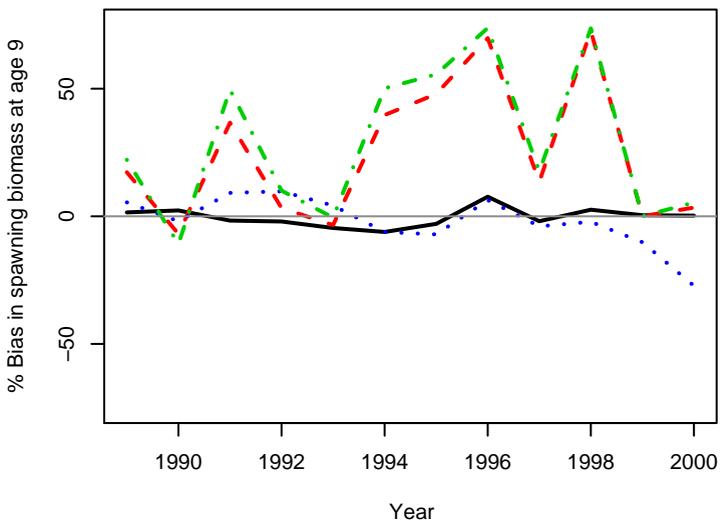
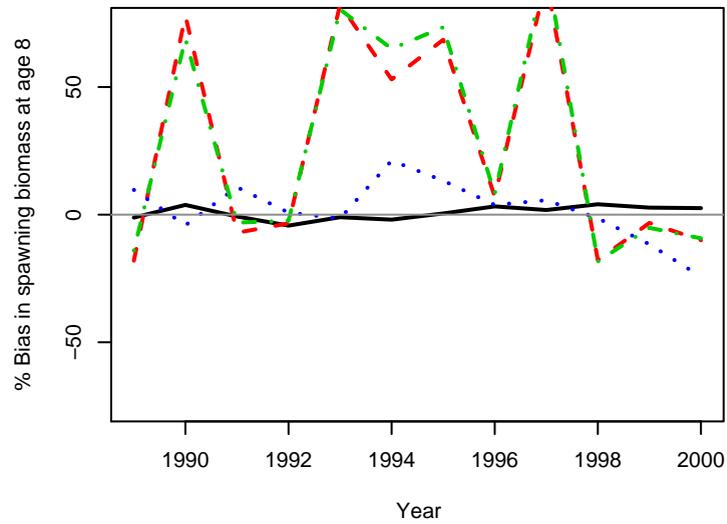
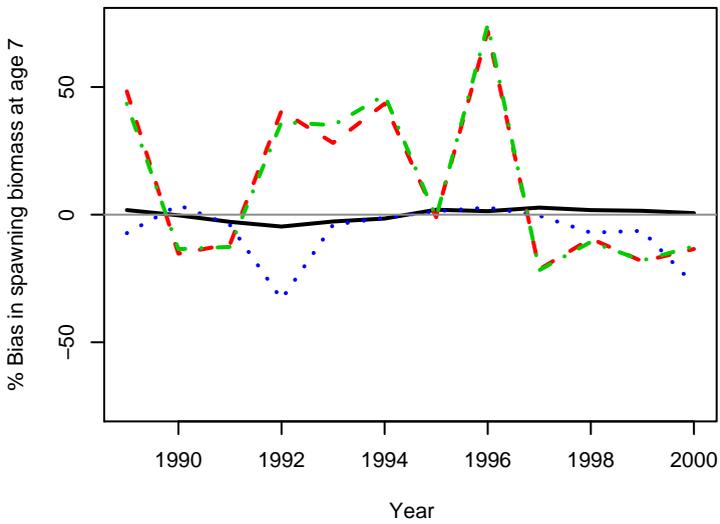


Figure WHAKE.9 (cont.)

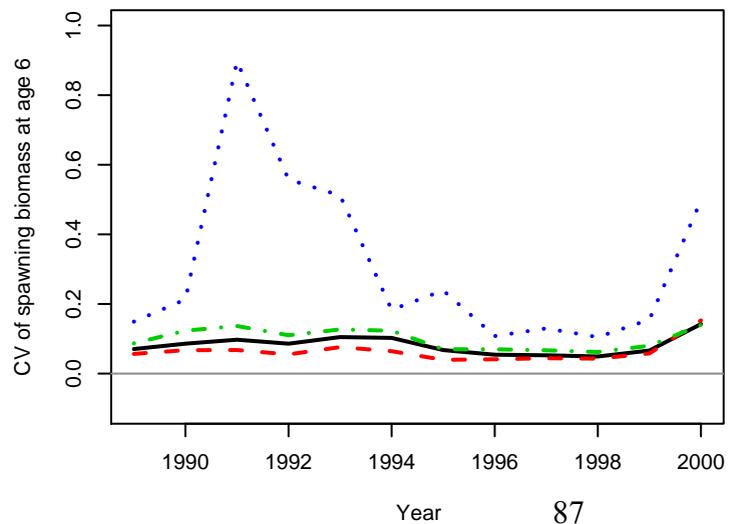
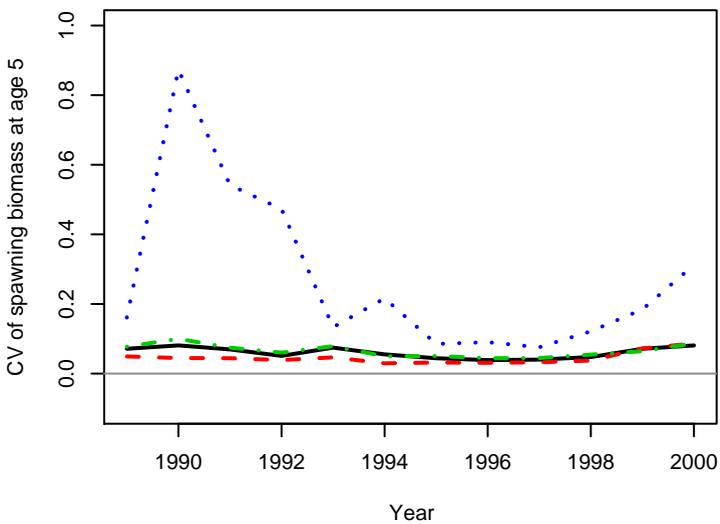
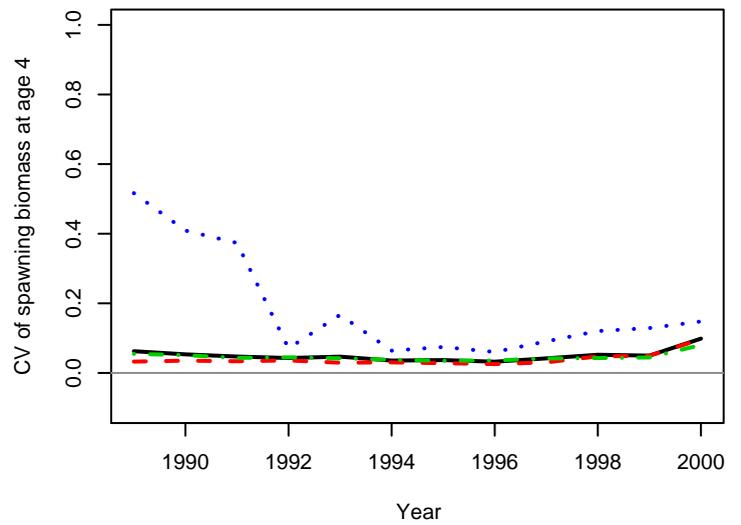
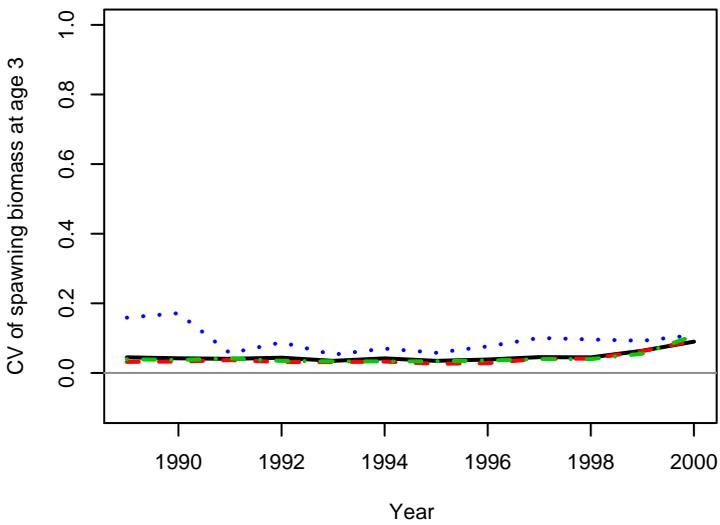
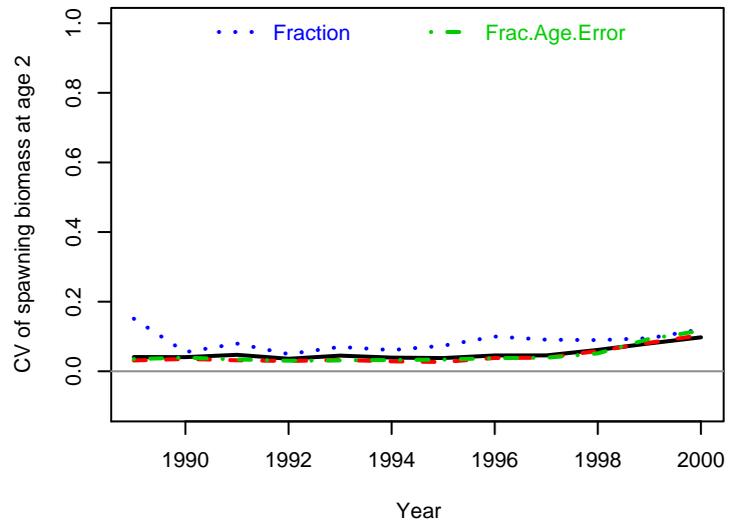
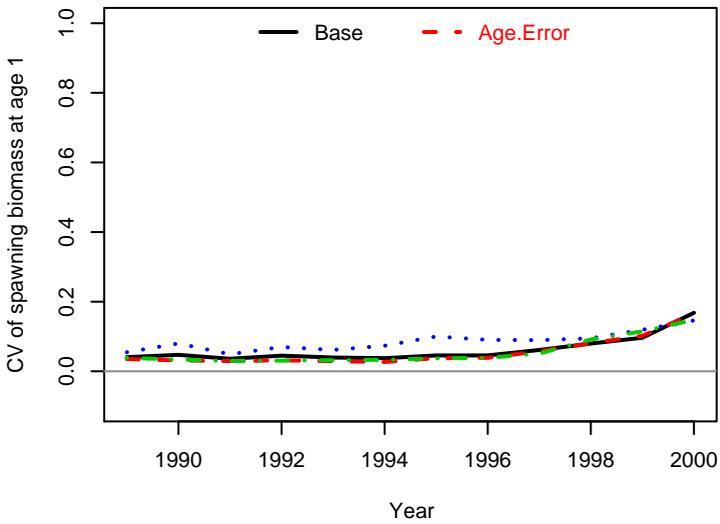


Figure WHAKE.10 CV of spawning biomass at age for all simulated data sets using VPA.

