NOT TO BE CITED WITHOUT PRIOR
REFERENCE TO THE AUTHOR(S)

Northwest Atlantic Fisheries Organization

Serial No. N5794 NAFO SCR Doc. 10/36

SCIENTIFIC COUNCIL MEETING – JUNE 2010

Random Retrospective Pattern in Fish Stock Assessment

by

Antonio Vázquez
Mónica Mandado

Instituto de Investigaciones Marinas
avazquez@iim.csic.es
mandado@iim.csic.es

Summary

Several indices were proposed to measure disagreement among results of a retrospective analysis. They were analyzed by numerical simulation in response to random variability in partial recruitment, catch at age numbers and survey indices. The potential use of a retrospective index as an indicator of accuracy in VPA results is explored.

Introduction

The retrospective problem is a systematic inconsistency among a series of estimates of population size, or related assessment variables, based on increasing periods of data (Mohn 1999). Legault (2009) studied the retrospective patterns produced by data with a particular characteristic, such as missing catch, an increase in natural mortality or changes in survey catchability. Inconsistency in retrospective results can also be produced by the underlying inaccuracy of input data and incompetence of the assumed VPA model between to explain stock dynamics, however a defined pattern or systematic deviations in retrospective is not expected if such underlying variability is at random. The objective of this paper is quantifying the inconsistency in the case these effects are random, and we will continue to call it "retrospective" even it has not a "pattern".

Three sources of retrospective were considered: inaccuracy in commercial catch at age, variability in survey indices at age, and disagreement between stock dynamics and the assumed VPA model due to partial recruitment does not fit assumptions. There are some other sources of variability in VPA analysis, i.e. the assumed constancy of natural mortality among years and ages, but they will not be considered here.

Inaccuracy of input data: commercial catch at age and survey indices at age. Commercial catch at age are usually considered exact data because of requirements of most common VPA methods, but, even in the hypothetical case where total catches in weight were accurately reported, length sampling and otoliths reading introduce an unavoidable variability in the final catch at age matrix. Survey indices at age have also an inherent high dispersion because of the haul to haul high variability, and because of length sampling and otoliths reading. Cerviño and Vázquez (2004) calculate CVs in the range of 18 to 129% for abundances at age from the Flemish Cap stratified random bottom trawl survey.

Model misspecification: disagreement between stock dynamics and the assumed VPA model is other unavoidable source of retrospective, when the model does not capture the real dynamics of the stock. In order to analyze such effects, ADAPT (Gavaris 1988) was preferred to other VPA methods because each one of its model assumptions is easily related to one quantifying parameter. However, the disagreement among results of a retrospective analysis and its measurement occurs with any VPA method; the use of ADAPT is not essential in this analysis even some of its properties emerge or, more precisely, performance of different ADAPT formulations can be judged by its retrospective characteristics.

The usual constraint used in ADAPT is that F at the last age is a function of the F at younger ages in the same year. The most simple case is $F_n = F_{n-1}$, being n the last true age, but F_n is usually assumed equal to an average of F at some more younger ages. In testing the effect of violation of this assumption, simulated populations can be created with a quantified disagreement between it and the model assumptions. Simulation based on ADAPT allows an independent evaluation of the effects of the two sources of random variability we consider: disagreement between stock dynamics and the assumed VPA model and random inaccuracy of input data. Three different ADAPT formulations were used, which implies three different degrees in which data requires model assumptions, consequently some marginal conclusions will be also obtained on which method is more appropriate.

Methods

A numerical simulation (Vázquez *et al.* 2009) was used to produce a simulated fish population with the following characteristics, which makes it a hypothetical stock not far from some other real ones. As a general remark, we tried that conclusions be independent of the particular characteristics of the simulated population; this means, all used values being irrelevant. We just simulate a plausible population.

Years: 1975 - 2007
Ages: 1 - 14 (age 14 being a plus group)
M: 0.2

a) Fishing mortality

$spF_{a,y}$ – simulated population Fishing mortality at age *a* year *y*

$$spF_{a,y} = spAF_y * PR_a$$

being:

$spAF_y$ = simulated population's Annual Fishing mortality at year *y*
values: $spAF_y = 0.2 + 0.5 * e$ e = Equal [0,1]

PR_a: Partial Recruitment at age a
values: $PR_a = N(\mu = spPR_a, \text{s.d.} = sdPR \text{ parameter})$ limited to [0.0001, 2]

$spPR_a$: simulated population's Partial Recruitment at age *a*
values: 0, 0, 0, 0.1, 0.2, 0.5, 1, 1, 0.75, 0.5, 0.5, 0.5, 0.5, 0.5

The sdPR parameter (standard deviation of Partial Recruitment) allows a quantified disagreement of the simulated population to the VPA model assumptions.

The ADAPT model assumes that F at the last age is a function of Fs at some younger ages in the same year, based on the hypothesis that partial recruitment is equal for those ages. If the above sdPR parameter were zero the simulated population would fit to the model assumptions, but the bigger the parameter, the bigger the discrepancy between them.

b) Abundances

$spN_{a,y}$: simulated population's abundance at age *a* year *y*

$$spN_{a+1,\,y+1} = spN_{a,y} / \exp(spF_{a,y} + M)$$

$spN_{1,y}$: simulated population's abundance of each cohort at age 1
$spN_{1,y}$ = Equal [0,10000] - random equally distributed numbers in the said interval

$spN_{a,1}$: simulated population's abundance of each cohort in year 1 at age a
$spN_{a,1} = e * 0.6^{a-1}$ e = Equal [0,10000]

c) Captures

$spC_{a,y}$ – simulated population's Catch in number at age *a* and year *y* (exact figure)

$$spC_{a,y} = (spN_{a,y} - spN_{a+1,\,y+1}) * spF_{a,y} / (spF_{a,y} + M)$$

$C_{a,y}$ – reported catch in number at age *a* and year *y*

$$C_{a,y} = spC_{a,y} * \varepsilon$$

$$\varepsilon = \log N(\mu = 1, cvCA)$$

cvCA = parameter (coefficient of variation of commercial Catch at Age)

d) Survey indices

Two independent and different surveys were simulated; even they are not identified with sub-indices in the following formulas:

$spI_{a,y}$ – simulated population's survey Indices for age *a* and year *y* (exact figure)

$$spI_{a,y} = spN_{a,y} * Q$$

being:

Q = survey catchability (for two fleets)

values: 0.02, 0.02, 0.03, 0.04, 0.1, 0.2, 0.2, 0.2, 0.1, 0.05, 0.02, 0.02, 0.02, 0.02

values: 0.3, 0.3, 0.25, 0.25, 0.2, 0.15, 0.15, 0.1, 0.05, 0.02, 0.01, 0.01, 0.01, 0.01

$I_{a,y}$ – observed survey indices for age *a* and year *y*

$$I_{a,y} = spI_{a,y} * \varepsilon$$

ε = logN (μ = 1, cvSI)

cvSI = parameter (coefficient of variation of Survey Indices)

The cvCA and cvSI parameters allow setting a predefined level of inaccuracy between reported catches and the exact figures, as well as between survey indices and the simulated abundance at age. Once the simulated population has been defined and it abundance at age, catch at age and survey indices has been calculated according to catch equation and prescribed parameters, observed catch at age and survey indices are calculated from the exact figures multiplying by the random factor ε:

e) SSB and Exploitable biomass

SSB_y – SSB at year *y*

$$SSB_y = \sum_{a=1,an} (spN_{a,y} * spSW_{a,y} * MO_{a,y})$$

being:

$spSW_{a,y}$ – Stock weight at age *a* and year *y*

values: 0, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1.1, 1.5, 1.9, 2.6, 3.1, 3.9, 5

$MO_{a,y}$ – Maturation ogive at age *a* and year *y*

values: 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1

FB_y – Exploitable biomass at year *y*

$$FB_y = \sum_{a=1,an} (spN_{a,y} * spSW_{a,y} * spPR_a)$$

In summary, three parameters are used: one to control departure of the simulated population from the assumed model (sdPR) and two others to generate assessment data (catch at age numbers and survey indices at age) with prescribed inaccuracy (cvCA and cvSI). Standard deviation was used for PR to maintain variability independent of size; however coefficient of variation was used for catch at age and survey indices to allow variability increase with size. Their tested values are:

sdPR : 0.0001, 0.1, 0.2, 0.5, 1.0
cvCA: 0.0001, 0.1, 0.2, 0.5, 1.0
cvSI : 0.0001, 0.1, 0.2, 0.5, 1.0

Catch at age and survey indices produced in this way were later analyzed with three different ADAPT formulations, all of them completed with a retrospective analysis.

