Southwest Pacific blue shark stock assessment
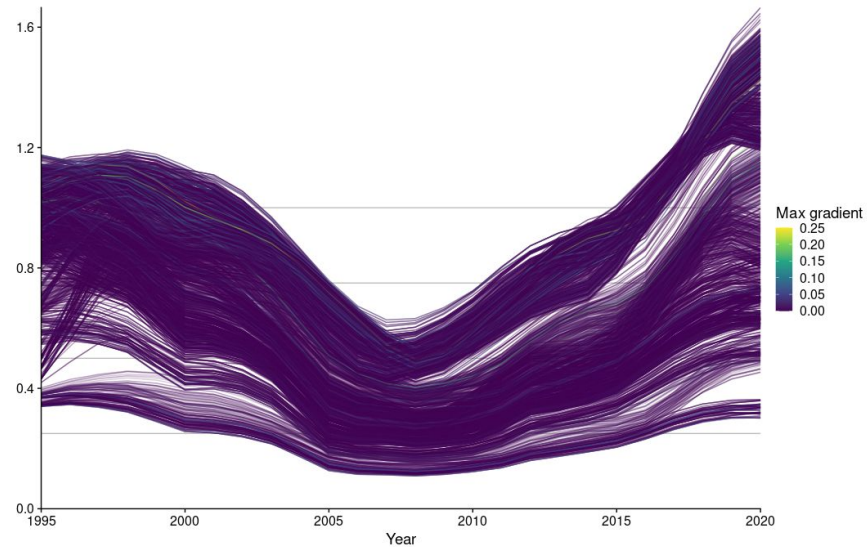
Philipp Neubauer, Kath Large, Stephen Brouwer

Agenda

1. Convergence
2. Key diagnostics for diagnostic case
3. Trimming the grid

1. Convergence

- All 3888 model have positive definite Hessian
- Nearly all models have small gradient
  - Some models have higher gradient, but don't have appreciably different trends:

1. Convergence

- Jitter applied to base-case (frac=0.2, N=10) suggests global convergence.

```
> jit.likes <- r4ss::SS_RunJitter('base', model = '../ss_linux',Njitter = 10)
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 1 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 2 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 3 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 4 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 5 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 6 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 7 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 8 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 9 = 778.779
arguments 'show.output.on.console', 'minimized' and 'invisible' are for Windows only
Likelihood for jitter 10 = 778.779
Table of likelihood values:
likesaved
778.779
    10
```
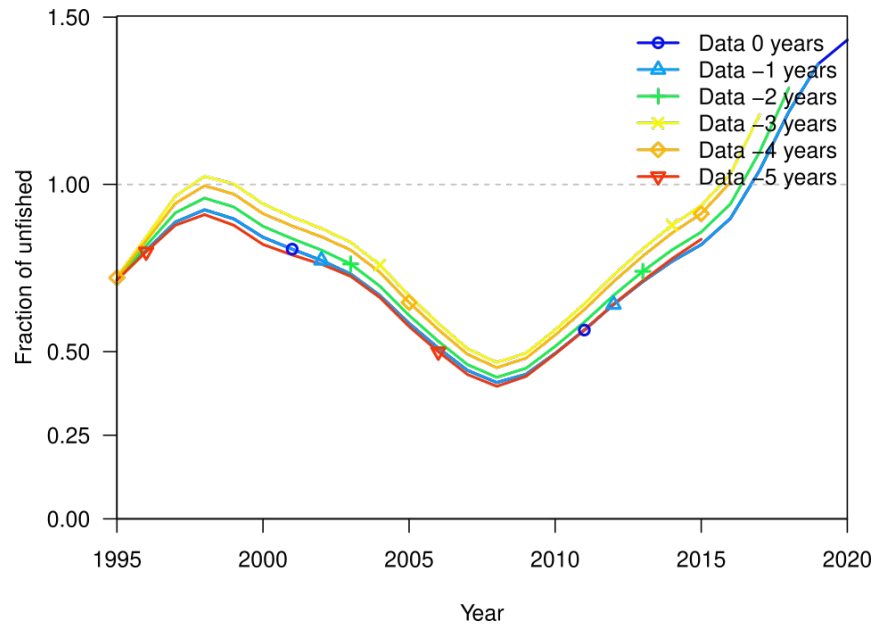
2. Diagnostics - diagnostic case
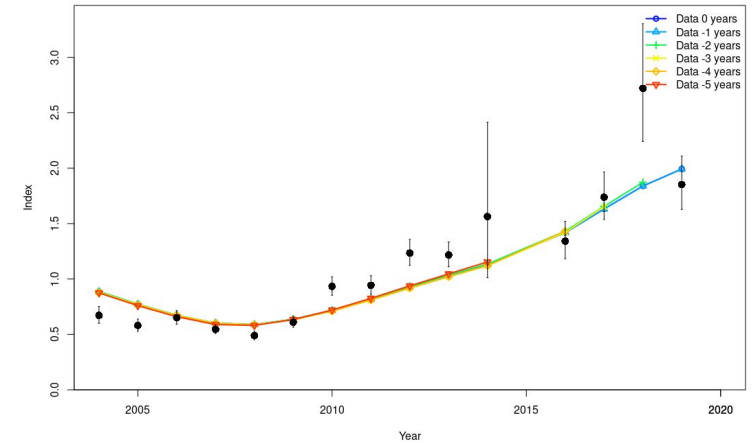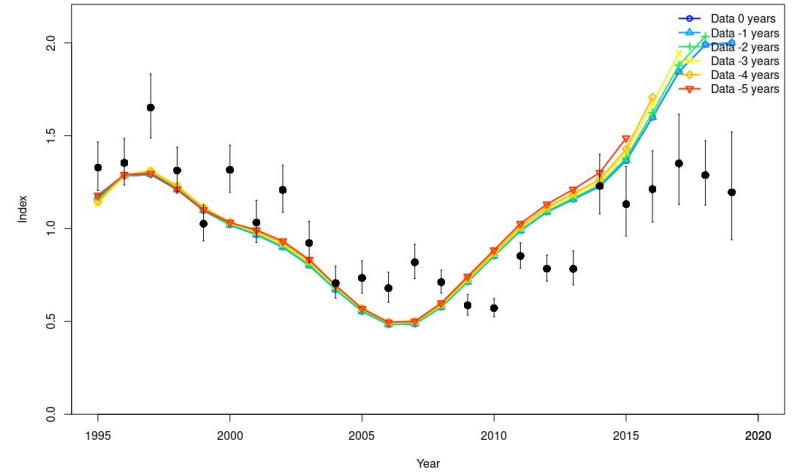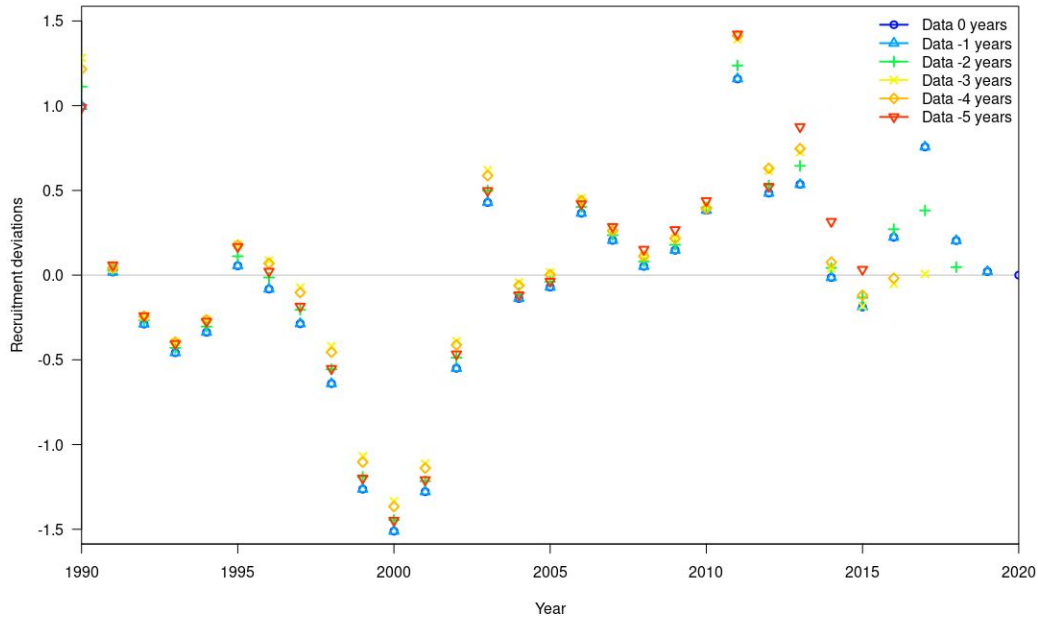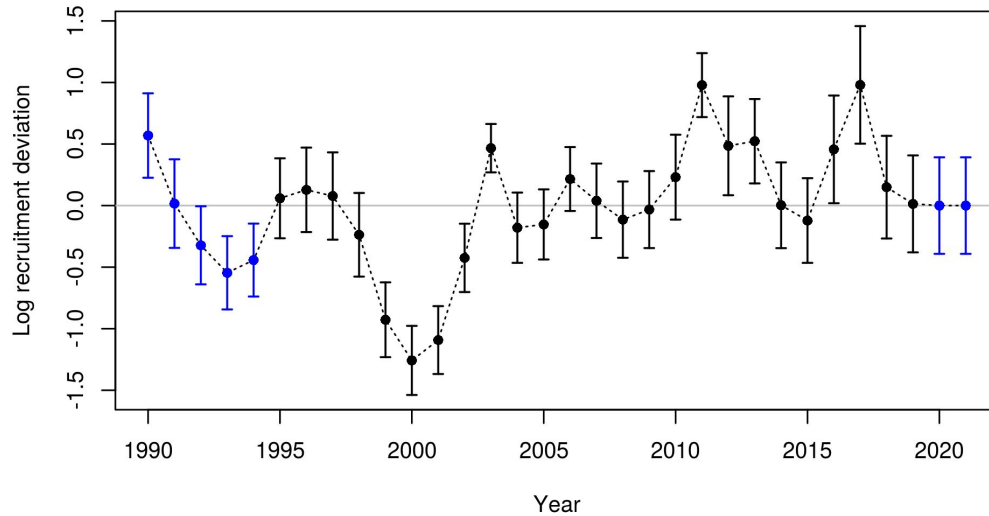
- Retrospectives applied to diagnostic case

2. Diagnostics - diagnostic case

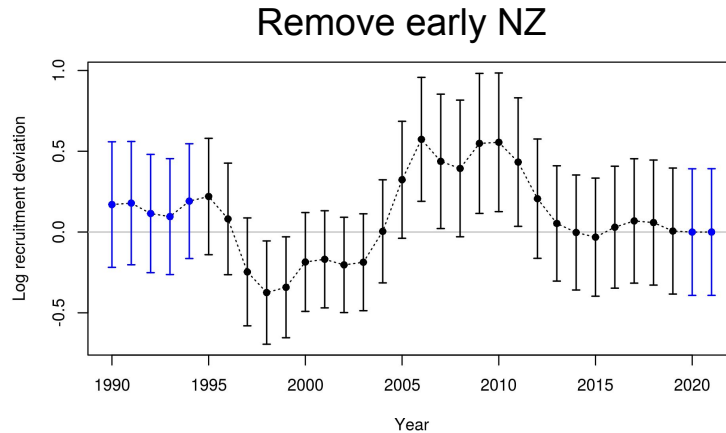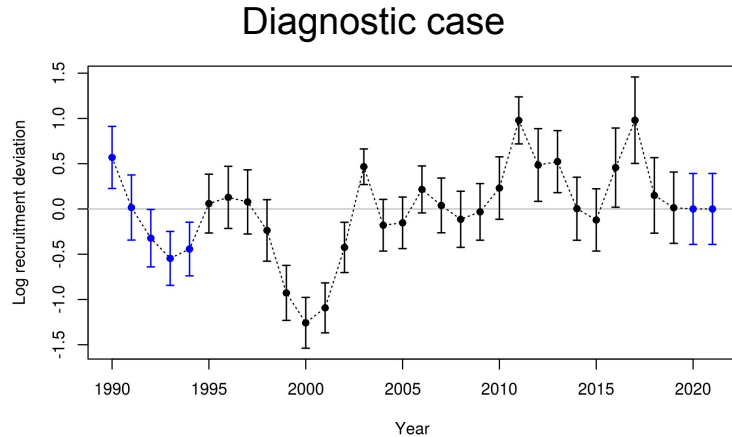● Retrospectives applied to diagnostic case: No consistent bias