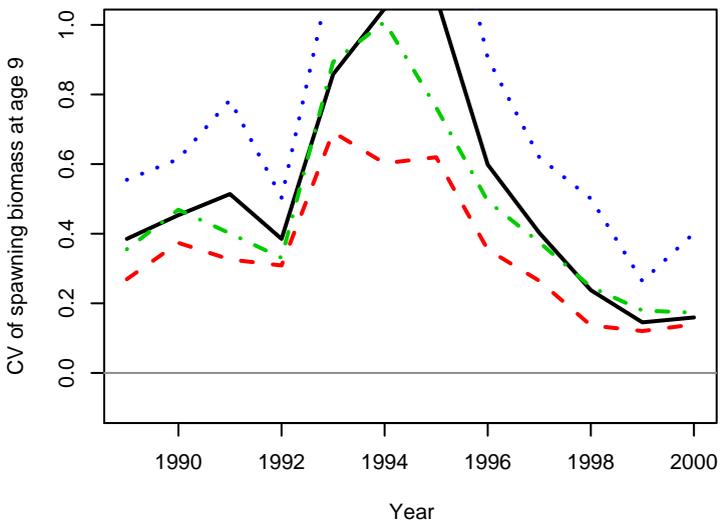
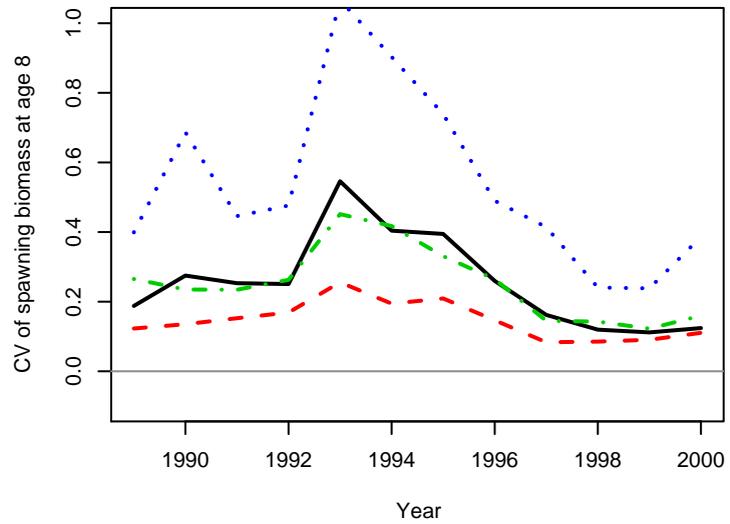
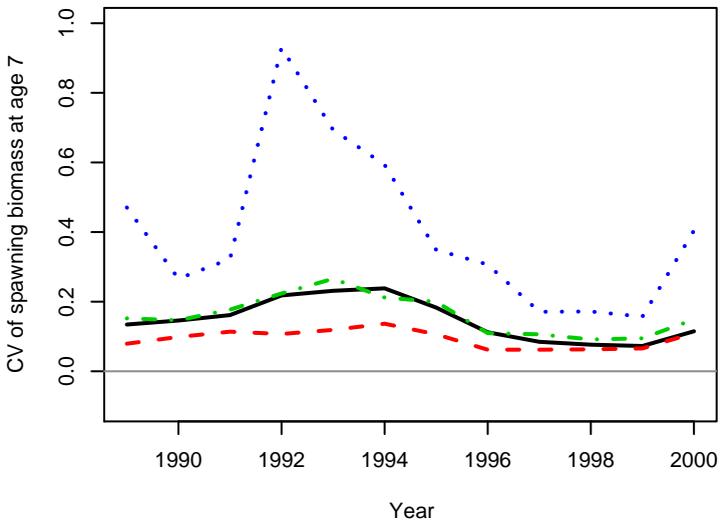


Figure WHAKE.10 (cont.)

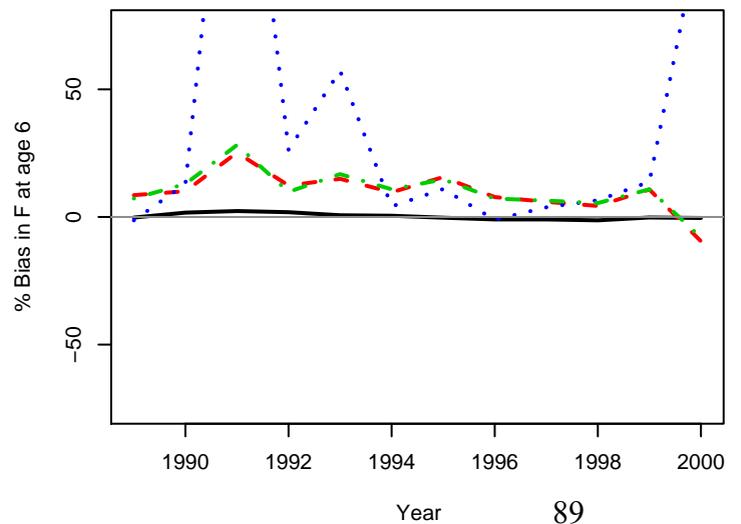
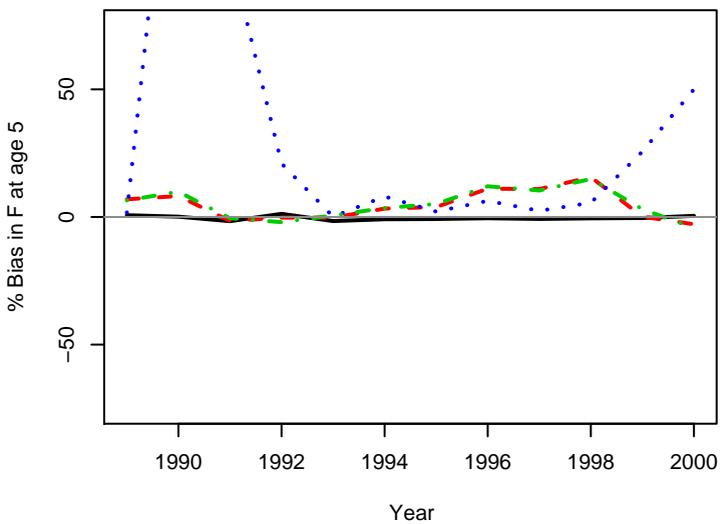
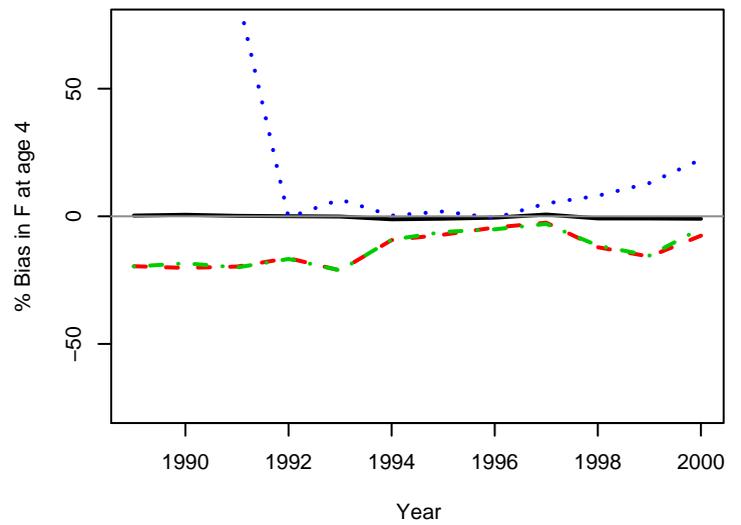
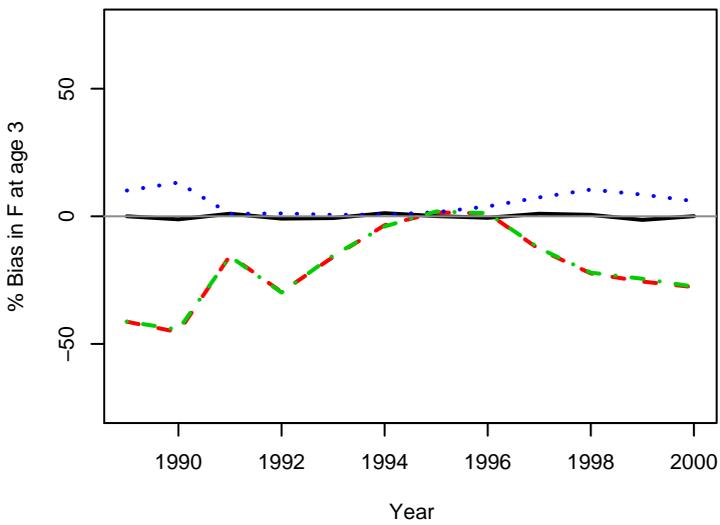
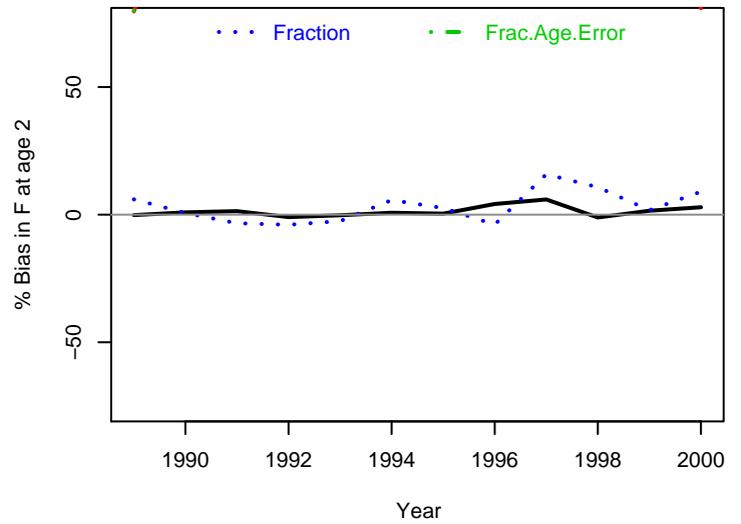
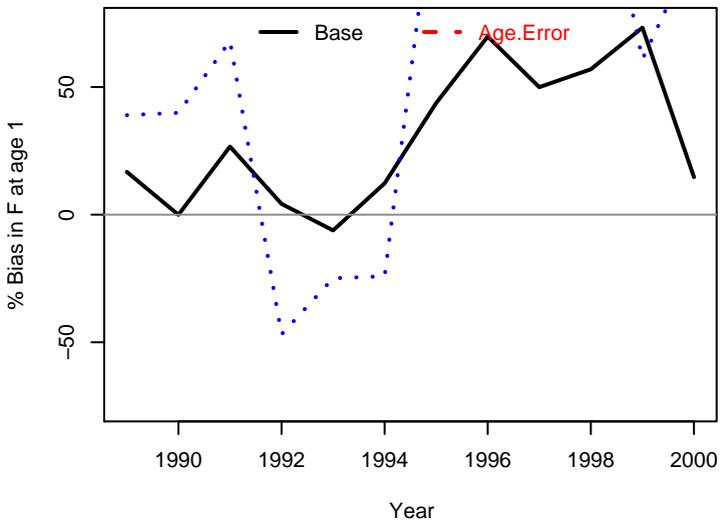
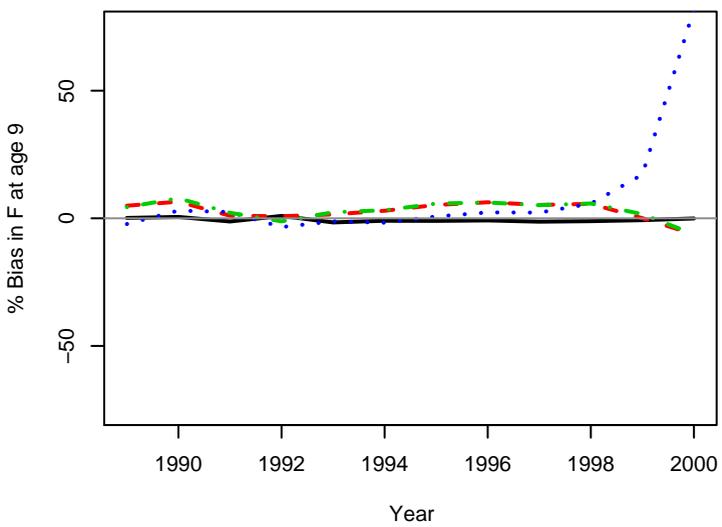
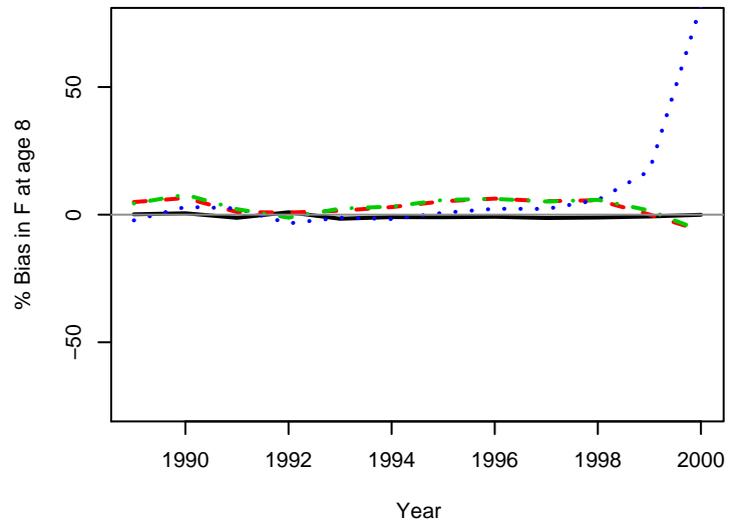
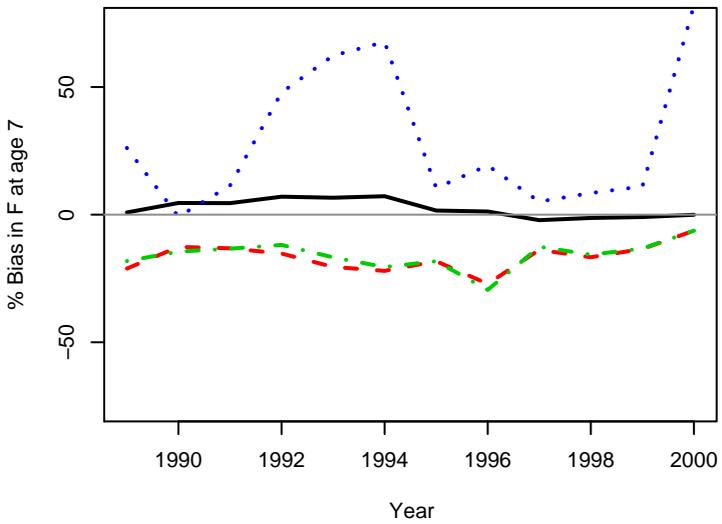