The three ADAPT's formulations differ on how many parameters are calculated:

- ADAPT-10: 10 parameters, which are the survivors of all cohorts occurring in the last year.
- ADAPT-9: 9 parameters: the survivors of all cohorts occurring in the last year excluding the oldest age.
- ADAPT-27: 27 parameters: the survivors of all cohorts (both occurring in the last year or elsewhere).

Common characteristics were:

- Objective function: sum of squares of [log (I / (Q*N))]

I-survey indices Q-survey catchability N-abundance
(sub-indices for survey, age and year have been excluded)

- Survey's catchability at age: geometric mean with epoch consideration:
- Partial recruitment: arithmetic mean of Fs
- Fishing mortality of the last age: equals an average of fishing mortalities of the 3 younger ages using Heincke (1913) method.
- VPA: exact solution
- Plus-group not considered.

These three ADAPT formulations have different model requirements because, even though they use the same criterion to calculate F at the last age, this procedure is applied only to those cohorts without a parameter for survivors. The ADAPT-27 formulation does not constraint F at the last age in most cases, so it will be almost insensible to variations of the above sdPR parameter.

A Monte Carlo simulation was carried out with different combinations of the above parameters. Each set of test parameters is submitted to 1000 iterations. Each iteration starts with a new simulated case, which includes new fishing mortalities, abundances, catches and survey indices. Each simulated case was analyzed with the three ADAPT formulations in order to obtain more comparable results among them. Retrospective analysis was done moving back (peeling) one year every time up to 5 years.

Several indices were used to measure the consistency between results of a retrospective analysis. We call them *retrospective indices*. In general, these indices are based in the comparison of a defined result in the peeled analysis with the same result in the reference analysis. Main retrospective indices compare abundances at the same age-year but calculated in different peels: full time (peel_0), one year peeled (peel_1), two years peeled (peel_2), …, up to five years peeled (peel_5). Other indices compare SSB or exploitable biomass of the same year but calculated in different peels. Fifteen retrospective indices were considered; they are the combination of three data sources (1, 2, 3) with five statistics (ρ, σ, π, σSSB, σFB):

Data sources for *retrospective indices*:

1 – Compares quantities of the last year of peel_p vs same figures in peel_p-1
2 – Compares quantities of the last year of peel_p vs same figures in peel_0
3 – Compares quantities in all years of peel_p vs same figures in peel_0

Statistics:

indices ρ – mean relative difference in abundance at age

$$\rho 1 = \sum_{p=1,5} [\sum_{a=1,na} (N^{p}_{a,yp} / N^{p-1}_{a,yp} - 1) / na] / 5$$

na = number of ages considered
$N^{p}_{a,y}$ = abundance at age *a*, year *y*, of peel_p
yp = last year in peel_p

$$\rho 2 = \sum_{p=1,5} [\sum_{a=1,na} (N^{p}_{a,yp} / N^{0}_{a,yp} - 1) / na] / 5$$

$$\rho 3 = \sum_{p=1,5} (\sum_{y=1,yp} [\sum_{a=1,na} (N^{p}_{a,y} / N^{0}_{a,y} - 1) / na] / yp) / 5$$

indices σ – squared root of mean of squared relative differences in abundance at age

$$\sigma 1 = \sum_{p=1,5} \text{sqrt} [\sum_{a=1,na} (N^{p}_{a,yp} / N^{p-1}_{a,yp} - 1)^2 / na] / 5$$

$$\sigma 2 = \sum_{p=1,5} \text{sqrt} [\sum_{a=1,na} (N^{p}_{a,yp} / N^{0}_{a,yp} - 1)^2 / na] / 5$$

$$\sigma 3 = \sum_{p=1,5} (\sum_{y=1,yp} \text{sqrt} [\sum_{a=1,na} (N^{p}_{a,y} / N^{0}_{a,y} - 1)^2 / na] / yp) / 5$$

indices π – squared root mean of squared logarithmic relative abundances

$$\pi 1 = \sum_{p=1,5} \text{sqrt} [\sum_{a=1,na} (\ln (N^{p}_{a,yp} / N^{p-1}_{a,yp}))^2 / na] / 5$$

$$\pi 2 = \sum_{p=1,5} \text{sqrt} [\sum_{a=1,na} (\ln (N^{p}_{a,yp} / N^{0}_{a,yp}))^2 / na] / 5$$

$$\pi 3 = \sum_{p=1,5} (\sum_{y=1,yp} \text{sqrt} [\sum_{a=1,na} (\ln (N^{p}_{a,y} / N^{0}_{a,y}))^2 / na] / yp) / 5$$

indices σSSB – squared root of mean squared relative differences in SSB

$\sigma SSB1 = \text{sqrt}\ [\sum_{p=1,5} (SSB^{p}_{yp} / SSB^{p-1}_{yp} - 1)^2 / 5]$

SSB_{yp} = SSB in year *yp* of peel_p
yp = last year in peel_p

$\sigma SSB2 = \text{sqrt}\ [\sum_{p=1,5} (SSB^{p}_{a,yp} / SSB^{0}_{a,yp} - 1)^2 / 5]$

$\sigma SSB3 = \text{sqrt}\ [\sum_{p=1,5} \sum_{y=1,yp} (SSB^{p}_{a,y} / SSB^{0}_{a,y} - 1)^2 / yp / 5]$

indices σFB – squared root of mean squared relative differences in FB

$\sigma FB1 = \text{sqrt}\ [\sum_{p=1,5} (FB^{p}_{yp} / FB^{p-1}_{yp} - 1)^2 / 5]$

FB_{yp} = Exploitable Biomass in year *yp* of peel_p
yp = last year in peel_p

$\sigma FB2 = \text{sqrt}\ [\sum_{p=1,5} (FB^{p}_{a,yp} / FB^{0}_{a,yp} - 1)^2 / 5]$

$\sigma FB3 = \text{sqrt}\ [\sum_{p=1,5} \sum_{y=1,yp} (FB^{p}_{a,y} / FB^{0}_{a,y} - 1)^2 / yp / 5]$

Index ρ2 is similar to Mohn's rho (Mohn 1999), but divided by the number of ages and number of peels. Index ρ3 is Woods Hole rho (Mohn 1999). Indices π follow Parma (1993).

Indices ρ should have a mean equal zero if variability of input data had a symmetric effect on results, having both positive and negative values, so they should not be good candidates to identify the effect of random variability. Indices σ and π are always positive; disagreement in one age is never balanced with some opposite disagreement in other age, as it happens in ρ indices; they are metrics in mathematical terms; they measure the distance between two positions in a *na* dimensional space, where each dimension is relative abundance at each age; if they increased monotonously with the tested error they would be better candidates as indicators of the effect of that error. Indices σSSB and σFB may have same properties as indices σ, but their variability is expected to be smaller.

Indices based on data sources 2 and 3 compare data (abundances, SSB or FB) in peel_p with the same figure in full time series (peel_0), based on the idea that the full time series is the main reference for all others peels. Indices based on data source 1 compare data in peel_p with the same figure in peel_p-1, assuming that full time series has nothing special and the important is the change from one year analysis to the next, or from one peel to the next. Indices based on data sources 1 and 2, in opposition to those based on data source 3, only compare the last year of each peel based on the idea that last year already contains the signal of the whole series because of the convergence current character of retrospectives; a circumstance that should be particularly true with random effects on input data due to the lack of pattern. For the above reasons, indices based on data source 1 are *a priory* the best candidates to better indicate random inaccuracy in input data and disagreement in VPA model.

Indices based on SSB and Exploitable biomass (FB) estimates were considered because these two variables are among the most important assessment results. A bad retrospective in abundance at age could be tolerated if the retrospectives in SSB and FB are good.

Distributions of retrospective indices are noticeable skewed, with several, even rear, huge values. In order to get steady statistics, distribution's tails were reduced up to +/- 5 standard deviations. Discarded cases were at a minimum.