2. Diagnostics - diagnostic case

- Retrospectives applied to diagnostic case: No consistent bias
- Recr. devs: pattern of low recruitment in 1990s

2. Diagnostics - diagnostic case

- Retrospectives applied to diagnostic case: No consistent bias
- Recr. devs: pattern of low recruitment in 1990s;
- Need for strong recruitment pattern disappears for runs with less extreme trends (i.e., runs removing EU or early NZ CPUE)
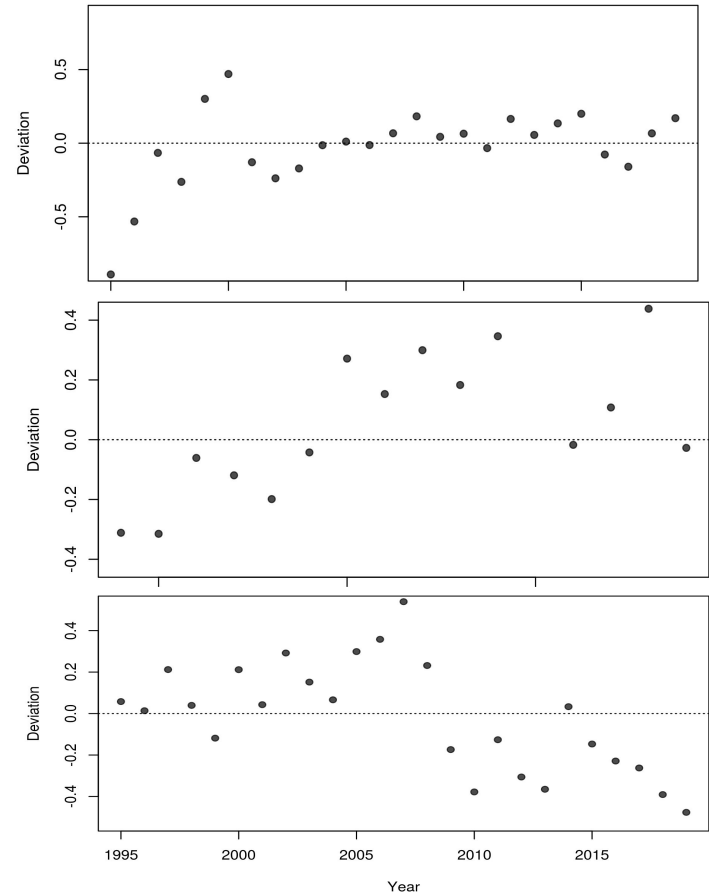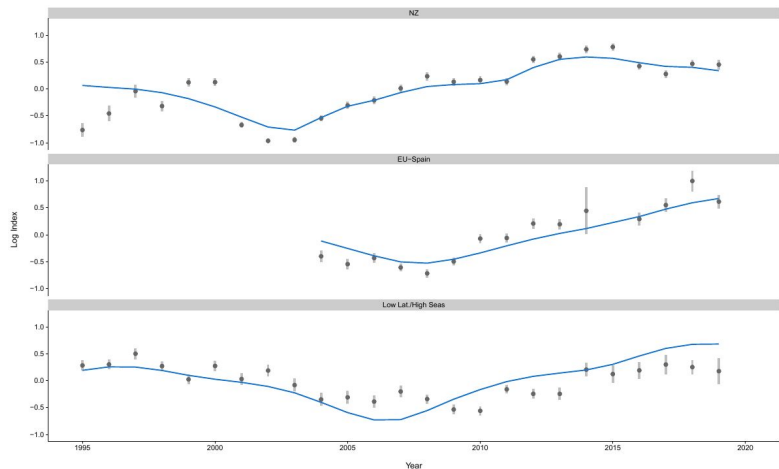

Diagnostic case

Remove early NZ

2. Diagnostics - diagnostic case

- Retrospectives applied to diagnostic case: No consistent bias
- Recr. devs: pattern of low recruitment in 1990s;
- Need for strong recruitment pattern disappears when removing early NZ
- Residuals look...not great. But not surprising either - cannot fit slightly different trends at once.
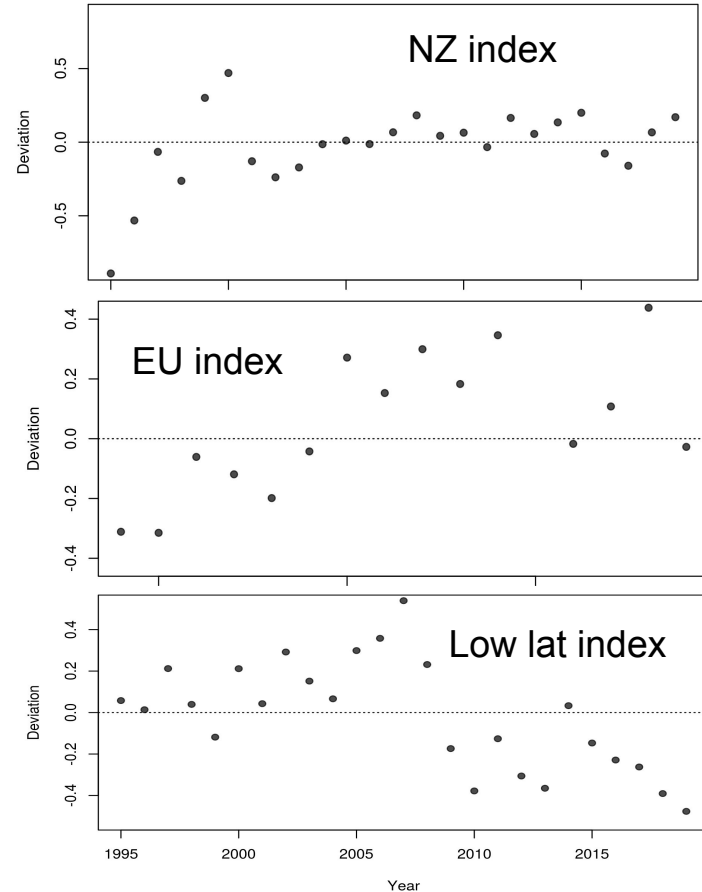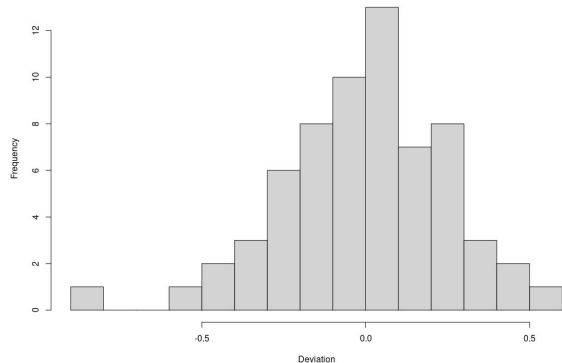
2. Diagnostics - diagnostic case