Figure WHAKE.11 Bias in F at age for all simulated data sets using VPA.



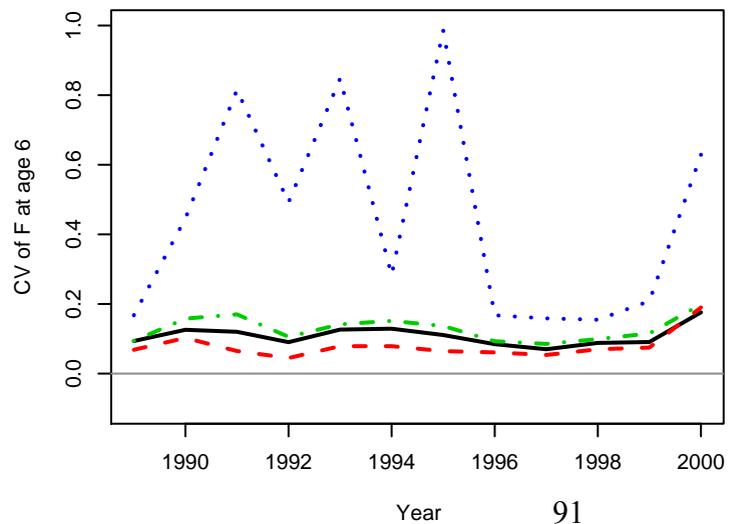
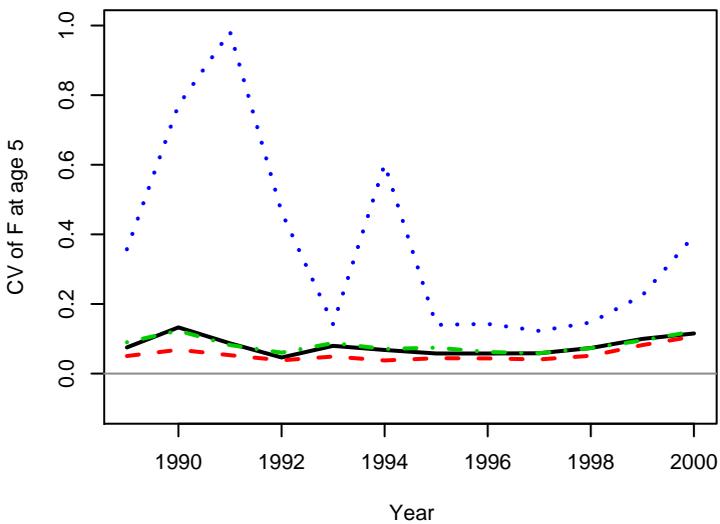
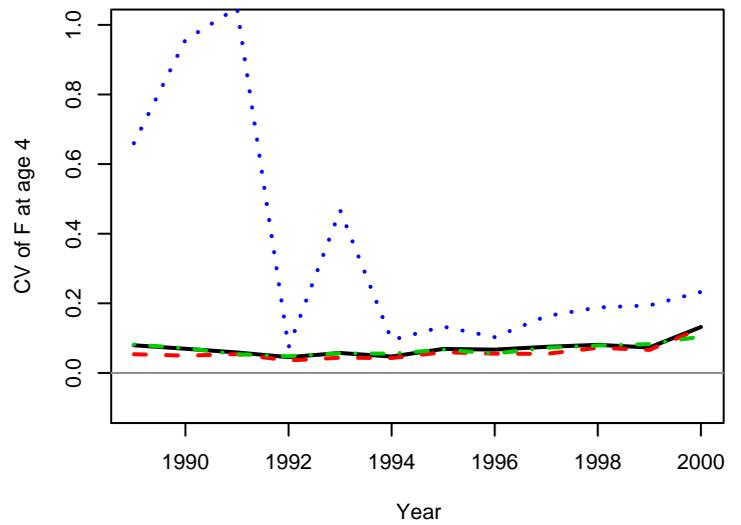
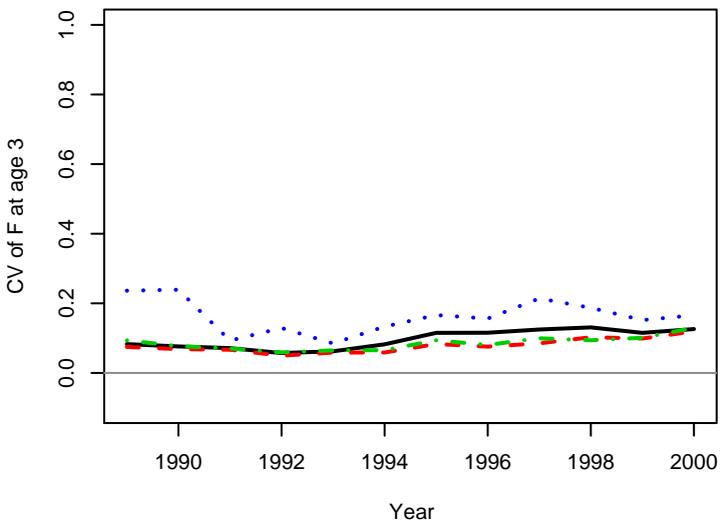
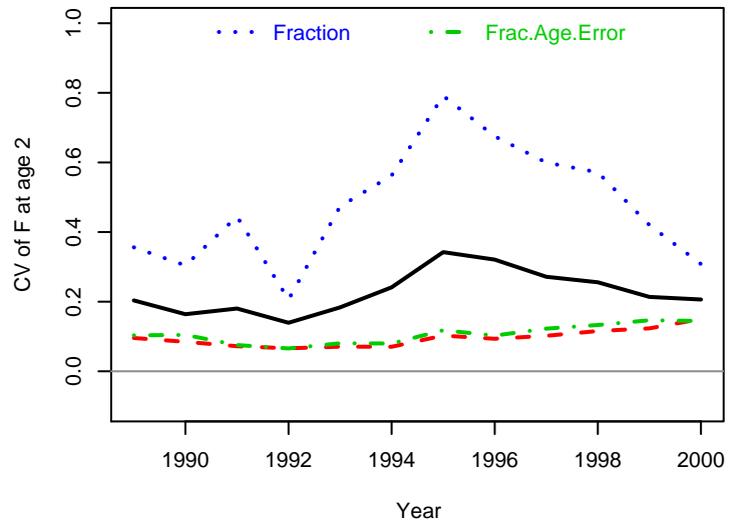
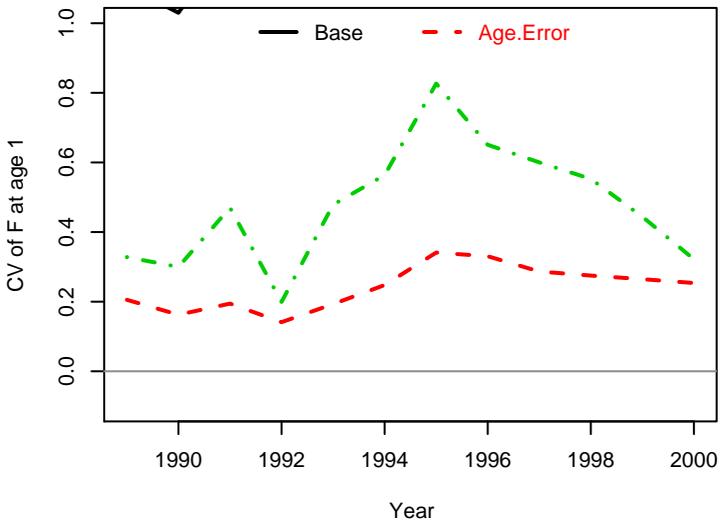
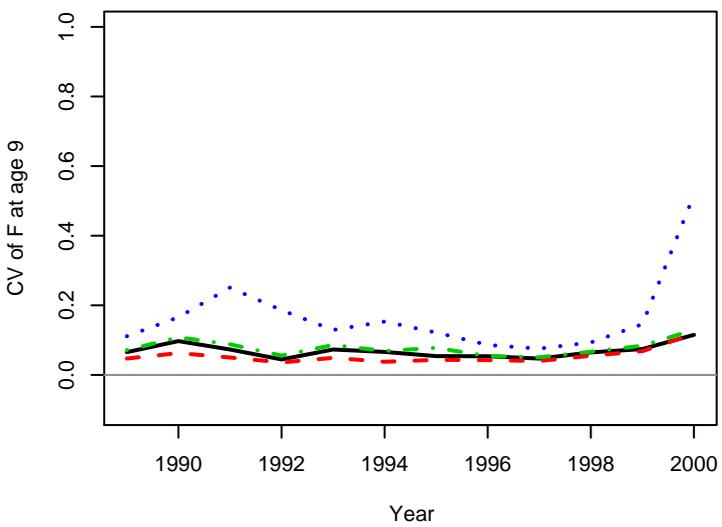
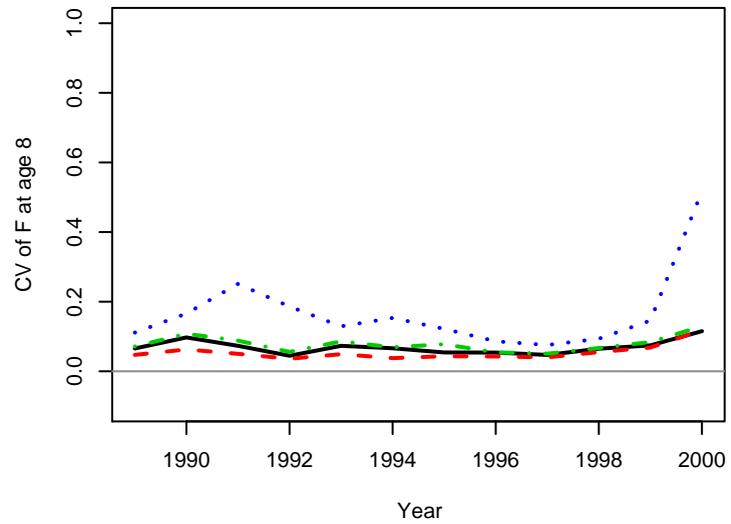
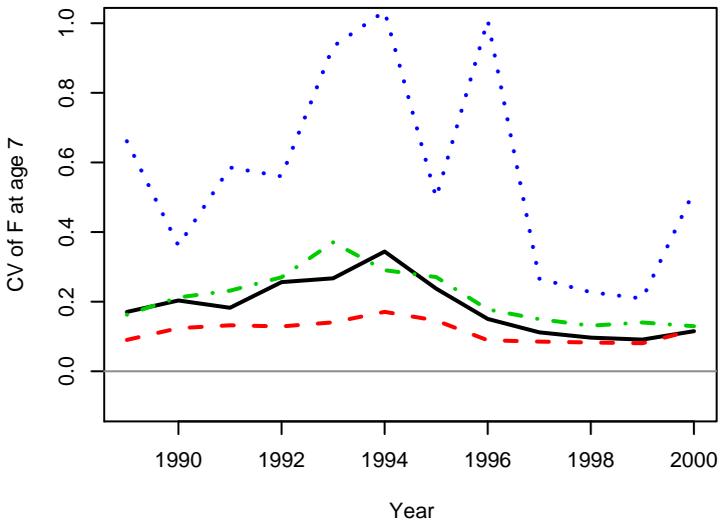