At this point, it is important to note that these retrospective indices do not compare VPA results with the "true" population, but between consecutive VPA results of the retrospective analysis. So, they tell nothing about accuracy of the results, but only indicate consistence of consecutive results. A low retrospective index doesn't mean an accuracy result; only a high retrospective index could indicate inconsistency in input data or VPA model assumptions. However, because this is a simulated exercise, it is possible to apply the same indices to measure the consistence between VPA results and the simulated population. Such indices may only compare one case: the full time series (peel_0) vs the simulated population, instead comparing 5 cases (peels) as retrospective indices do. However these new indices and the retrospective indices are in the same scale because all of them are mean relative values. These new indices may measure the distance between an ADAPT solution and the simulated population; they should be indicators of the capacity of the VPA method, the ADAPT's formulation in this case, to produce results that are close to or far from the simulated population; we call them *bias indices* and represent them with the same names used for retrospective indices plus an asterisk. Having only two sets to compare: simulated population (SP) and full results (FR), data sources were changed to:

<u>Data sources for *bias indices*</u>:

1 – Compares quantities of the last year of FR vs same figures in SP
2 – Compares quantities of the last 5 years of FR vs same figures in SP
3 – Compares quantities in all years of FR vs same figures in SP

Results

The fifteen retrospective indices were analyzed in relation to increases of each parameter: cvCA, cvSI and sdPR, which represent variability in catch at age, survey indices and partial recruitment respectively. The following figures illustrate the main facts that were systematically observed in results with multiple combinations of the three parameters.

The first group of analysis was done with the three parameters getting simultaneously one of the five test values: 0.0001, 0.1, 0.2, 0.5, and 1.0. Figure 1 shows the behaviour of the fifteen retrospective indices vs these inaccuracy levels. Each graph contains one index as calculated by the three ADAPT formulations. Each point represents a mean of 1000 different cases from the Monte Carlo simulation. It is observed that all indices coincide or are quite close to cero when the three parameters are negligible (0.0001). This is a primarily test on the goodness of the ADAPT solution and an unavoidable condition for the calculation routines in use. This figure also shows that ADAPT formulations 10 and 9 have quite similar behaviour, but the ADAPT-27 formulation has the lower values, which means it produces the more consistent retrospective.

Figure 2 shows the behaviour of the fifteen bias indices for the same group of analysis, i.e. the same cases as in Figure 1. Again, all indices coincide or are quite close to cero when the three parameters are equal 0.0001. It also shows that ADAPT formulations 10 and 9 have quite similar behaviour, but the ADAPT-27 formulation has the higher values, which means it produces the more biased results. These results make ADAPT-27 formulation inadvisable and exclude it from later consideration; even it remains represented in all the figures (blue lines).

A second group of analysis was done with only variation in the sdPR parameter, the one related to partial recruitment, and being cvCA = 0.2, a low value, and cvSI = 0.3, also a low value. Figure 3 shows the behaviour of the fifteen retrospective indices vs the sdPR parameter. Similar results are founded with other values of the cvCA and cvSI parameters. The behaviour of the indices based on the ADAPT-27 formulation is different to the other two formulations; it indicates that this formulation is insensitive to year to year variation in partial recruitment, as it was previously supposed. For other ADAPT formulations, indices increase when the sdPR parameter increases, but an asymptote seems to exist at higher sdPR values.

Figure 4 shows the behaviour of the fifteen bias indices for the same cases as in Figure 3. Most indices bases on the ADAPT-10 formulation are below the ADAPT-9 formulation, particularly indices σ, π, σSSB with data sources 1 and 2. It indicates that the ADAPT-10 formulation produces the more accurate results in variable PR circumstances. The corresponding retrospective indices also produce the smaller values, as Figure 3 shows.

Another group of analysis was done with only variation in the cvCA parameter, the one related to catch at age accuracy, and being sdPR = 0.2 and cvSI = 0.2. Figure 5 shows the behaviour of the fifteen retrospective indices vs the cvCA parameter. The behaviour of the indices based on the ADAPT-10 and ADAPT-9 formulations are similar and, in general, they increase monotonously with cvCA, indicating a high sensibility of these indices to accuracy in catch at age input data. Figure 6 shows that the behaviour of the bias indices is similar to the retrospective equivalent.

A fourth group of analysis was done with only variation in the cvSI parameter, the one related to survey indices precision, and being sdPR = 0.2 and cvCA = 0.2. Figure 7 shows the behaviour of the fifteen retrospective indices vs the cvSI parameter. Ones again the behaviour of the indices based on the ADAPT-10 and ADAPT-9 formulations is similar, being the first preferable. Results indicate that these indices are quite insensible to variations in the cvSI parameter in this range of values. Figure 8 indicates that bias indices show the same insensibility.

In order to make comparison between pair of results, an additional fifth group of analysis was done, similar to the second one, but with more extreme conditions: sdPR varied in the 5 usual levels, being cvCA = 0.5 and cvSI = 0.8. These five groups of analysis totalizing 25,000 cases (5 groups of analysis, 5 levels, 1000 iterations) contain a wide range of situations and will be used as a representative sample of possibilities. Even each case was analyzed with the three ADAPT formulations, only results of the ADAPT-10 has been considered hereafter. Figure 9 shows the relationship between bias indices for SSB and Exploitable Biomass (FB). These indices are expected to be the best indicators of the overall bias of the results because of the relevance of SSB and FB. However, the relationship among them is poor when only the last year is considered as data source. The agreement is the highest when 15 years are considered as data source and, consequently, we will use σSSB3* and σFB3* as the best indicators of overall bias.

Discussion

Retrospective results are usually considered to identify a pattern: it seems clear that any strong pattern may indicate problem in input data or VPA assumptions. However the occurrence of a pattern is not the only issue to consider; the retrospective is not an aesthetic problem of the results. From a general point of view, it would be good that retrospective could be linked to the goodness of fit of the analysis done. That was the reason why bias indices were introduced, because they are direct measurement of the agreement between results and real stock's data or the bias of a VPA solution. Bias indices cannot be calculated in real world; however, if a close relationship could be proved between a retrospective and a bias index, the first could be used to estimate the second. Including this relationship, a good retrospective index, to be useful as indicator of problems in the input data or in the whole analysis, should have the following properties:

1. It should be zero when input data inaccuracy and VPA model disagreement are both null.
2. It should be a positive magnitude.
3. It should increase when input data inaccuracy increased and when VPA model disagreement also increased, preferably with a linear relationship.
4. It should have the lowest dispersion for a given error of input data.
5. It should be correlated with some bias index.

The first condition was satisfied by all tested indices. It was checked systematically as a way to verify the routines.

The second condition is justified in this analysis because the effects of random variability are only tested, so the sign of output is irrelevant. This condition eliminates ρ indices from further consideration, even they remain a useful tool to detect retrospective patterns.

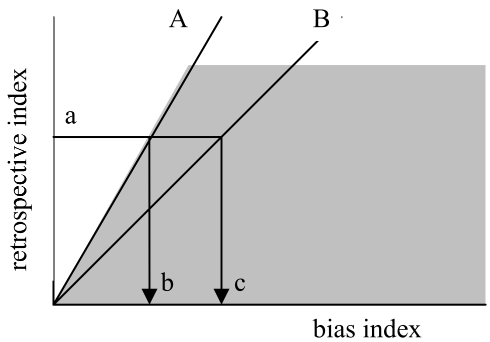
The third condition was satisfied for remaining indices: σ, π, σSSB and σFB. In general, linearity was mostly observed in relation to inaccuracy in catch at age (cvCA parameter); it was weak with inaccuracy in survey indices (cvSI parameter), and it seems to be asymptotic with variability in partial recruitment (sdPR parameter). These conclusions are obviously linked to the characteristics of the problem we analyzed, however, taking into account that our case is a common case, it is expected that a similar or close behaviour occurs in other cases.

The fourth and fifth conditions are linked: a good retrospective index should allow testing the hypothesis that some bias index is below or higher than certain level. A preliminary approach indicates that an unbiased solution cannot produce a high retrospective index; however a biased solution is compatible with a low retrospective. In other words: a low retrospective is not indicative of an unbiased solution and only a high retrospective indicates bias in some extent. Bias indices can only be calculated in a simulation like the one here presented. In a real case only retrospective indices can be calculated.

Best candidates for a retrospective-bias indices pair might be σSSBn-σSSBn* and σFBn-σFBn* because they consider most important VPA results. If they measure the same characteristic of results, they may be well correlated. Figure 9 contains graphs with the scattering plots for pairs of these indices. It shows that the σSSB1-σSSB1* relationship is too weak, but the σSSB3-σSSB3* one is fairy good. The σSSB1 and σFB1 consider 5 peels, σSSB2* and σFB2* consider last five years of data, and σSSB3* and σFB3* consider last fifteen. The more the number of data the best they catch the bias.

The relationship between retrospective and bias indices seems to behave as it is illustrated in the following scheme, where the grey area represents allowed occurrence. Line A represents the left most limits of possible values (99^{th} percentile) and line B represents the most probable values (median). There is not right hand limits. The white space between the ordinates axis and the A line is an exclusion area. A retrospective value like "*a*" would mean a bias greater than *b* and most probably so high as *c*. If line A was quite close to the ordinate axis such retrospective index would be useless to make any judgment on the bias of the results. It would be desirable that line A be closer as possible to line B in order to improve prediction capabilities.