- Retrospectives applied to diagnostic case: No consistent bias
- Recr. devs: pattern of low recruitment in 1990s;
- Need for strong recruitment pattern disappears when removing early NZ
- Residuals look...not great. But not surprising either - cannot fit slightly different trends at once.
- Aggregate diagnostics not too far from normal?

# 2. Diagnostics - diagnostic case

- Retrospectives applied to diagnostic case: No consistent bias
- Recr. devs: pattern of low recruitment in 1990s;
- Need for strong recruitment pattern disappears when removing early NZ
- Residuals look...not great. But not surprising either - cannot fit slightly different trends at once.
- Can do better if we remove some conflicting data (e.g., no-EU)
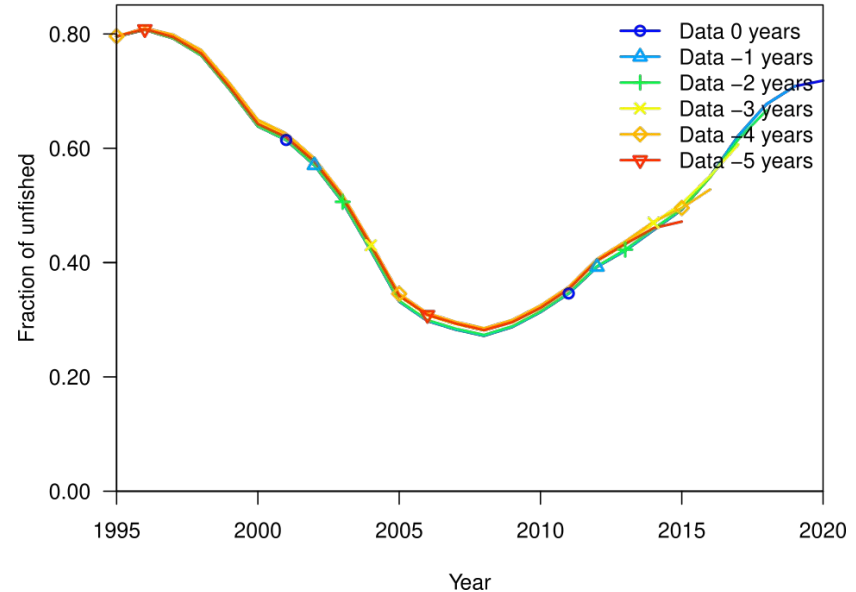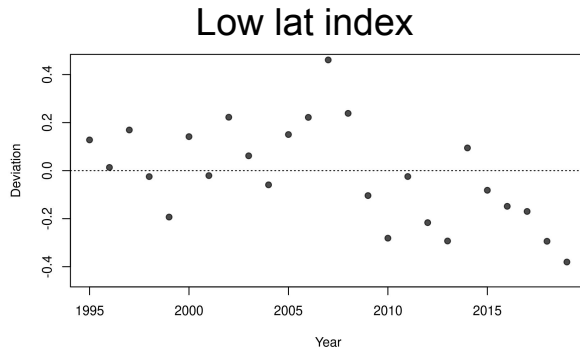
Low lat index

3. Trimming the grid: followed EU proposal:

Axes retained -

- ○ Low Lat. - include EU?
- ○ High Lat. - include early NZ?
- ○ Initial F - high?
- ○ Discard fraction (low/base/high)

SB/SB_{MSY}

F/F_{MSY}
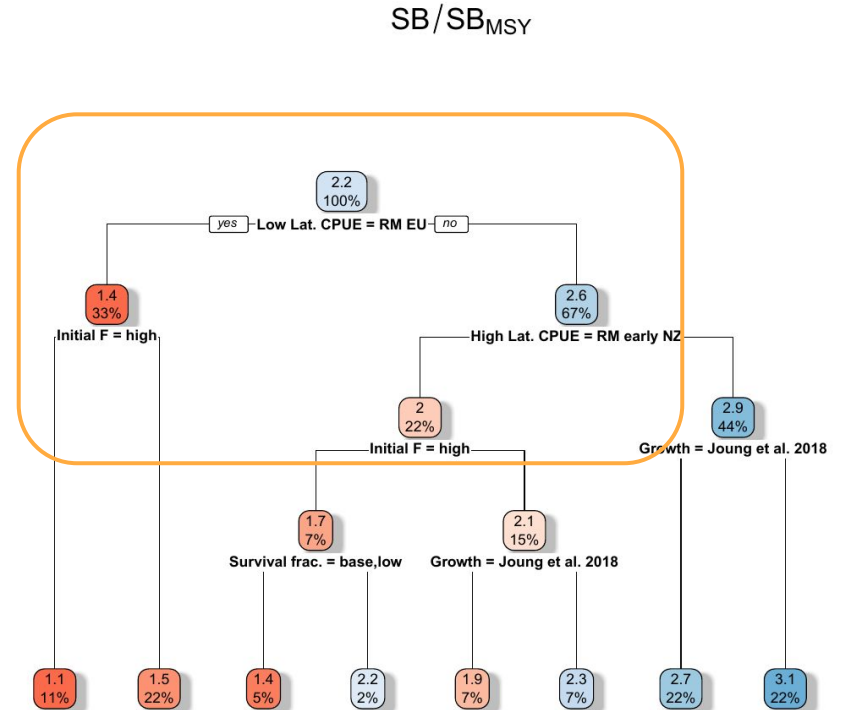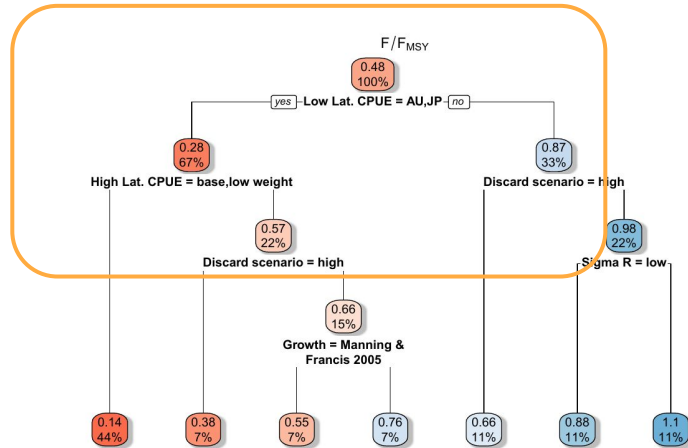
3.   Trimming the grid: followed EU proposal:

Axes retained -
- ○ Low Lat. - include EU?
- ○ High Lat. - include early NZ?
- ○ Initial F - high?
- ○ Discard fraction (low/base/high)
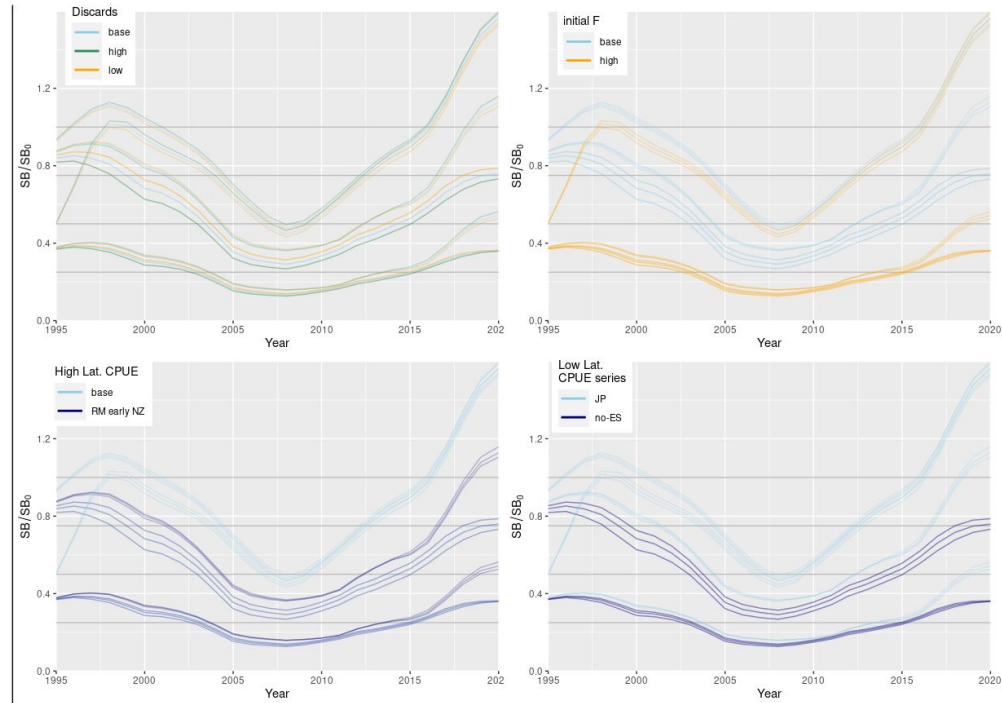
624 models remaining

3. Trimming the grid: followed EU proposal given

Axes retained -
- ○ Low Lat. - include EU?
- ○ High Lat. - include early NZ?
- ○ Initial F - high?
- ○ Discard fraction (low/base/high)

624 models remaining - ran out of time to run retrospectives/jitter etc on an exhaustive set.

Ran jitter and retrospectives on main grid axes that accounted for most variability in outcomes.
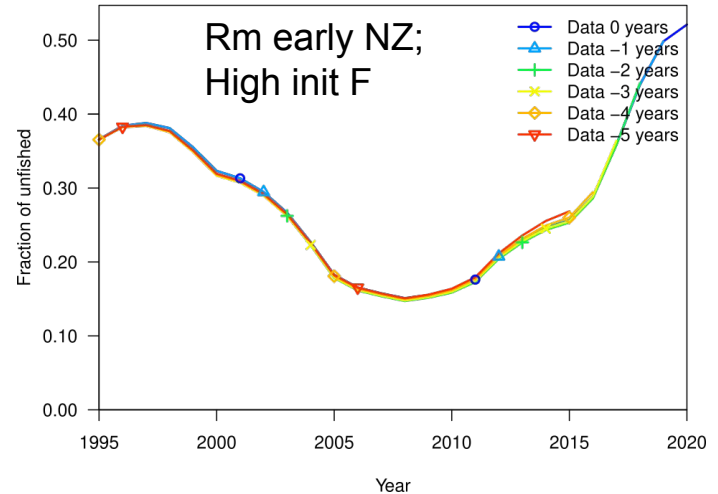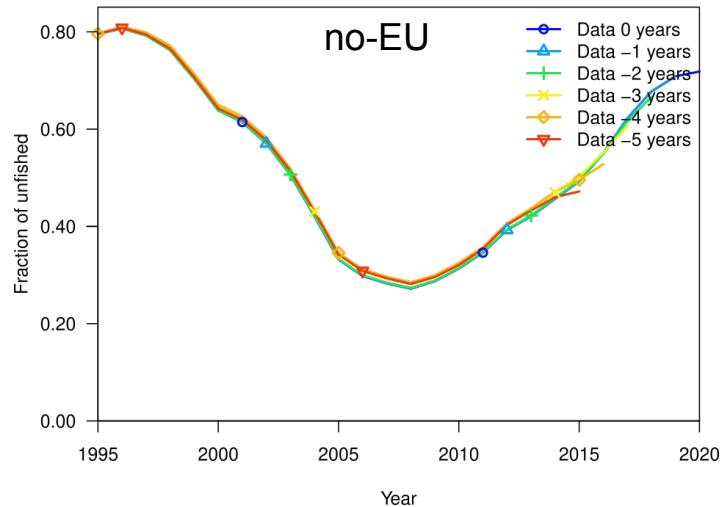
All jitter-runs suggested global optimum reached

Retrospectives showed no discernible patterns. Unclear how best to report on retrospectives: metrics like Mohn's Rho are not always indicative of problems (Hurtado-Ferro et al. 2015 - "The value of Mohn's ρ is not related to either the sign or magnitude of bias in the estimate of terminal year biomass").