Figure WHAKE.12 CV of F at age for all simulated data sets using VPA.



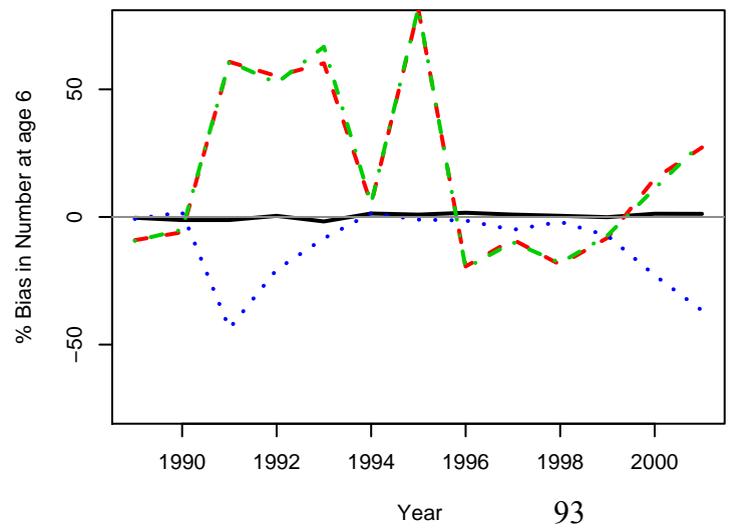
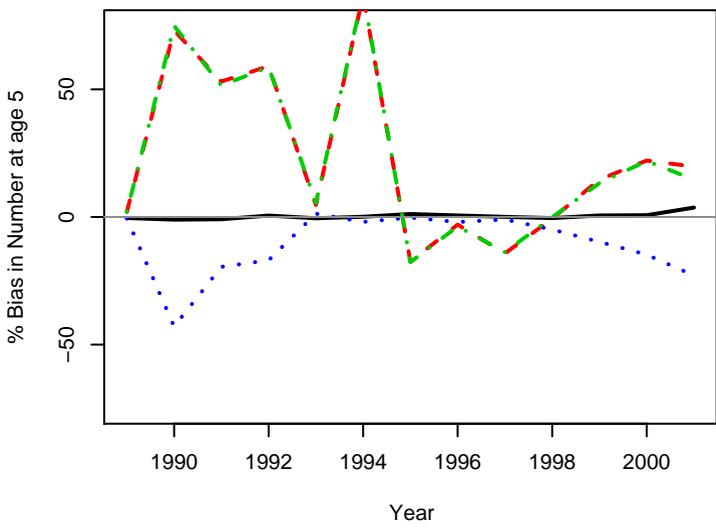
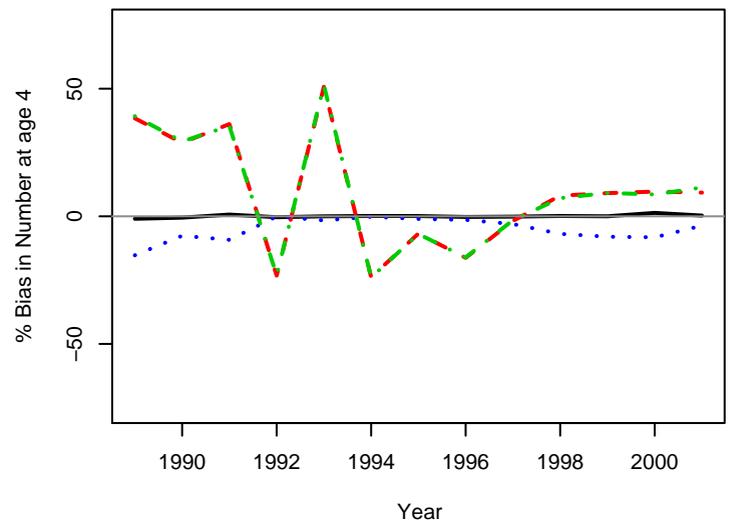
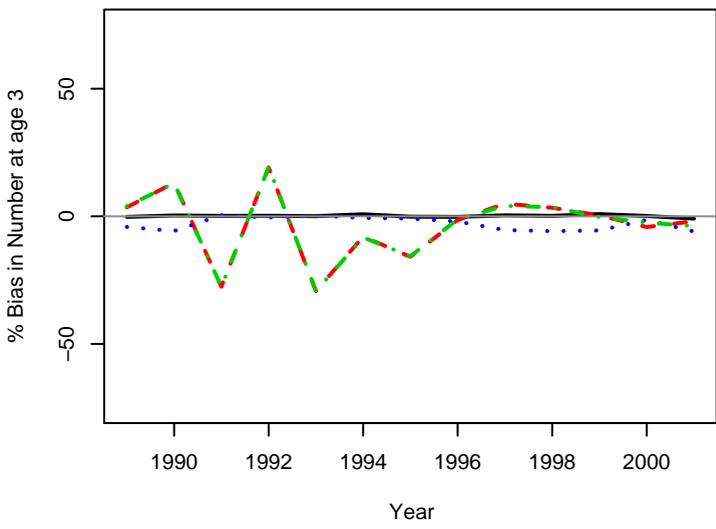
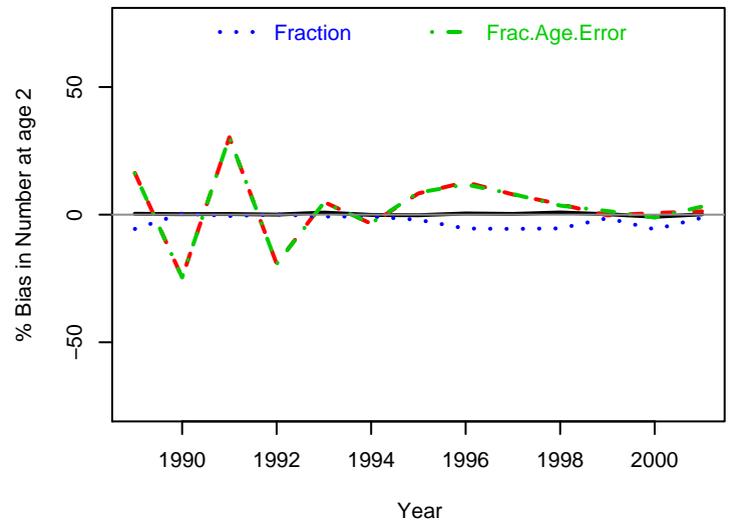
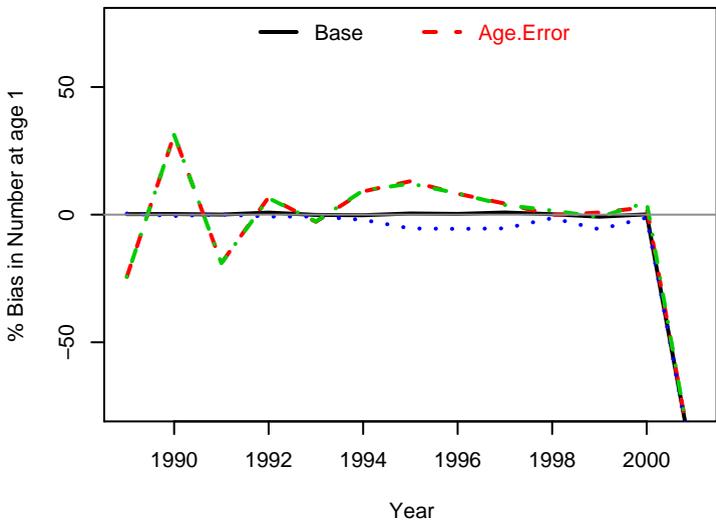


Figure WHAKE.13 Bias in number at age for all simulated data sets using VPA.