Based on the above scheme a search was done for those pairs of indices (one *retrospective* and one *bias*) that produce the less disperse plots; the 99^{th} percentile (line A) and the median (line B) were calculated. The quotient of these two magnitudes (99^{th} percentile and median) is a measure of the 99^{th} percentile in units of the median and is higher than 1;



the lower value the closer the 99th percentile is to the median, which implies less disperse plots. The next table shows results of the survey on lower quotients. It can be observed that best quotients are found among σ1, π3 and σ2*, σ3* and π3*, or σSSB3* and σFB3*, among those indices related to SSB and FB. Figures 10 and 11 shows behaviour of these pairs:

indices' pair		slope of 99th percentile	slope of median	quotient (99th percentile / median)
σ1	π3*	1.30	0.63	2.06
σ1	σ3*	1.48	0.64	2.31
π3	π3*	1.16	0.46	2.50
σ1	σ2*	1.88	0.72	2.61
σ1	σSSB3*	2.17	0.64	3.38
π3	σSSB3*	1.73	0.46	3.73
σ1	σFB3*	6.96	1.82	3.83
σFB3	σSSB3*	0.81	0.20	3.98
π3	σFB3*	5.30	1.30	4.08
σFB1	σFB3*	2.38	0.58	4.08
σSSB1	σSSB3*	2.16	0.53	4.09

The σ1-π3* pair presents the highest quotient, but the interpretation of the π3* bias index is not simple (π3* is the squared root of the 15 years mean squared logarithmic relative abundance). However σFB3* or σSSB3* (squared root of the 15 years mean squared relative differences in SSB or FB) are relative SSB or FB mean inconsistency, which is easier to interpret; e.g. σSSB3* = 0.2 roughly means that a SSB estimate is 20% higher or lower that the real figure as a mean. This easy way to interpret makes σSSB3* preferable to other bias indices.

Interpretation of the σ1-σSSB3* pair relationship, based on the analysis already done on the ADAPT-10 formulation and highlighted in the above table, would be summarized in the beside table: for each measurable σ1 value exists a wide range of possible σSSB3* values, with a lower limit at the level of the 99th percentile and a most probable value related with the median. Let it be 0.2 an acceptance upper limit for σSSB3*, this table indicates that any σ1 value above 0.43 is unacceptable, but values higher than 0.13 should be considered insecure. Note that this method could be used to reject a VPA analyses when retrospective indices are high, but it cannot validate any results when indices are low.

	σSSB3*	
σ1	(slope 2.17) 99th perctl	(slope 0.64) median
0.1	0.05	0.16
0.13	0.06	0.20
0.2	0.09	0.31
0.3	0.14	0.47
0.4	0.18	0.63
0.43	0.20	0.67
0.5	0.23	0.78
0.6	0.28	0.94
0.7	0.32	1.09
0.8	0.37	1.25

Note that the above method could be used to reject a VPA analyses when retrospective indices are high, but it cannot validate any results when indices are low.

The question arise of how much portable are these conclusions to other VPA methods or for stocks with different characteristics. Even the issue is not solved, some conclusion could be generalized:

- As it was initially envisaged, any level of bias permits a low retrospective, the lowest retrospective index does not imply the lowest bias, so low retrospective is not a guarantee of goodness of fit.

- Based on the same input data, retrospective indices are dependent of the VPA method in use. They should have a different meaning because of their own relationship between retrospective indices and accuracy of results. An example is the ADAPT-27 formulation, where every cohort (with enough survey data to compare) has its own parameter (the survivors), which produces solutions that are very far away of the simulated population, even its retrospective indices are better than the other two formulations. The ADAPT-10 and ADAPT-9 formulations produce retrospective indices with similar coefficients of variation, however the first one was preferred for complete analysis because, in general, it shows smaller retrospective and bias indices, even differences are low.

- Each VPA method in use introduces an additive component to the retrospective and bias indices. Inaccuracies in catch at age and survey indices have their own contributions to bias of results, and they cannot be hidden by any VPA method. Based on the same input data, expected bias of any solution cannot be smaller that some level due to the contribution of inaccuracy in catch at age and survey indices. Different VPA formulations can modify the component due to the incompetence of the VPA formulation, but they cannot cancel the bias due to input data. Any very low retrospective index is misleading.

- Described retrospective indices do not consider year to year variability, but only make comparison among quantities of the same year, so they can be applied even there are yearly tendencies. A retrospective having a pattern has at least an additional situation besides the random retrospective due to inaccuracy of input data, but it does not invalidate these retrospective indices.

- Retrospective indices can be used to test the goodness of fit in a VPA analyses.

Acknowledgements

We thank to Jean-Claude Mahé, from IFREMER, France, for his valuable comments.

This study was supported by the European Commission (Program for the Collection of Data in the Fisheries Sector).

References

Cerviño, S. and A. Vázquez – 2004. A Survey-based Assessment of Cod in Division 3M. *NAFO SCR Doc.* 04/53.

Gavaris, S.– 1988. An adaptive framework for the estimation of population size. *Can. Atl. Fish. Adv. Comm. (CAFSAC) Res. Doc.* 88/29.

Heincke, F.– 1913. Investigations on the plaice. General report I. The plaice fishery and protective regulations. Part. I. *Rapp. P.-v. Réun. CIEM*, 17A:1-153 + Annexes

Legault C.M. (Chair) – 2009. Report of the Retrospective Working Group, January 14-16, 2008, Woods Hole, Massachusetts. *US Dept Commer, Northeast Fish Sci Cent Ref Doc.* 09-01; 30 pp.

Parma, A.M.– 1993. Retrospective cartch-at-age analysis of Pacific halibut: Implications on assessment of harvesting policies. *Proc.Int. Symp. On Management Strategies for Exploited Fish Populations*: 247-265.

Mohn, R.– 1999. The Retrospective Problem in Sequential Population Analysis: An Investigation Using Cod Fishery and Simulated Data, *ICES Journal of Marine Science*, 56, 473-488.

Vázquez, A., R. Piñeiro and M. Mandado – 2009. Inquiry into ADAPT's performance. *NAFO SCR Doc.* 09/49

