Modelling partitions using MULTIFAN-CL

- space, tagged populations, species, stocks, and gender
- coding implementation and recommendations

Nick Davies, Dave Fournier, John Hampton

Outline

- Introduction to MULTIFAN-CL
- Partitioning:
  - Space
  - Tagging
  - Species/Stocks/Sexes
- Software development and testing
- Improvements possible and development recommendations

1. Introduction to MULTIFAN-CL

What is MULTIFAN-CL?

- An integrated age-structured, length-based population model
  - **Integrated** - integrates the fit over multiple data types
  - **Age-structured** - core population state is dimensioned by age
  - **Length-based** because it is conditioned on:
    - size composition
    - tag recaptures at length

Particularly useful for tropical species for which **age-specific data is generally unavailable**

Brief history

- First **MULTIFAN** – estimated growth from cohort LF
- Placed into an **age-structured model** that fitted to catch-at-length data, **SPARCLE** (South Pacific Albacore Research Catch-at-Length Estimators) in 1993
- **MULTIFAN-CL** was introduced in CJFAS in 1998; followed by the YFT assessment in 2001; simulation tested in 2005
- Used in WCPO stock assessments for nearly 20 years

Fournier, D.A., Hampton, J., and Sibert, J.R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, Thunnus alalunga. *Can. J. Fish. Aquat. Sci.*, 55:2105-2116.

Hampton, J., and D.A. Fournier. 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (Thunnus albacares) in the western and central Pacific Ocean. *Mar. Fresh. Res.*, 52:937-963.

Labelle, M. 2005. Testing the MULTIFAN-CL assessment model using simulated tuna fisheries data. Fisheries Research 03/2005; DOI: 10.1016/j.fishres.2004.08.035

Team

- Primary developer – Dave Fournier
- Development support and testing – Nick Davies
- Direction – John Hampton and Graham Pilling
- Software support (Viewer, Repository) – Fabrice Bouyé

License: Oceanic Fisheries Programme, Pacific Community, (SPC) and Otter Research Ltd

Overview of the code project

- C++
- 250 source files
- Dependent libraries:
  - ADMB – data structures and auto-diff; in-house version
  - OpenBLAS – matrix routines
  - QD – precision
  - PVM3 – multi-threading
- Generic – using input flags (5 types) act as “switches” to control:
  - Program operation, processes, likelihoods and parameterisations
  - Fitting procedure
  - Simulation projections
  - Reporting
- Website: www.multifan-cl.org – for binaries and documentation

2. Space partition

Rationale:

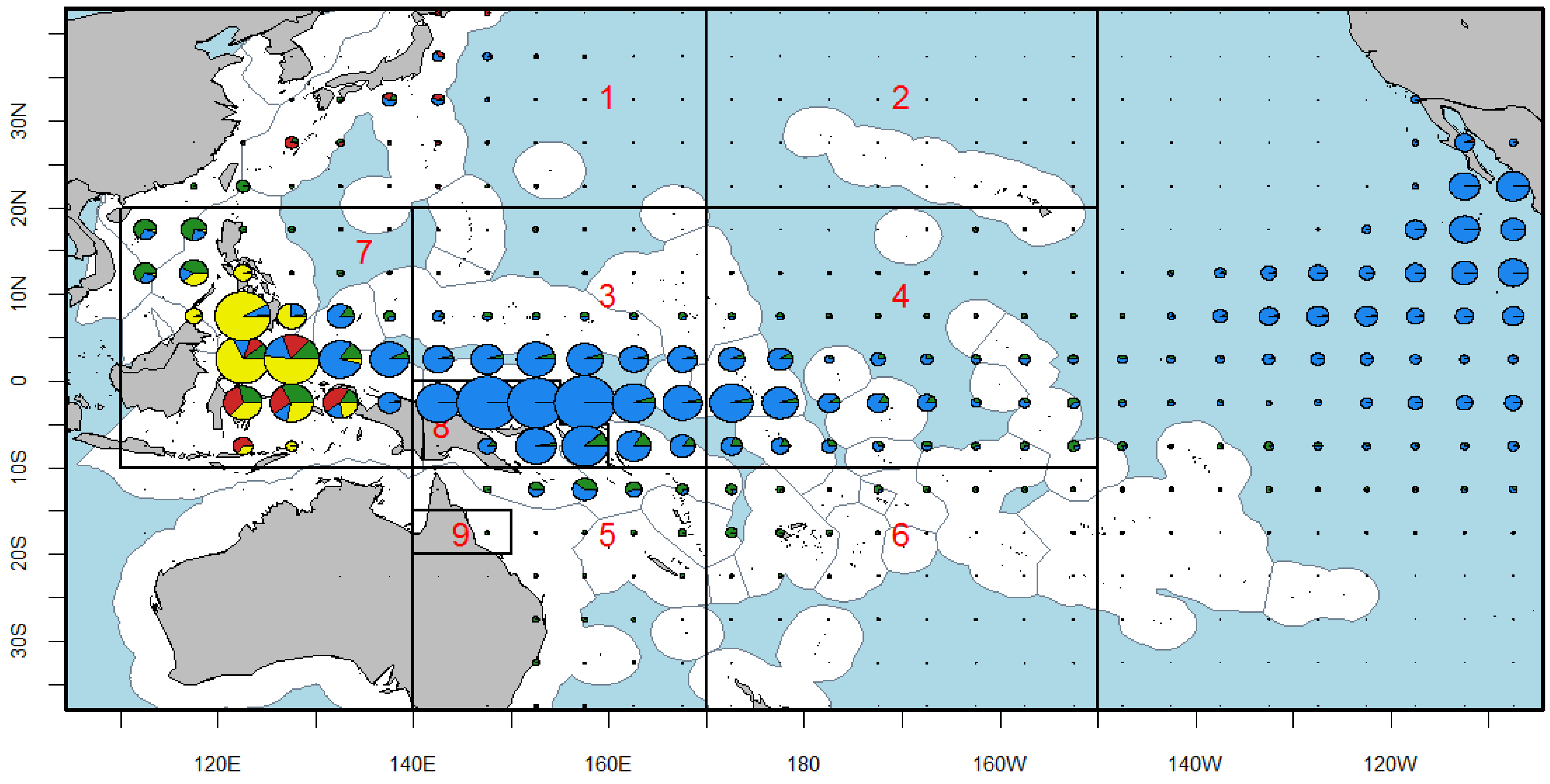
- to explicitly describe spatial processes that lead to heterogeneity within the fish stock
- to account for heterogeneity in fleet structure or management measures