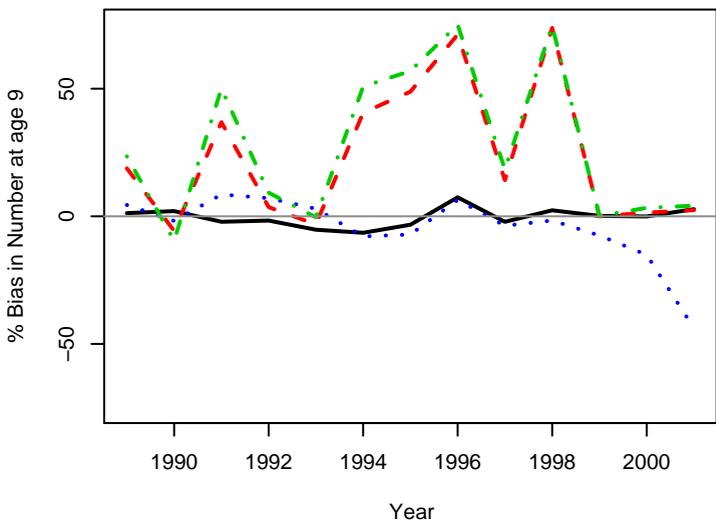
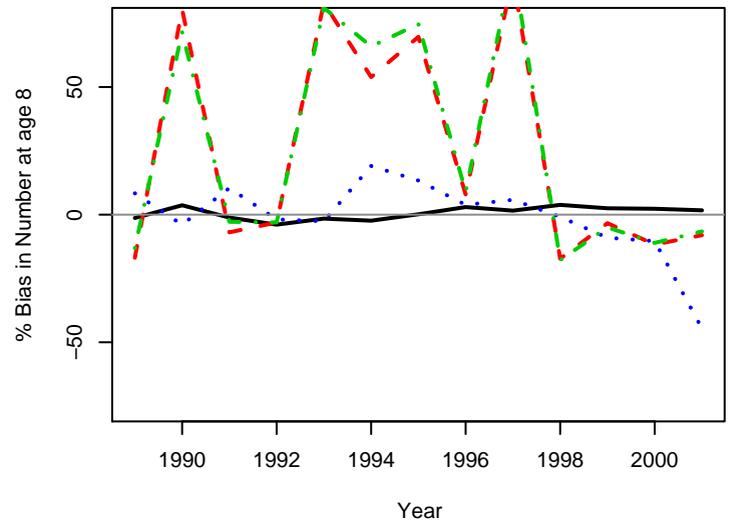
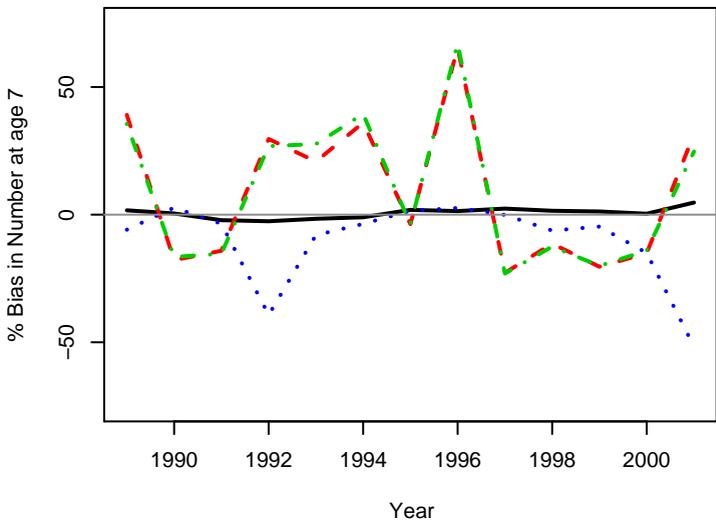


Figure WHAKE.13 (cont.)

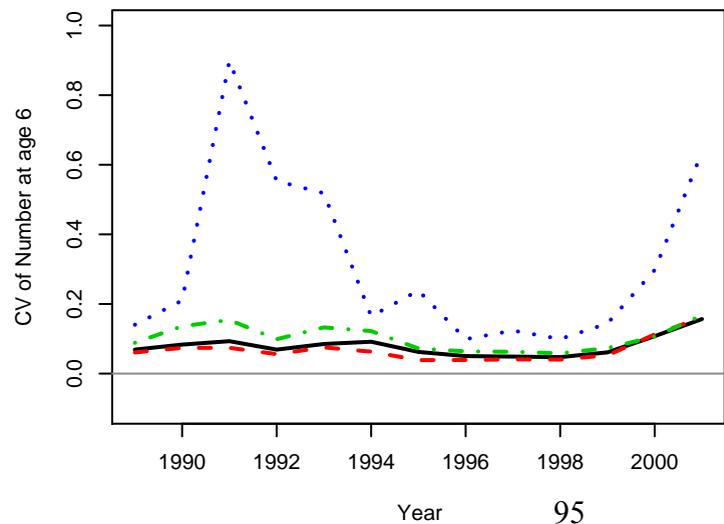
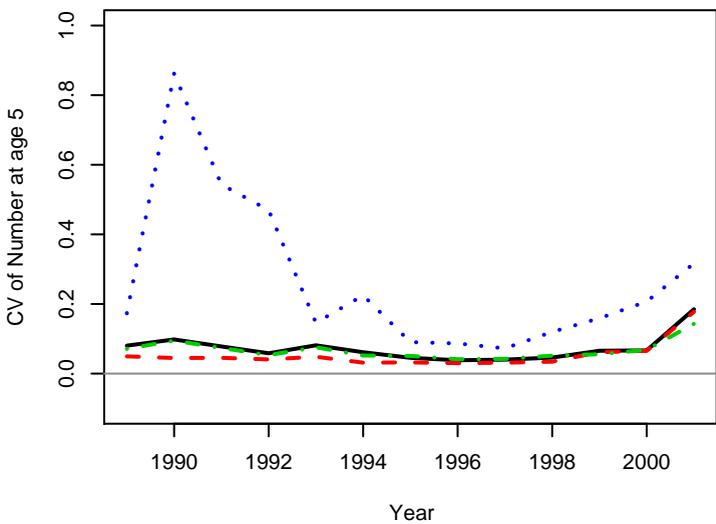
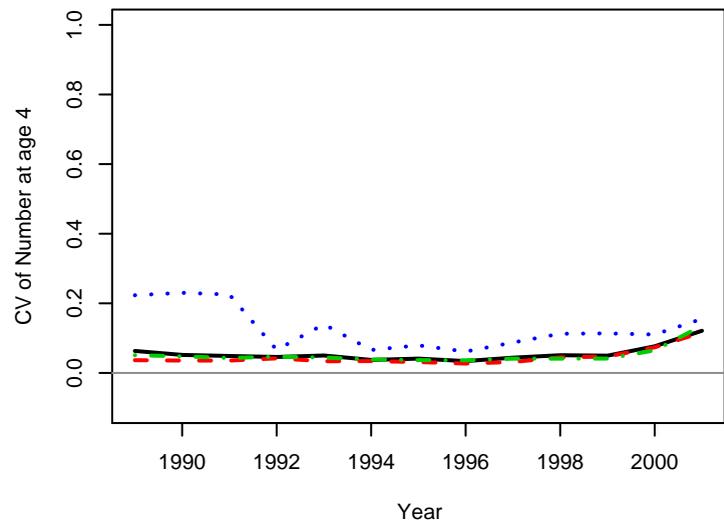
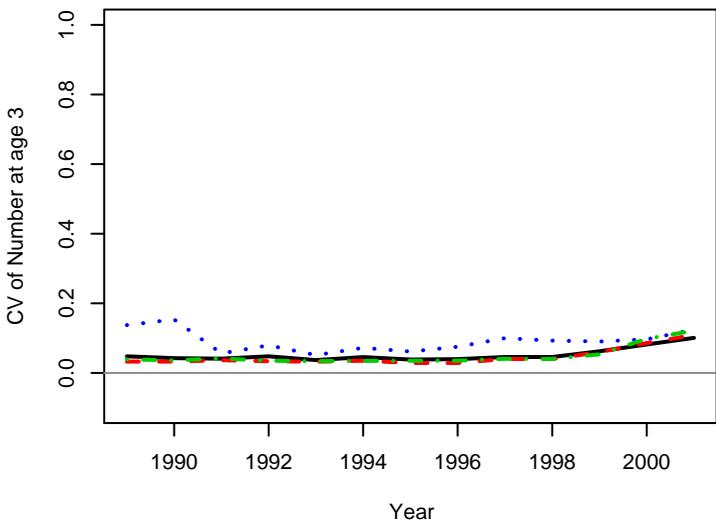
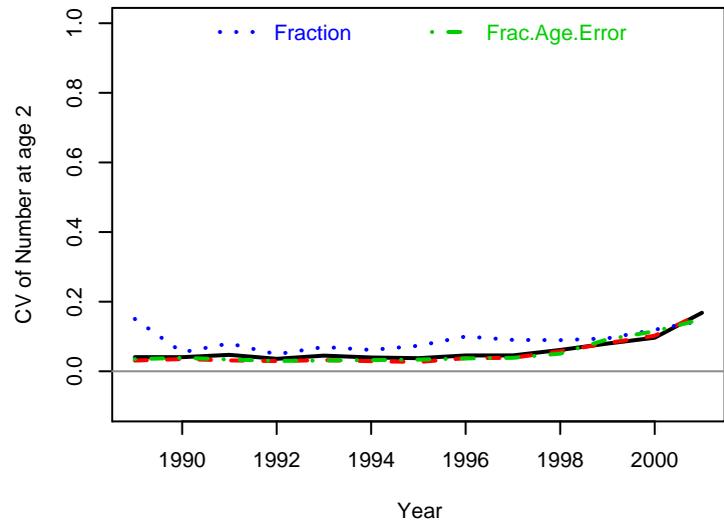
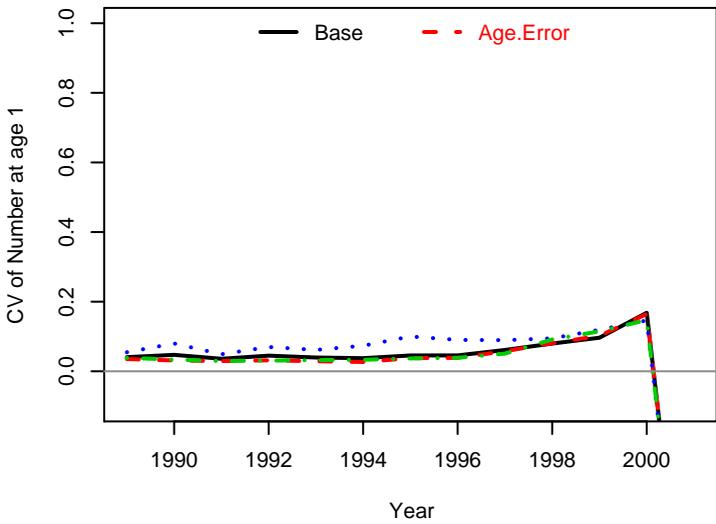


Figure WHAKE.14 CV of number at age for all simulated data sets using VPA.