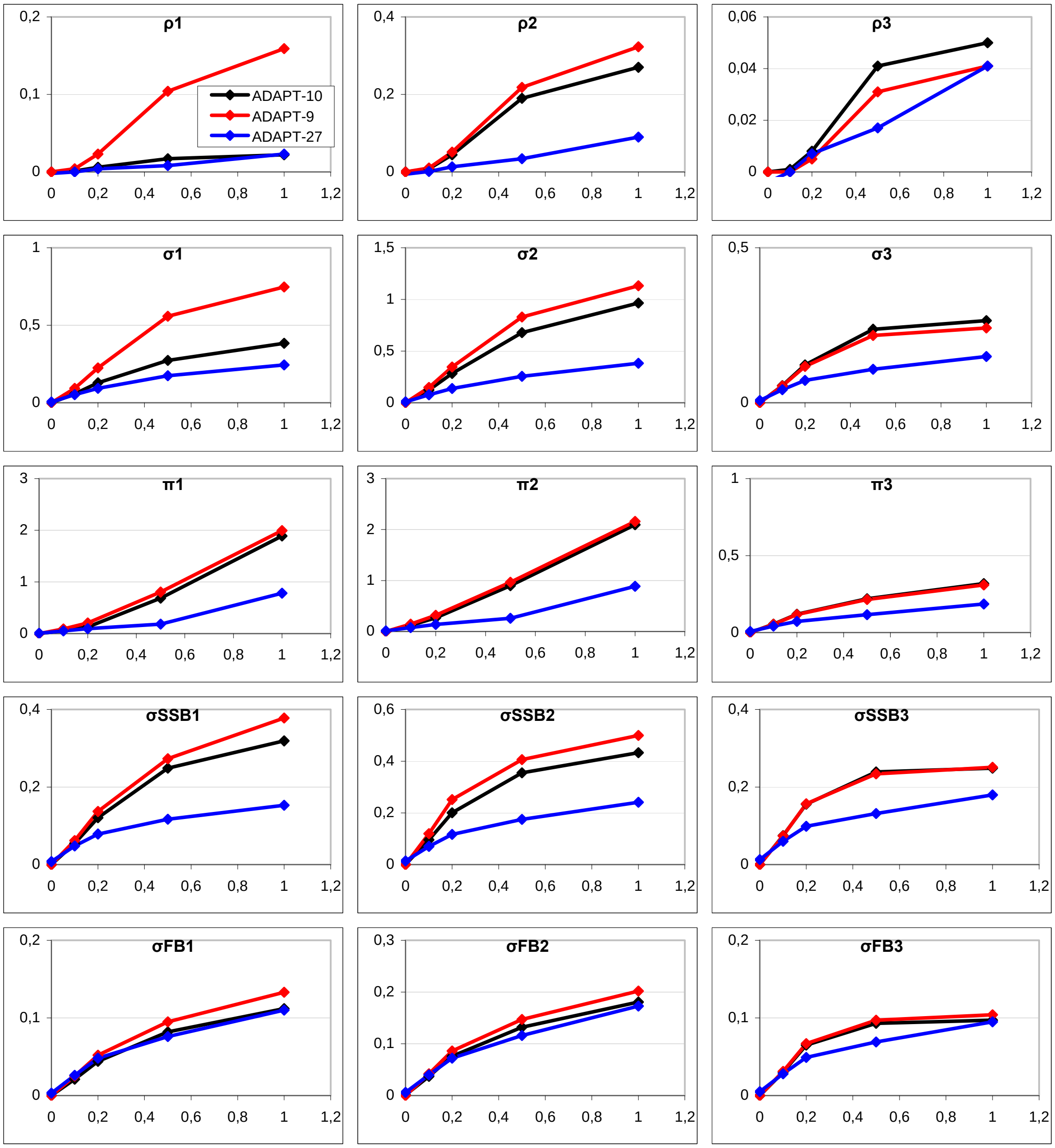


FIGURE 1 – Retrospective indices vs cvCA, cvSI and sdPR parameters, all of them equal to the common level indicated in the abscissa axis (0.0001, 0.1, 0.2, 0.5 and 1.0). Each graph corresponds to one retrospective index: ρ, σ, π, σSSB and σFB, with one data source: 1, 2 and 3. Each graph contains 3 lines that correspond to the three ADAPT formulations considered: the one with 10 parameters in black, with 9 parameters in red and with 27 parameters in blue. Each point is an average over 1000 runs.

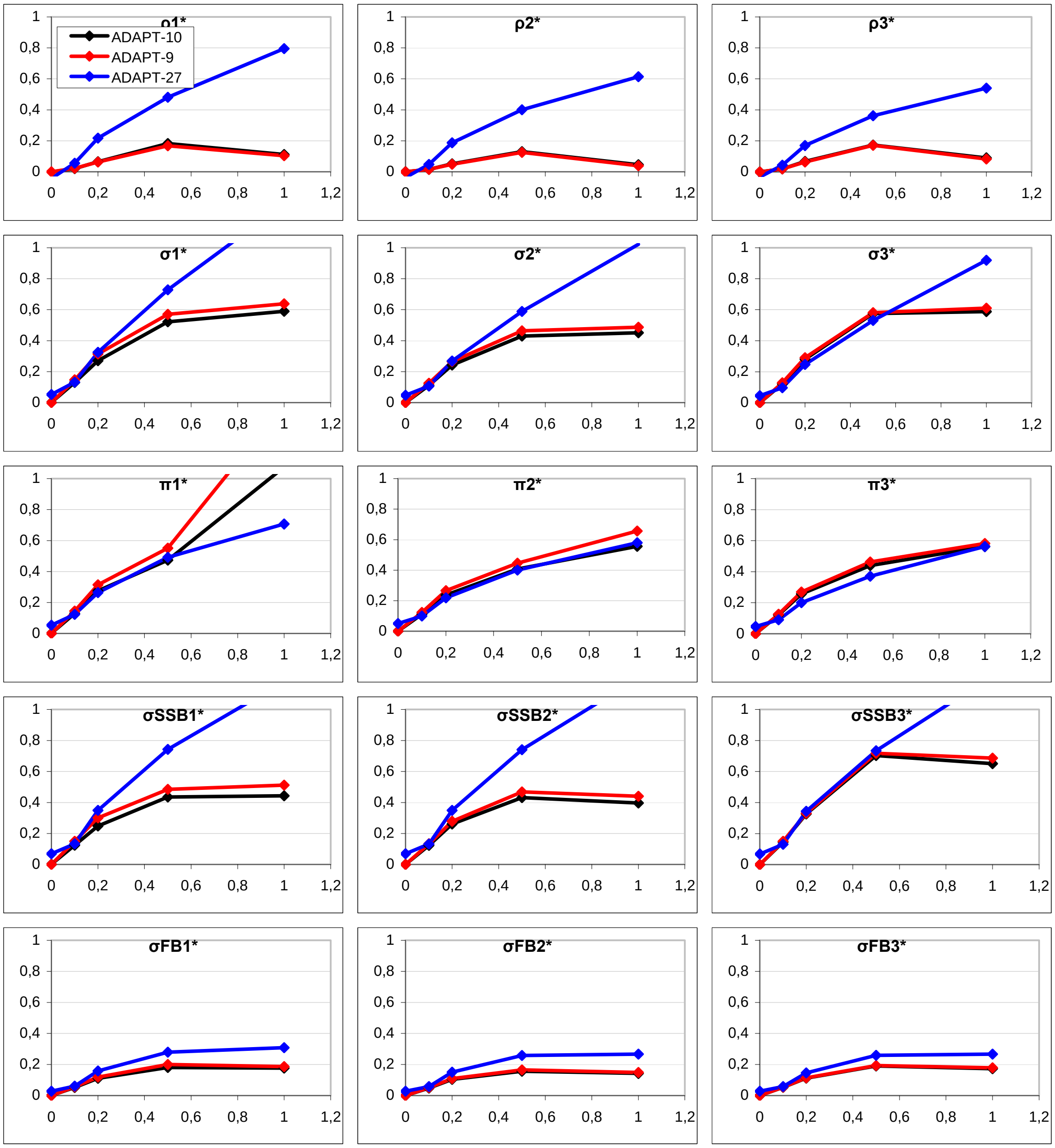


FIGURE 2 – Bias indices that correspond to the same cases as in Figure 1.