3. Trimming the grid: followed EU proposal given

Axes retained -

- ○ <mark>Low Lat. - include EU?</mark>
- ○ <mark>High Lat. - include early NZ?</mark>
- ○ <mark>Initial F - high?</mark>
- ○ Discard fraction (low/base/high)

624 models remaining - ran out of time to run retrospectives/jitter etc on an exhaustive set.

3. Trimming the grid: followed EU proposal given

Axes retained -
- ○ Low Lat. - include EU?
- ○ High Lat. - include early NZ?
- ○ Initial F - high?
- ○ Discard fraction (low/base/high)

624 models remaining - ran out of time to run retrospectives/jitter etc on an exhaustive set.

Ran jitter and retrospectives on main grid axes that accounted for most variability in outcomes.

All jitter-runs suggested global optimum reached

Retrospectives showed no discernible patterns. Unclear how best to report on retrospectives: metrics like Mohn's Rho are not always indicative of problems (Hurtado-Ferro et al. 2015 - "The value of Mohn's ρ is not related to either the sign or magnitude of bias in the estimate of terminal year biomass").

3. Trimming the grid:

Some comments with regards to US proposal:

- Retrospectives are difficult to interpret - when are they OK, when are they bad enough to reject a run? Case by case?

- Normality test assumes data are unbiased with obs error, but indices are not "data", they are an imperfect representation of relative biomass. Are we happy to drop one to get a better fit/normal residuals in an other?

- When uncertainty lies mainly in the inputs, classical model diagnostics/selection tools are not useful other than to delimit models that do not work.

- **As the trimming has made little difference to the outcome, it may be better to use the full grid to provide management advice to avoid unintended bias through ad-hoc grid selection with minimal time for analysis and due consideration.**

Proposed SC report figures based on the grid (examples from full grid)

Stock status metrics table

Table 4: Summary of reference points for 3888 grid models in the structural uncertainty grid

|	Mean	Median	Min	10%	90%	Max
C_{latest}	6010	6188	3219	3580	8454	10349
C_{recent}	6815	7234	4007	4263	9135	9788
MSY	23902	13234	5462	7451	50727	311628
SB_0	45150	27894	10148	13508	91763	455076
SB_{MSY}	21202	13201	4686	6303	42881	210296
SB_{recent}/SB_0	54566	22758	6774	8599	119091	605252
SB_{recent}	47464	18385	5800	7638	106751	560768
SB_{latest}/SB_0	1.03	1.08	0.30	0.58	1.49	1.66
SB_{recent}/SB_0	0.88	0.87	0.27	0.49	1.21	1.29
SB_{latest}/SB_{MSY}	2.19	2.28	0.64	1.23	3.15	3.61
SB_{recent}/SB_{MSY}	1.88	1.84	0.57	1.06	2.57	2.80
F_{MSY}	0.153	0.152	0.132	0.135	0.168	0.182
$F_{lim,AS}$	0.231	0.228	0.199	0.205	0.253	0.274
$F_{crash,AS}$	0.318	0.312	0.274	0.282	0.346	0.377
F_{latest}	0.073	0.066	0.002	0.013	0.153	0.216
F_{recent}	0.089	0.075	0.002	0.015	0.191	0.282
F_{latest}/F_{MSY}	0.48	0.43	0.01	0.08	1.00	1.29
F_{recent}/F_{MSY}	0.58	0.48	0.01	0.10	1.24	1.68
$F_{latest}/F_{lim,AS}$	0.32	0.28	0.01	0.06	0.66	0.87
$F_{recent}/F_{lim,AS}$	0.38	0.32	0.01	0.06	0.82	1.13
$F_{latest}/F_{crash,AS}$	0.23	0.21	0.01	0.04	0.48	0.63
$F_{recent}/F_{crash,AS}$	0.28	0.23	0.01	0.05	0.59	0.83