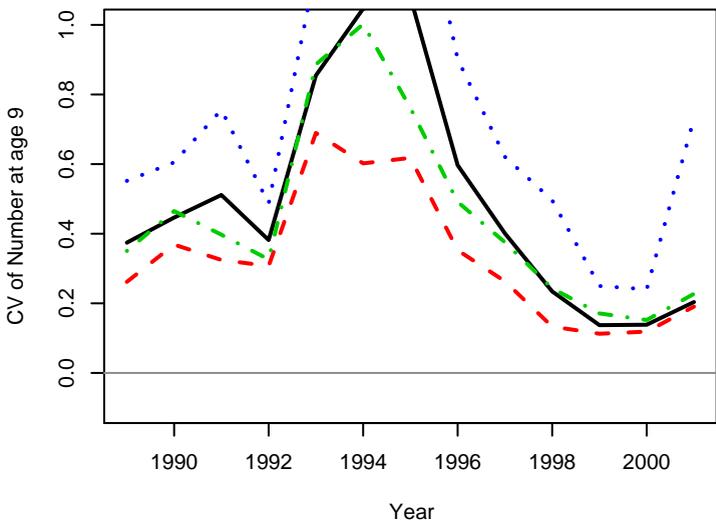
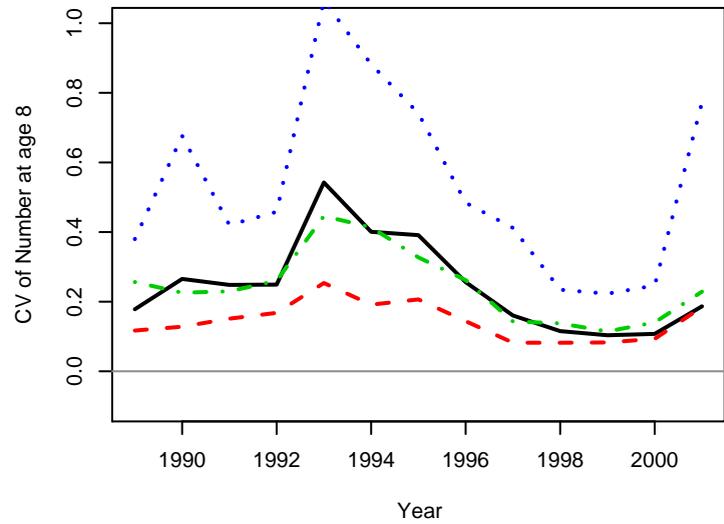
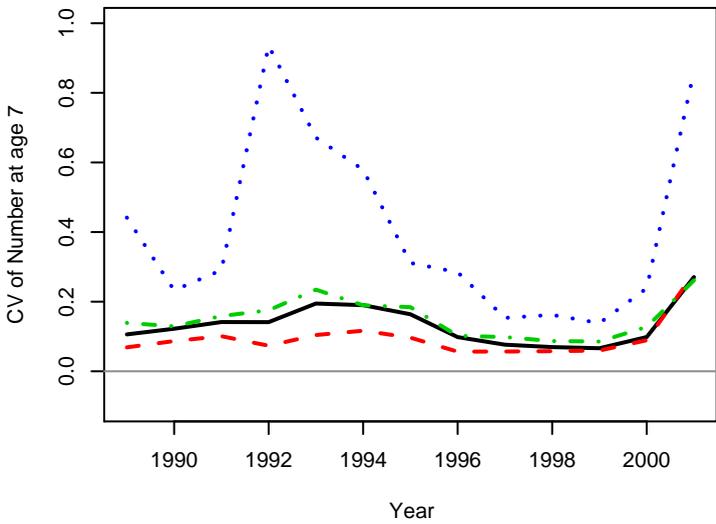


Figure WHAKE.14 (cont.)

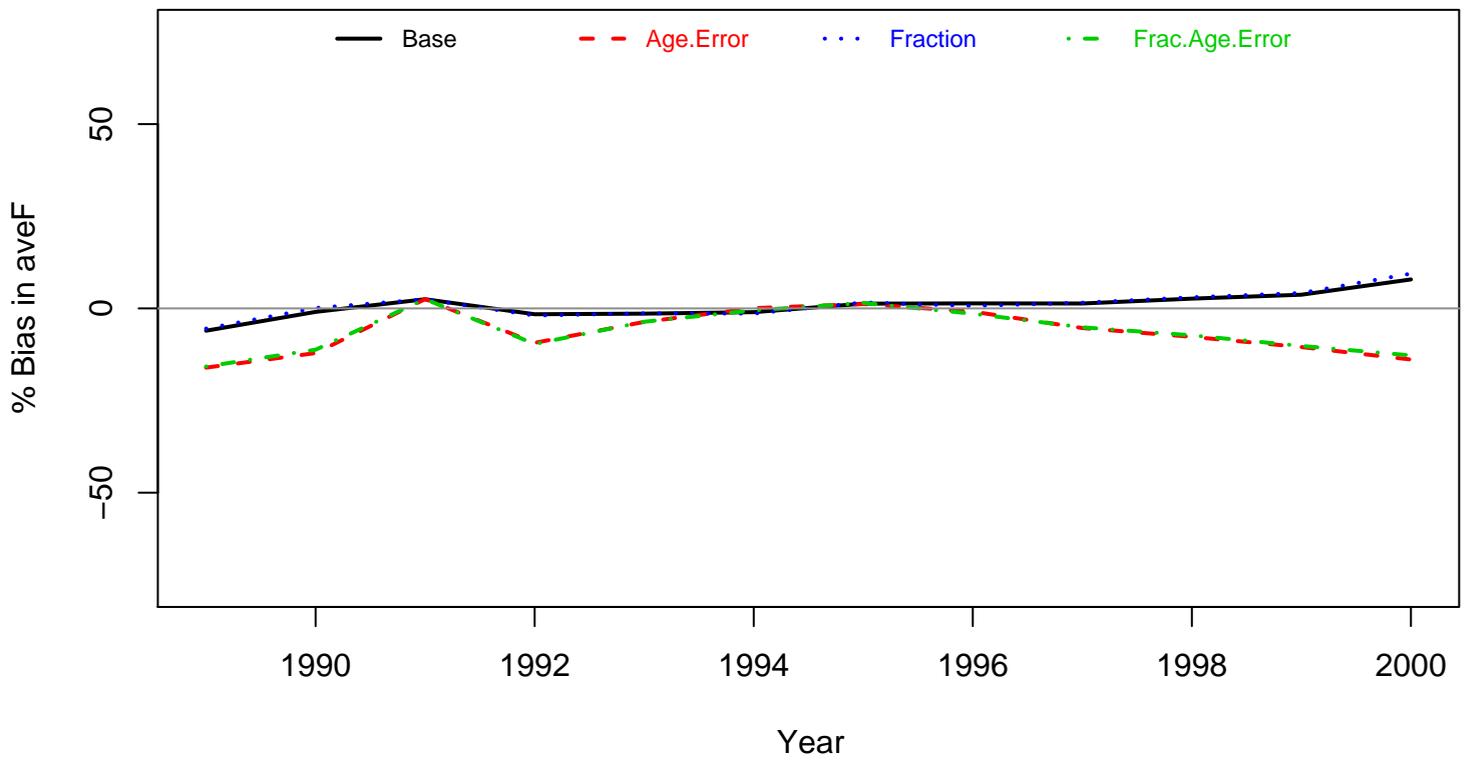


Figure WHAKE.15 Bias in spawning biomass and average F for all simulated data sets using ASAP.⁹⁷

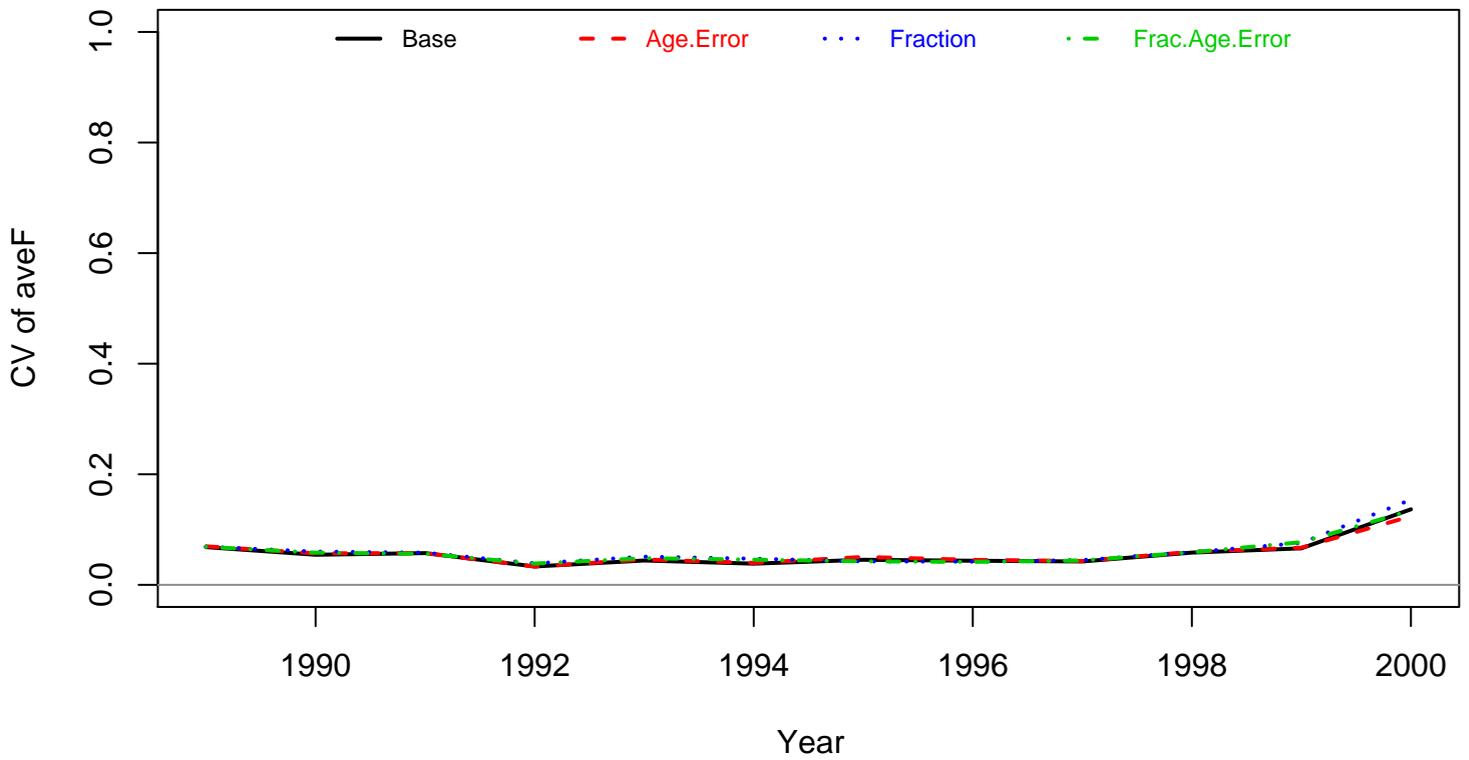
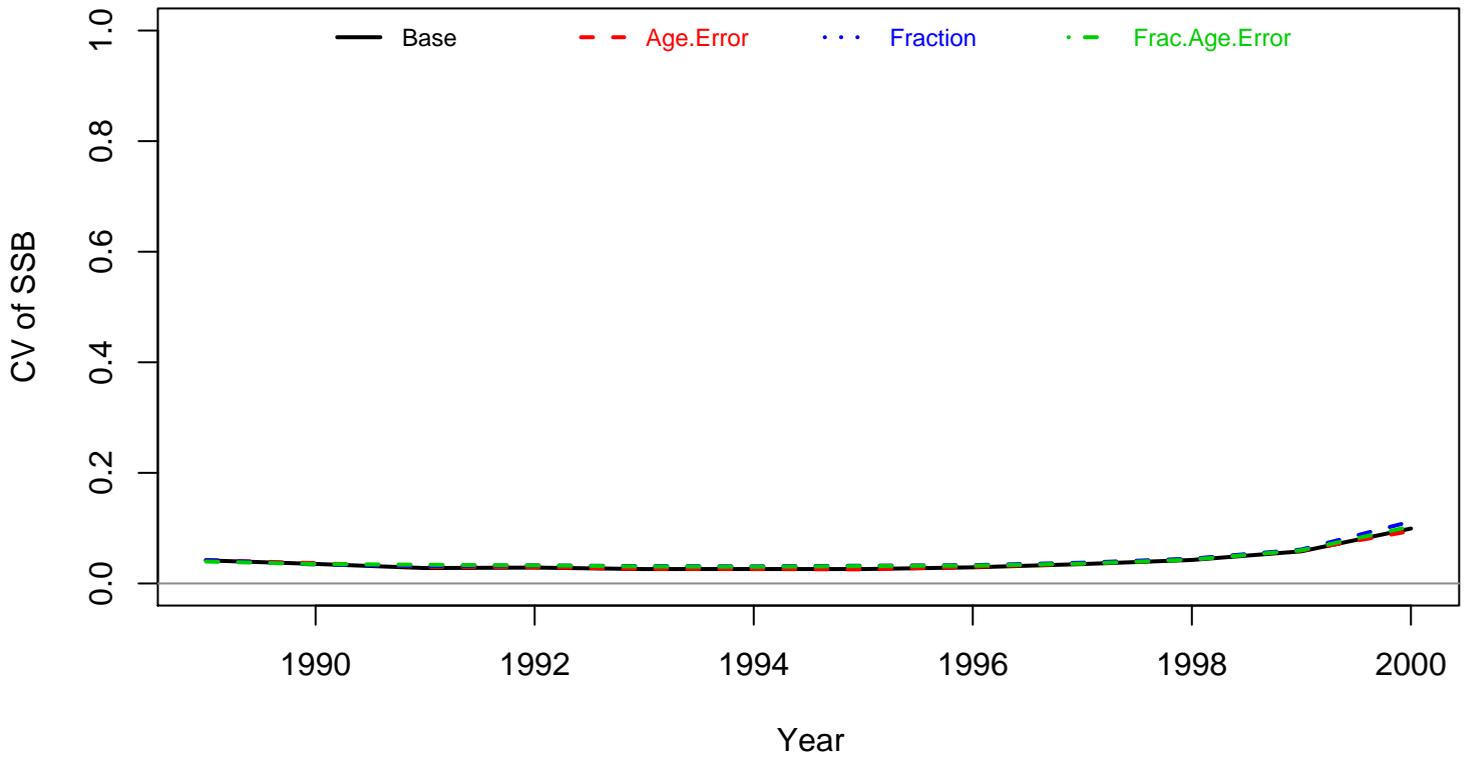