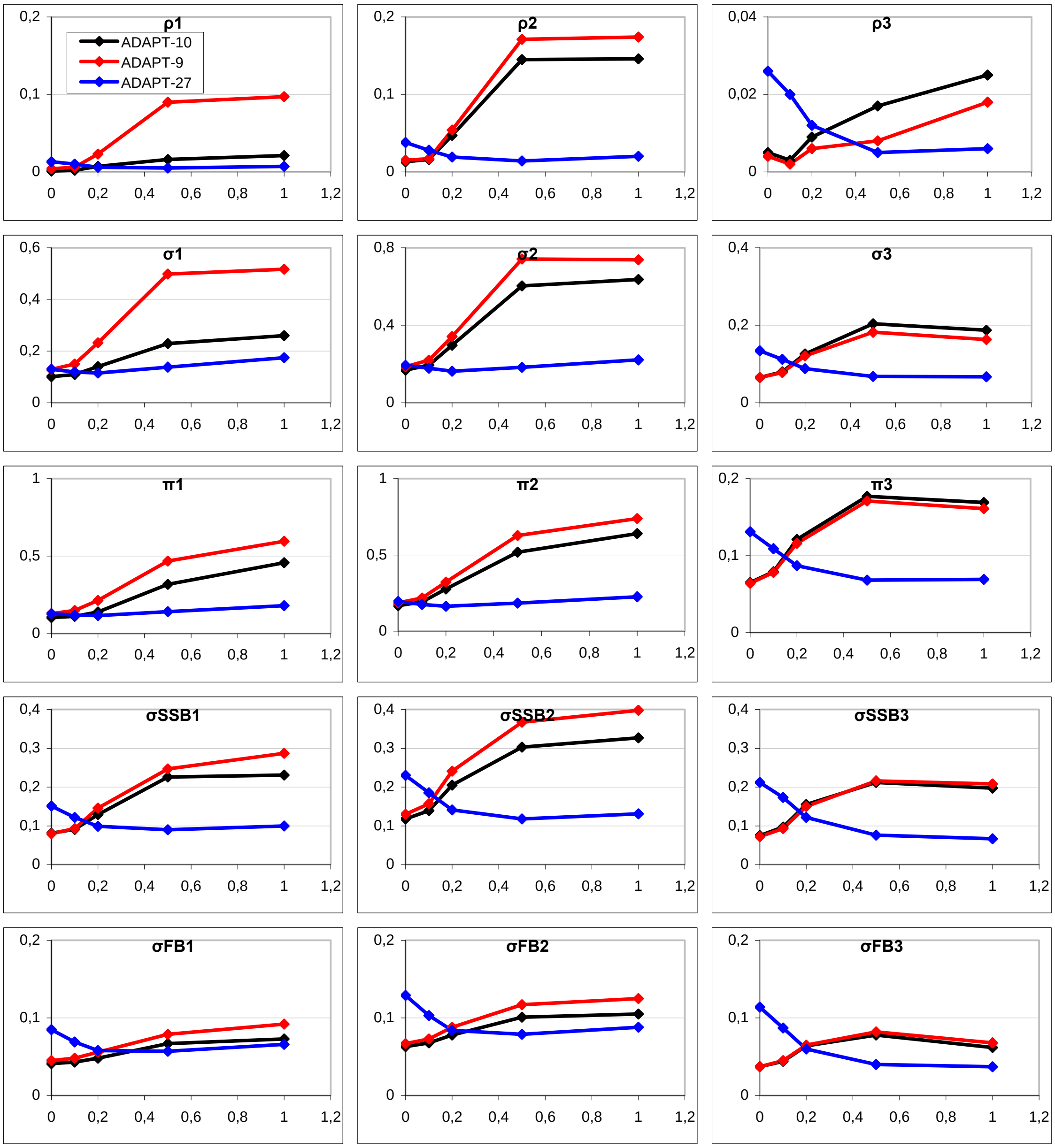


FIGURE 3 – Retrospective indices vs sdPR parameter (0.0001, 0.1, 0.2, 0.5 and 1.0). cvCA parameter equal 0.2 and cvSI parameter equal 0.3 in all cases. Other details are as in Figure 1.

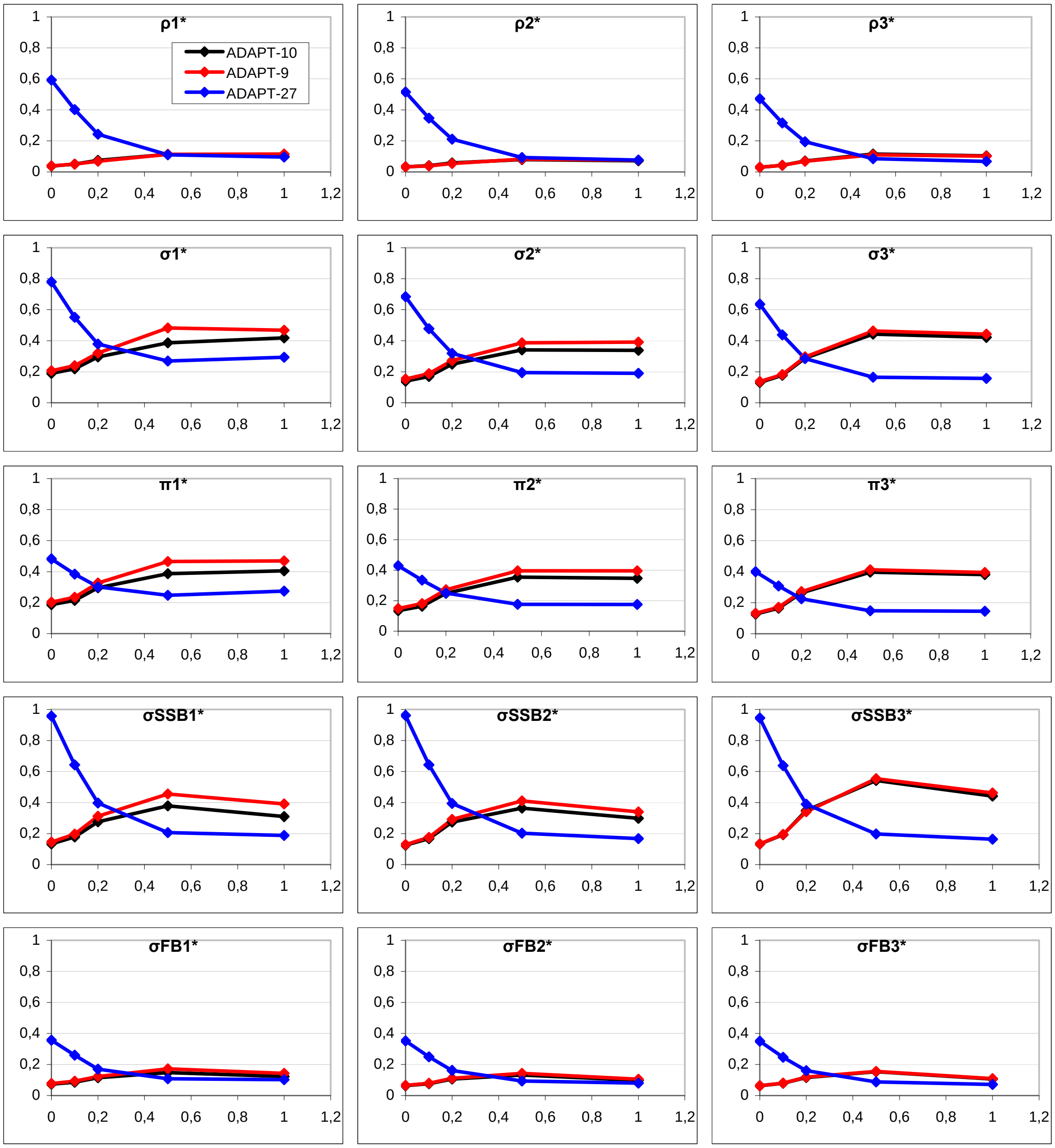


FIGURE 4 – Bias indices that correspond to the same cases as in Figure 3.

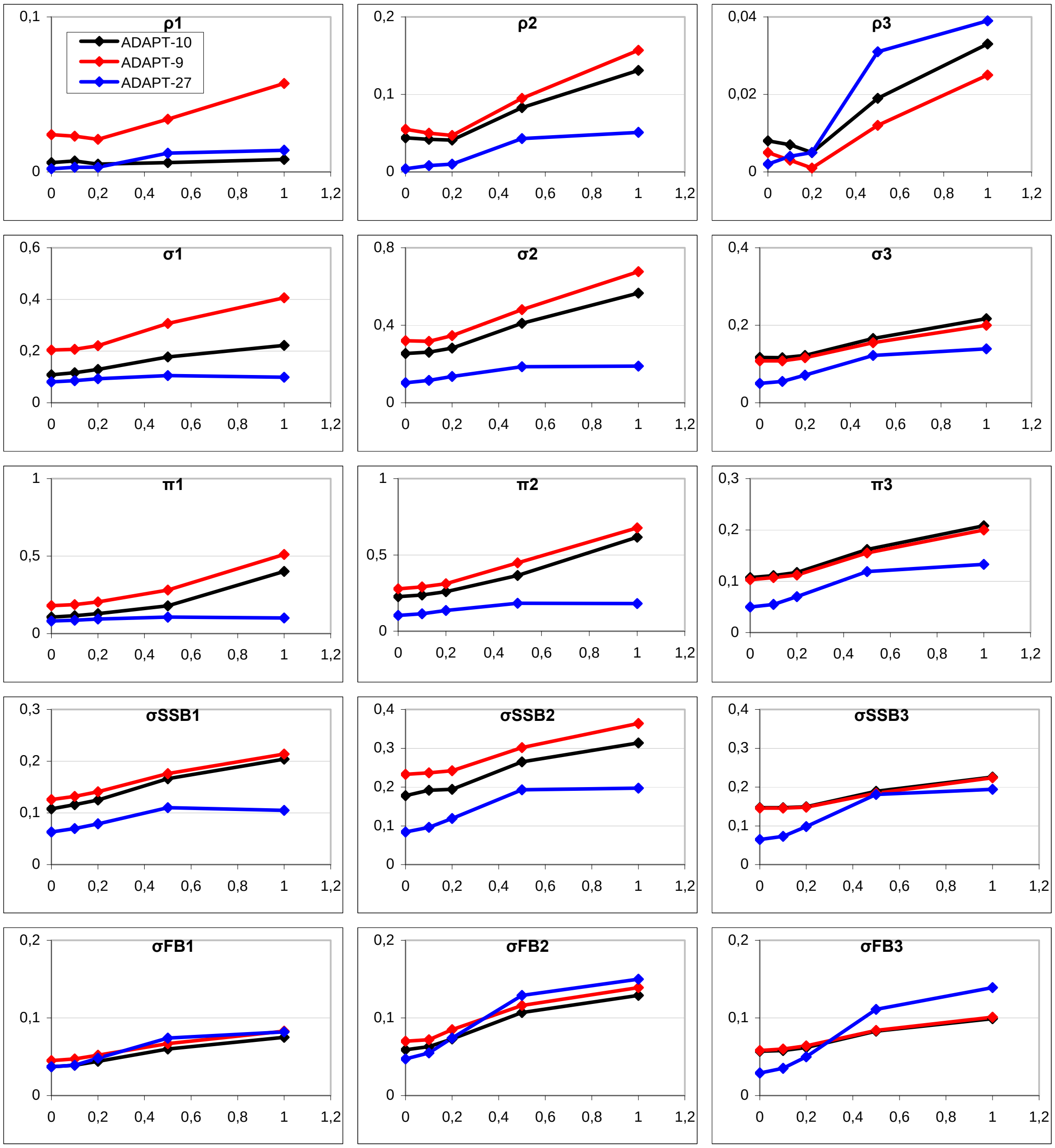


FIGURE 5 – Retrospective indices vs cvCA parameter (0.0001, 0.1, 0.2, 0.5 and 1.0). cvSI parameter equal 0.2 and sdPR parameter equal 0.2 in all cases. Other details are as in Figure 1.