Catch plots

Mortality including post release mortality
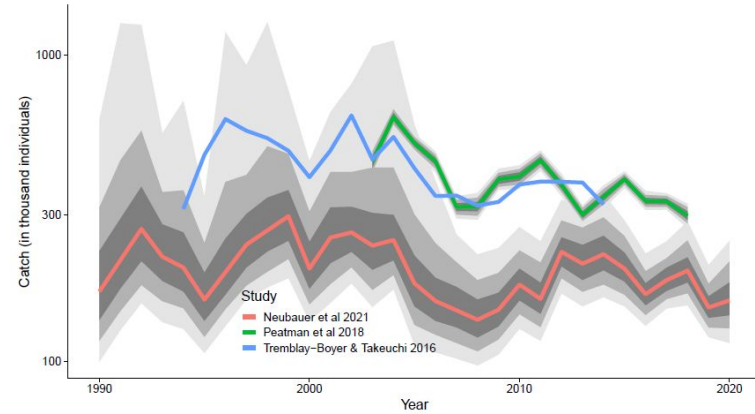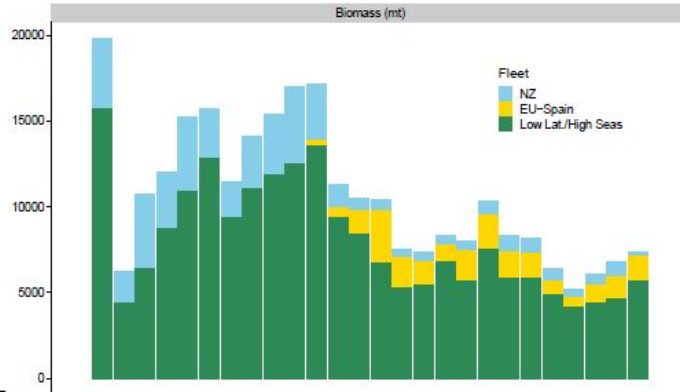
Recruitment plot
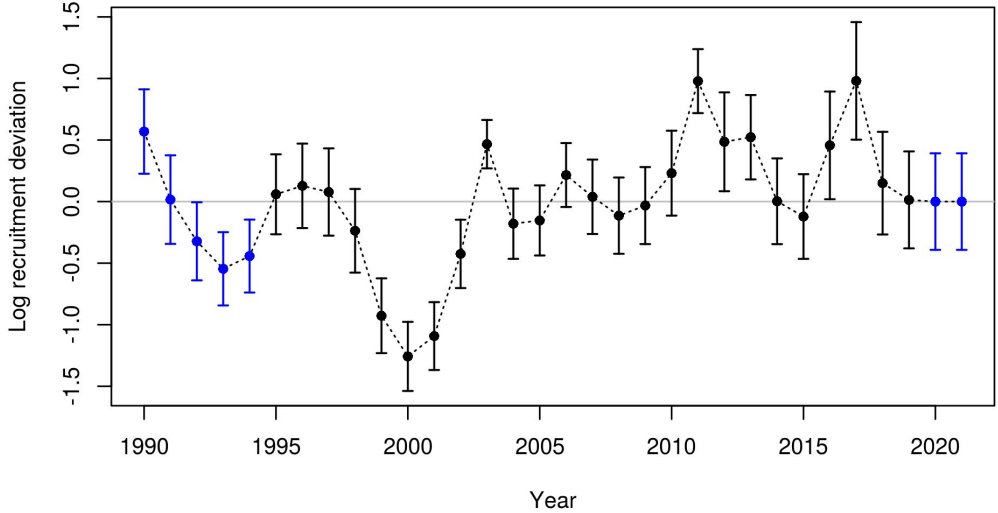
Fishing mortality
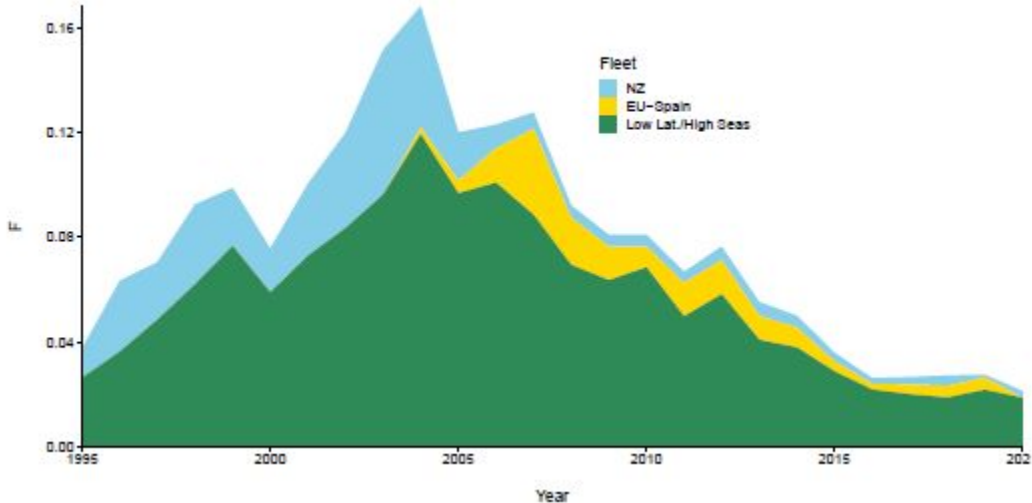