Figure WHAKE.16 CV of spawning biomass and average F for all simulated data sets using ASAP. ⁹⁸

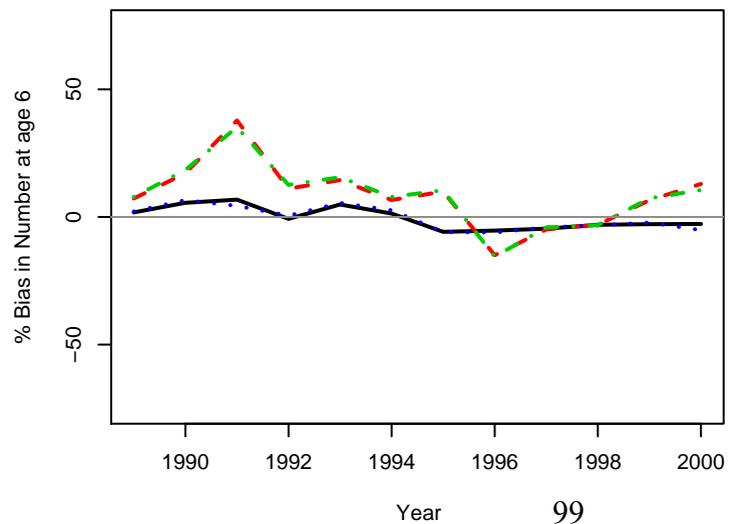
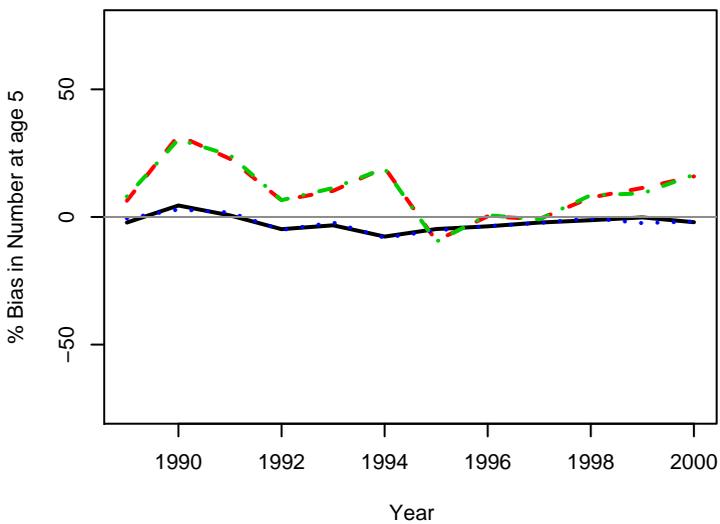
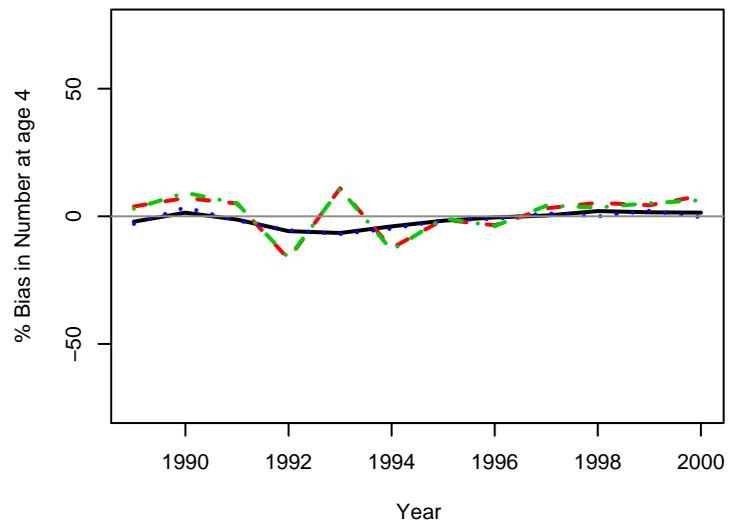
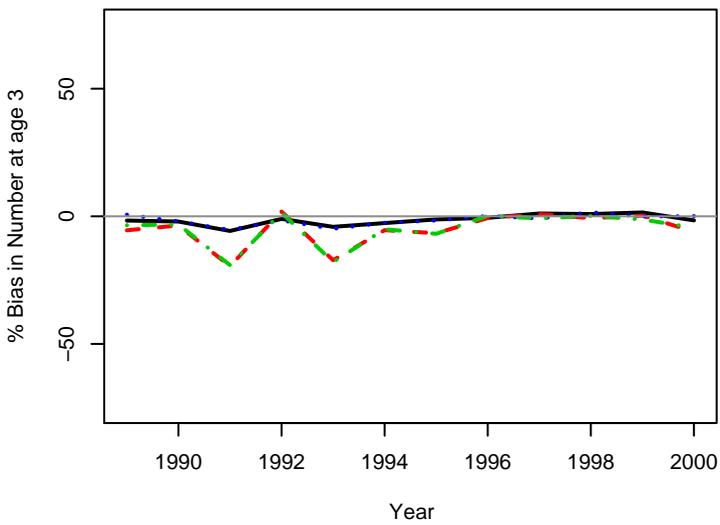
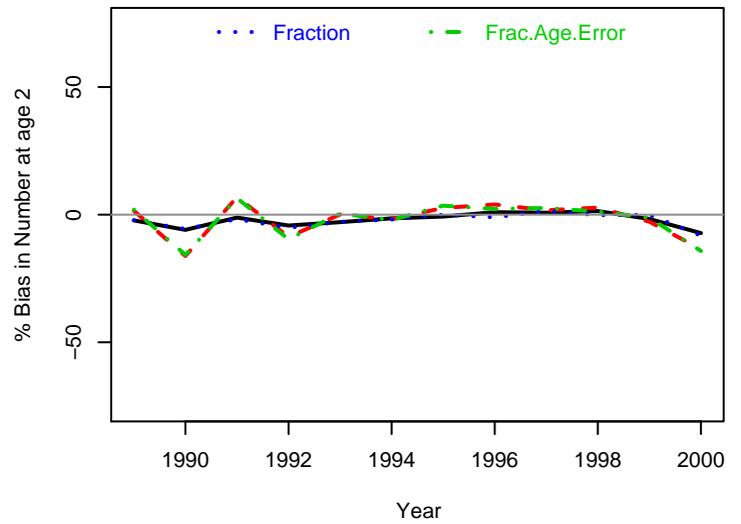
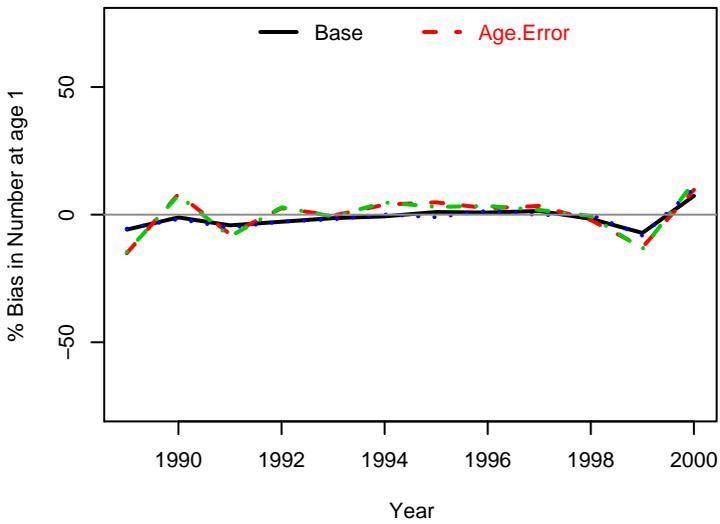


Figure WHAKE.17 Bias in number at age for all simulated data sets using ASAP.