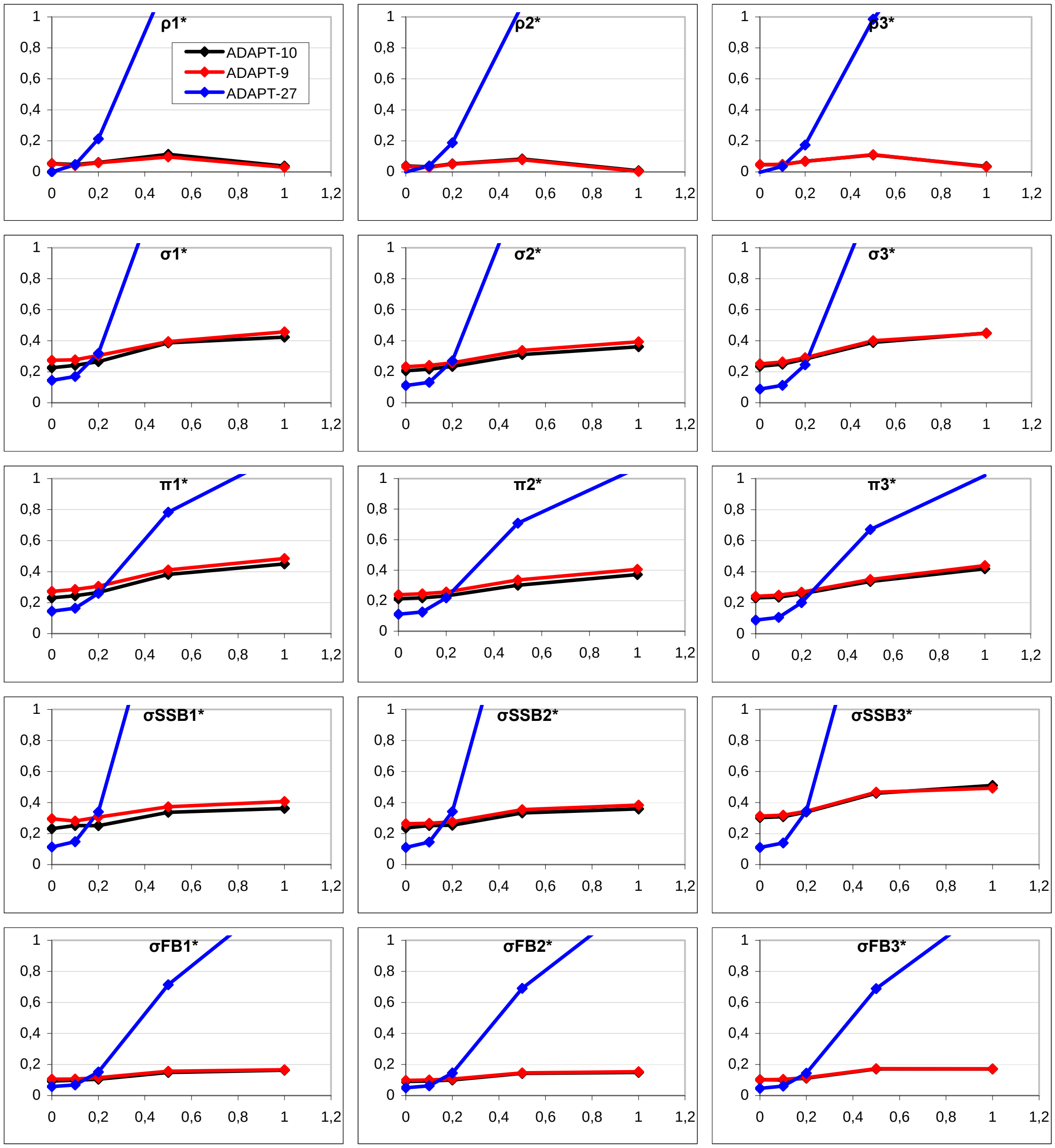


FIGURE 6 – Bias indices that correspond to the same cases as in Figure 5.

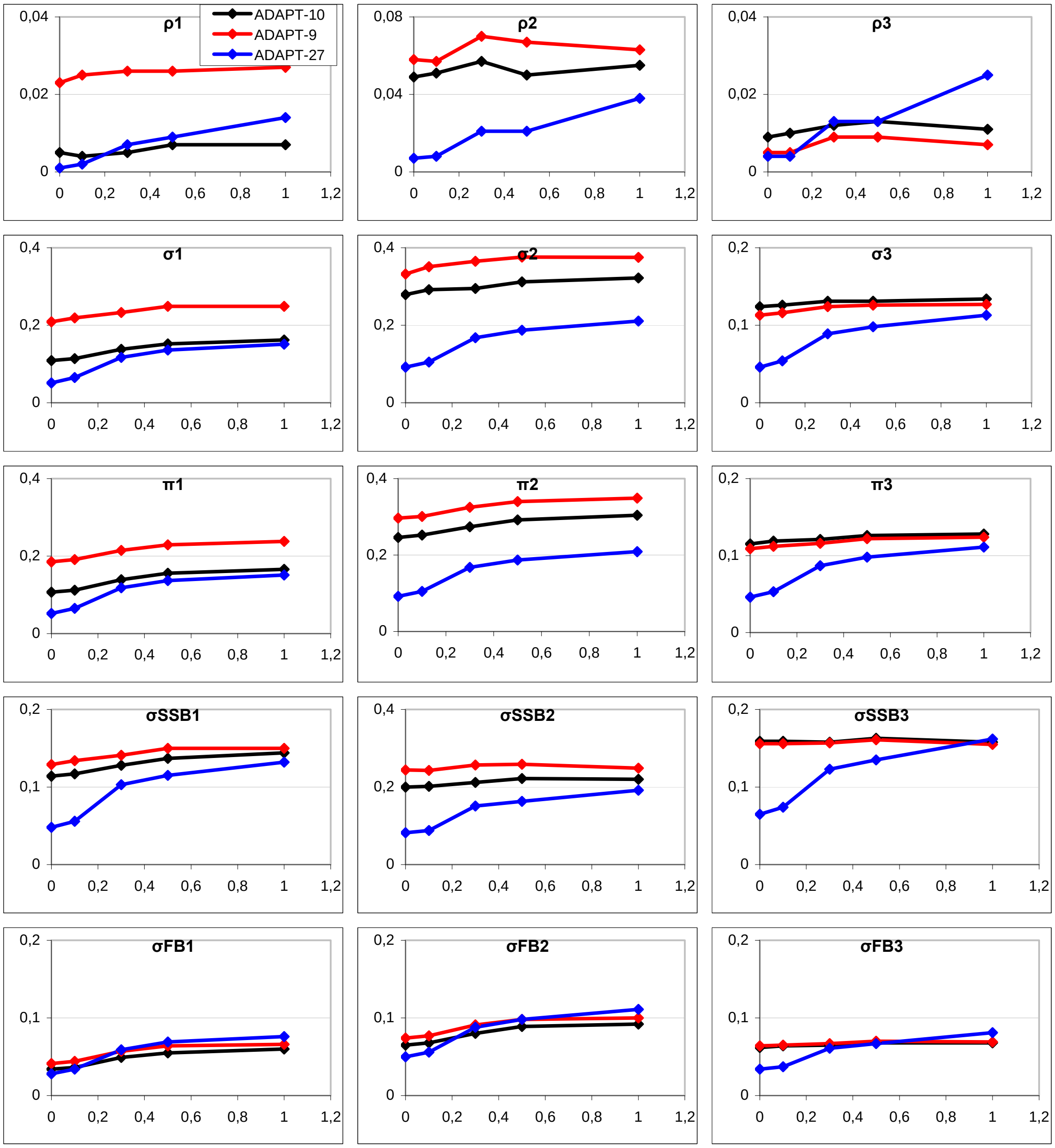


FIGURE 7 – Retrospective indices vs cvSI parameter (0.0001, 0.1, 0.2, 0.5 and 1.0). cvCA parameter equal 0.2 and sdPR parameter equal 0.2 in all cases. Other details are as in Figure 1.