Spawning biomass
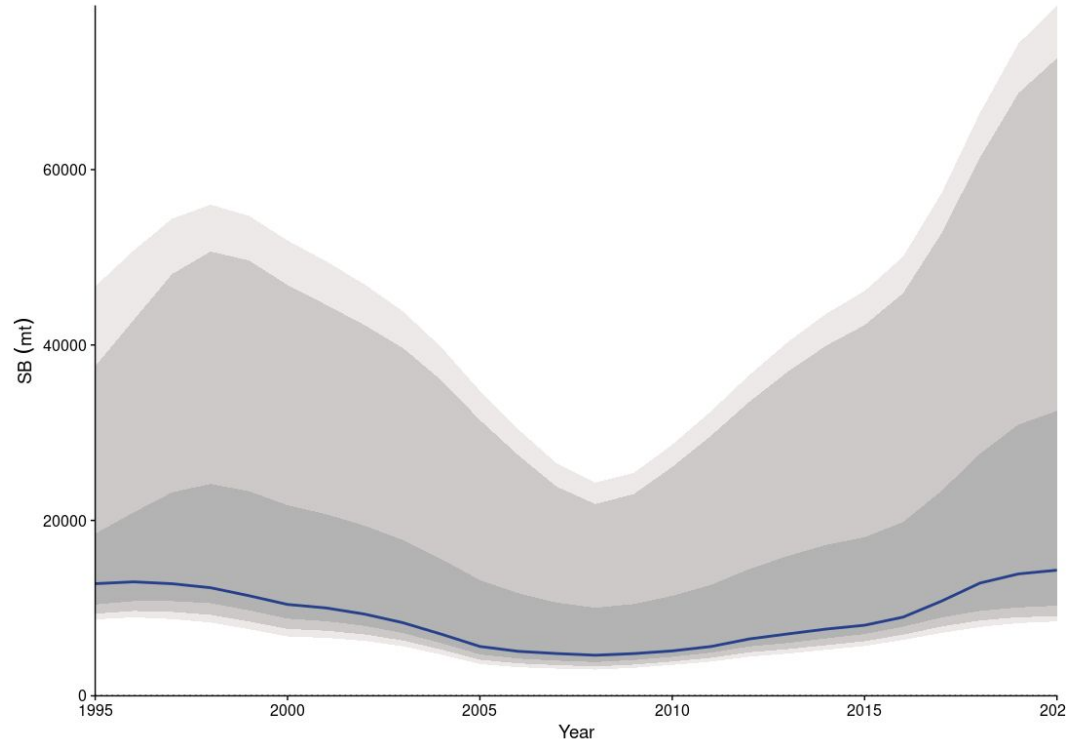
Depletion plot
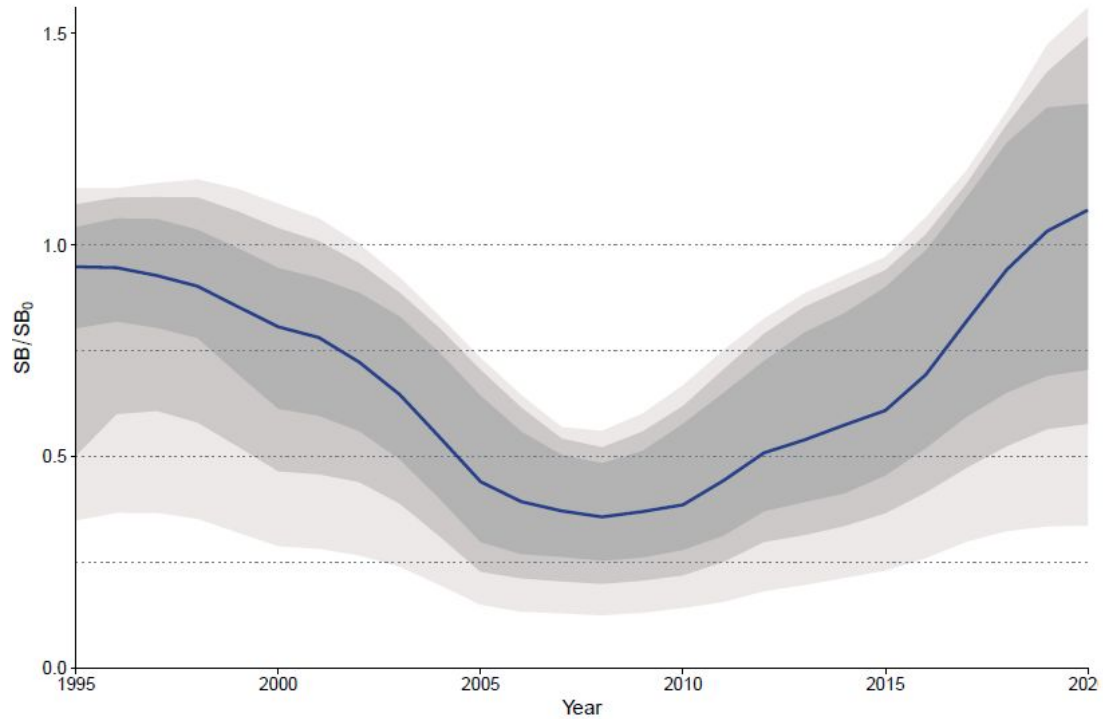
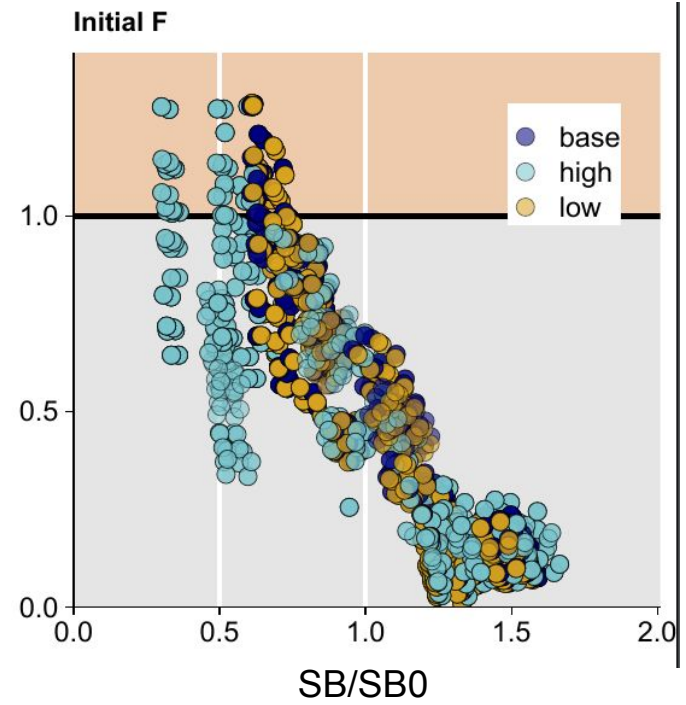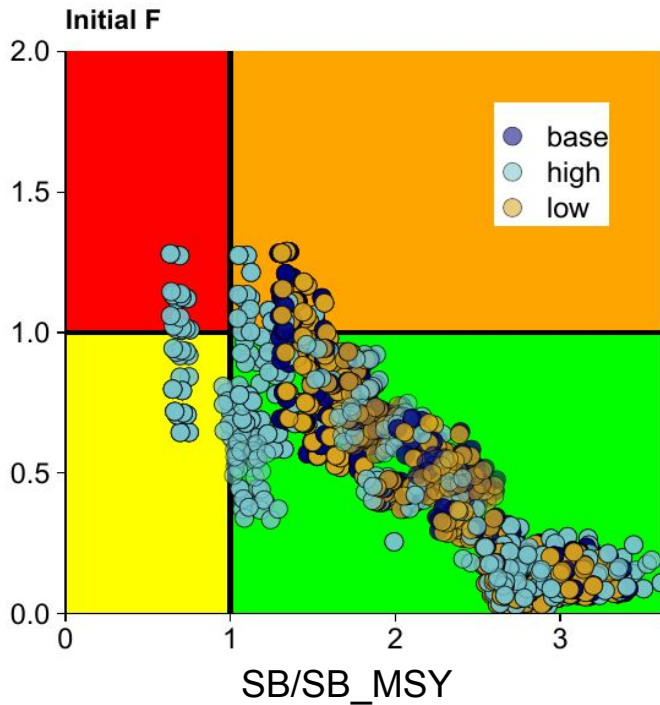# Kobe plots

Optional alternative reference points plot