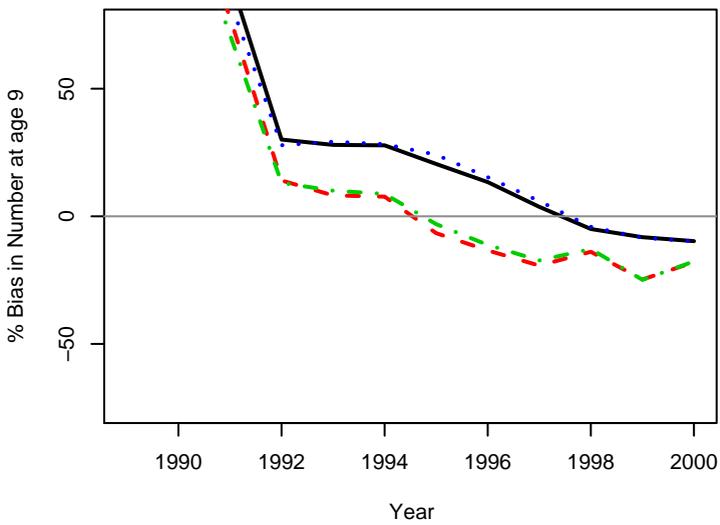
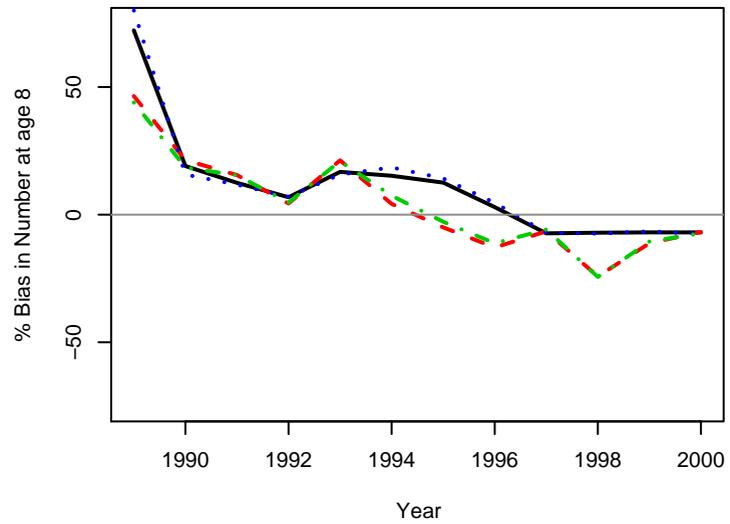
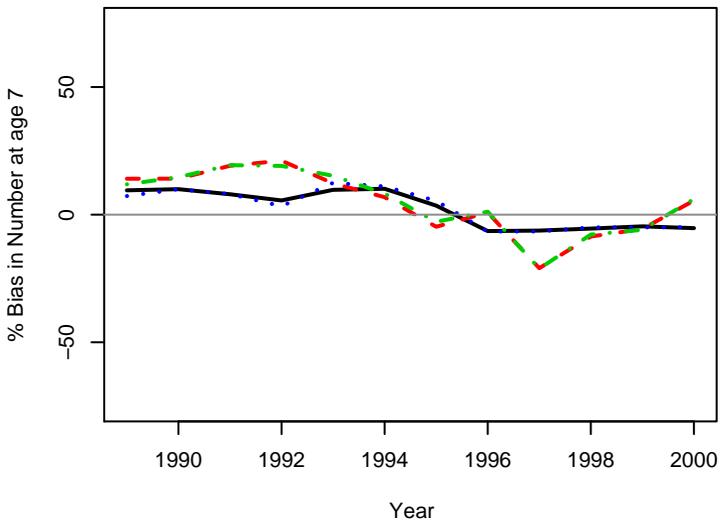


Figure WHAKE.17 (cont.)

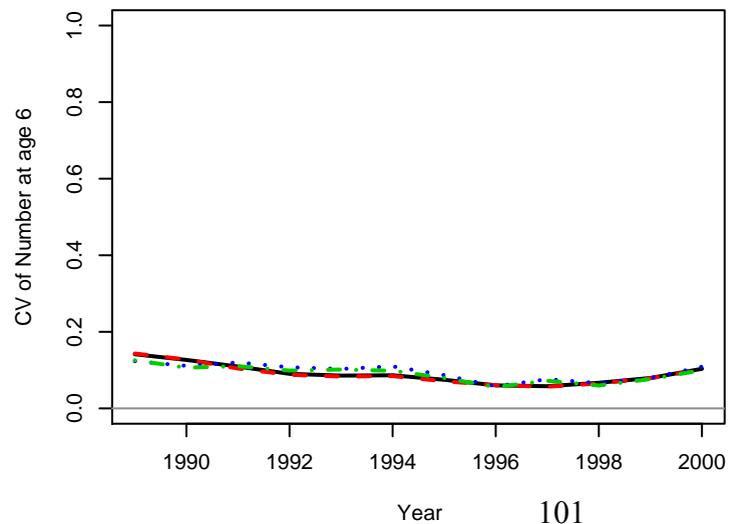
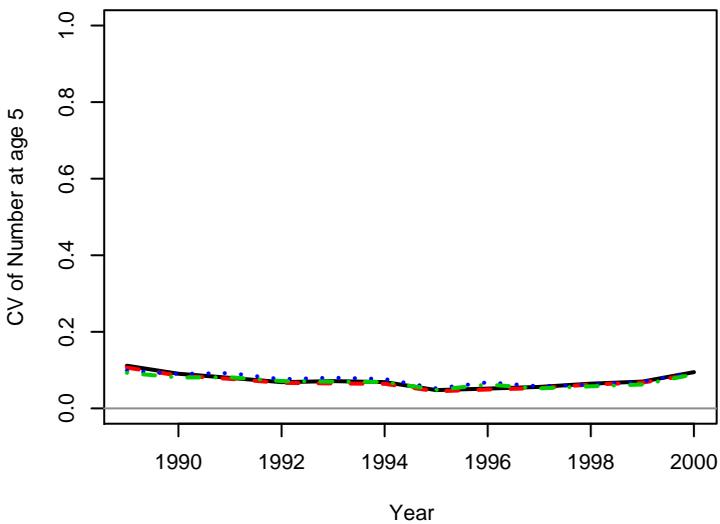
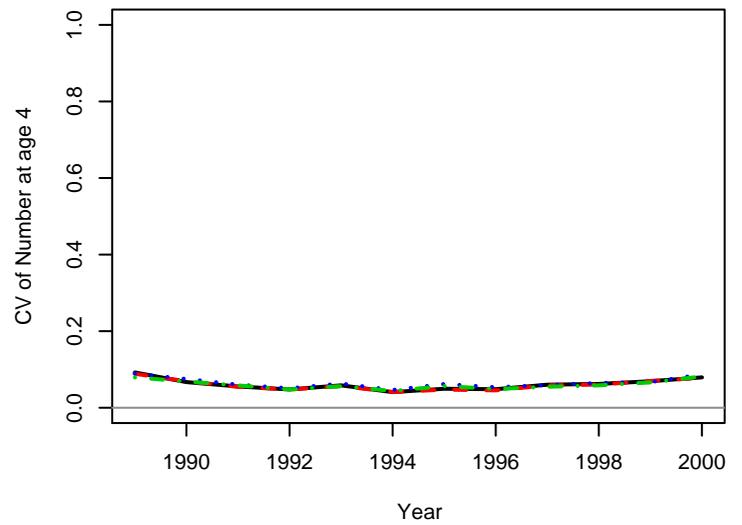
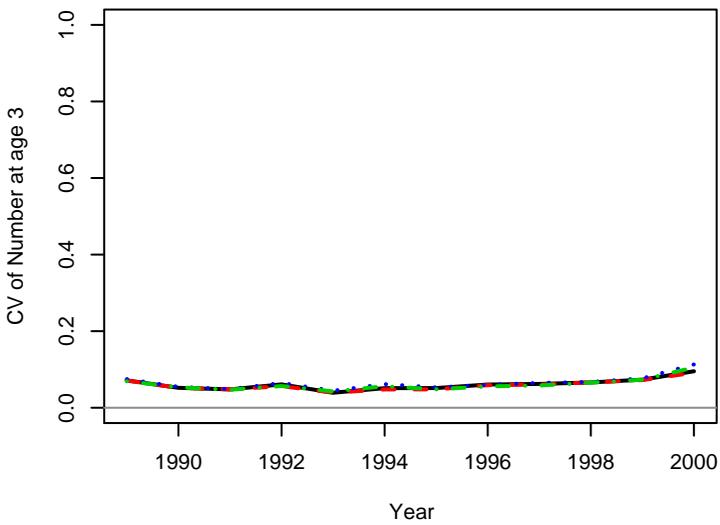
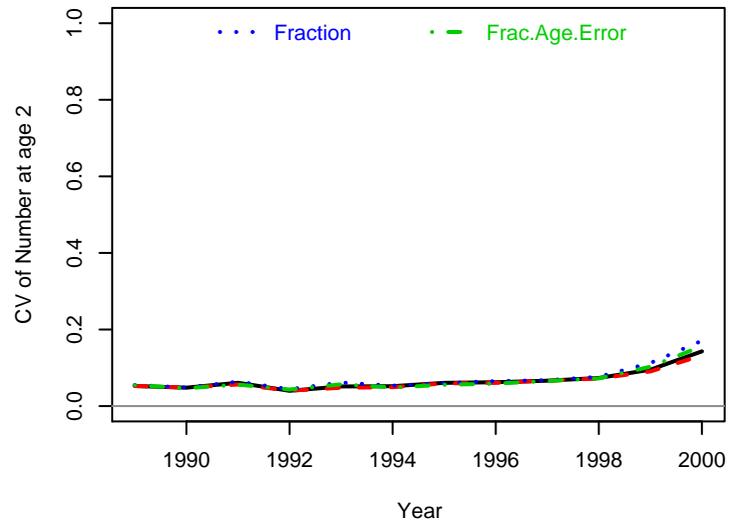
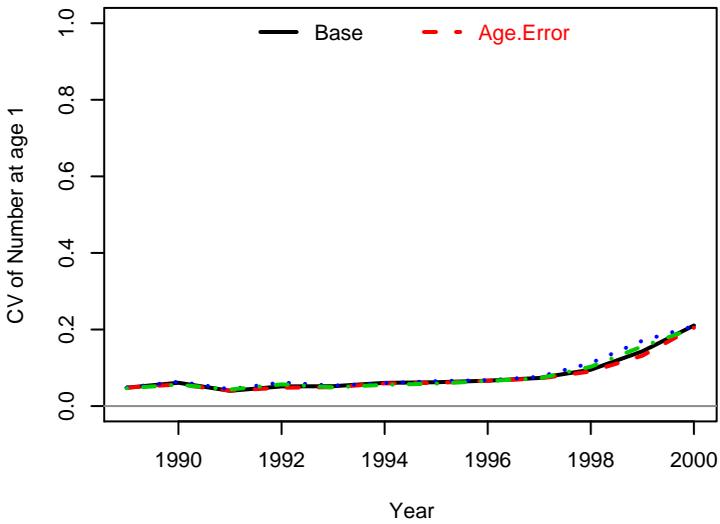


Figure WHAKE.18 CV of number at age for all simulated data sets using ASAP.