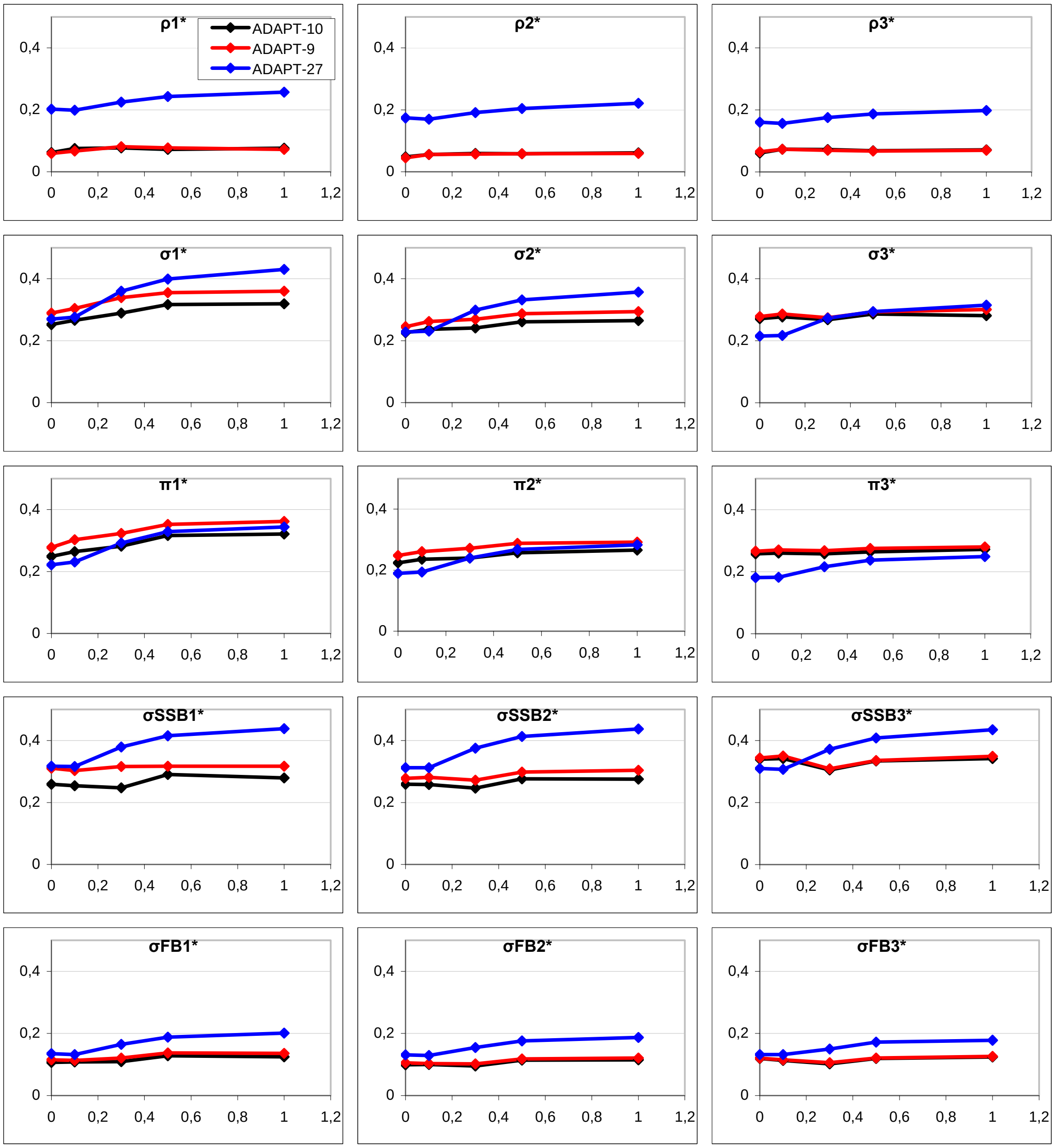


FIGURE 8 – Bias indices that correspond to the same cases as in Figure 7.

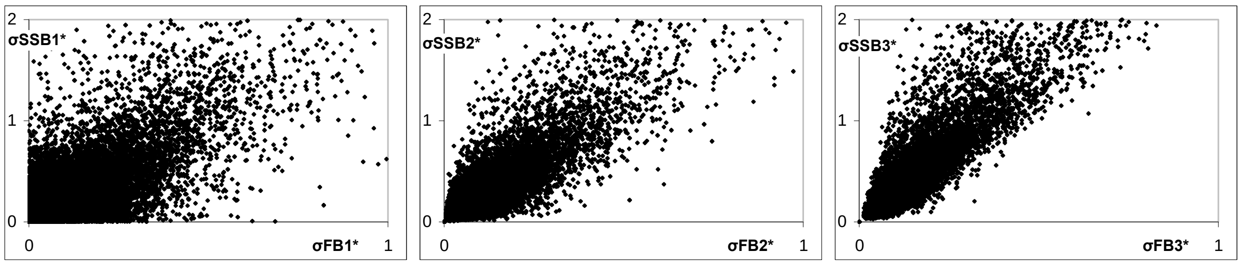


FIGURE 9 – Relationship between three pairs of bias indices: σSSB1*-σFB1* (SSB last year vs Exploitable Biomass last year), σSSB2*-σFB2* (five years), σSSB3*-σFB3* (fifteen years).

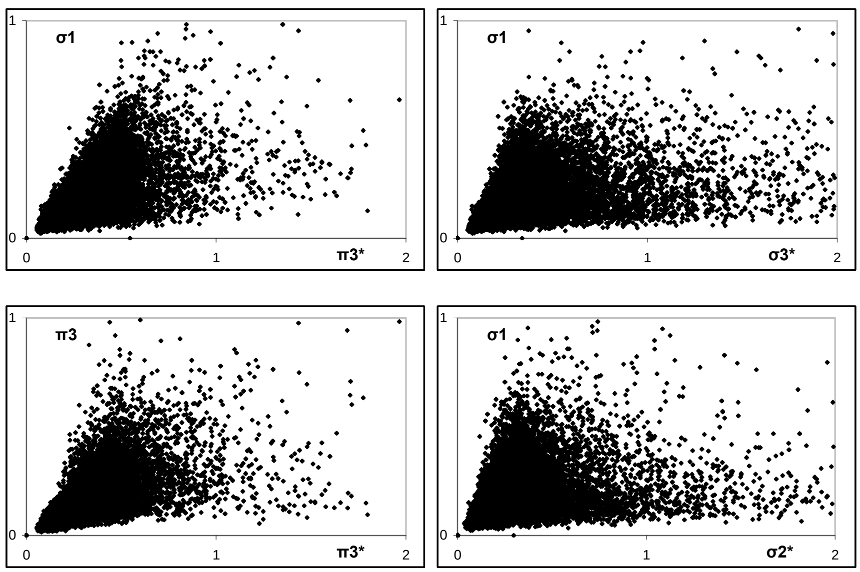


FIGURE 10 – Relationship between those pairs of retrospective-bias indices with the lowest ratio between 99th percentile and median of the slope (see text for details).

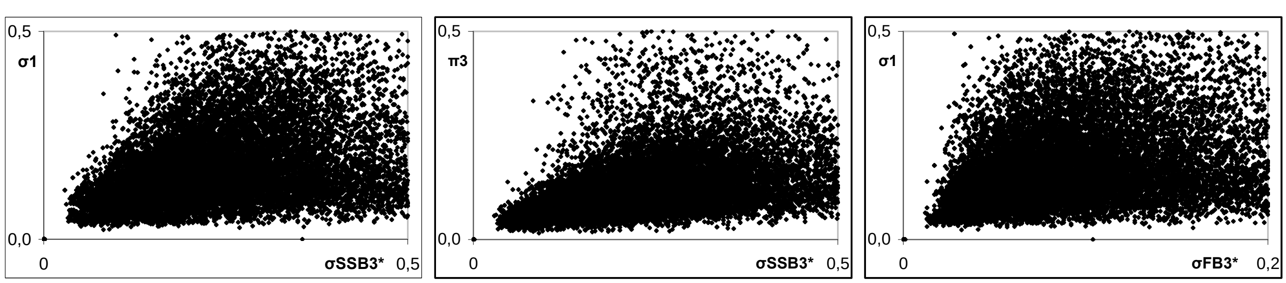


FIGURE 11 – Relationship between pairs of retrospective-bias indices involving SSB and Exploitable biomass (FB).