In summary - to explicitly describe the variable effects of fishing mortality on the stock by area

Spatial complexity in WCPO

Heterogeneity:

- fisheries
- biology



Space partition

- Region is the basis dimension of the population data structure

N(**1,nregions**,1,nyrs,1,nage)

- Region: a subset within the stock unit in which the sub-population within a time period is **discrete** from other subsets
- Movement among regions is possible.

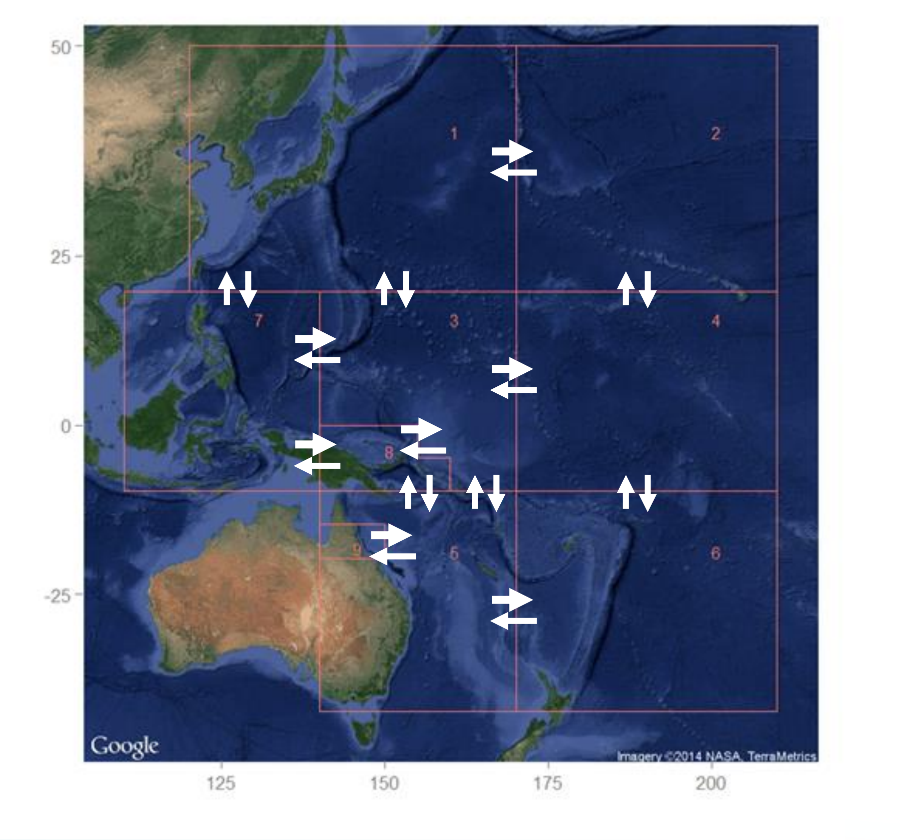
Movement parameterisation

- Temporal:
  - Flexibility for number of movements per year (e.g. each quarter)
  - Flexibility for grouping (shared) movements for particular time periods, e.g. 1 2 2 1 rather than 1 2 3 4
- Spatial:
  - coefficients estimated are region boundary-specific, so flexibility for:
    - number of regions
    - adjacent regions

Movement processes in MULTIFAN-CL cont.

Movement matrix in respect of regions

e.g. yellowfin tuna



Destination region

Incidence matrix

Region of origin	R2	R3	R4	R5	R6	R7	R8	R9
R1------------------	**1**	**1**	0	0	0	**1**	0	0
R2------------------		0	**1**	0	0	0	0	0
R3------------------			**1**	**1**	0	**1**	**1**	0
R4------------------				0	**1**	0	0	0
R5------------------					**1**	0	**1**	**1**
R6------------------						0	0	0
R7------------------							**1**	0
R8------------------								0

- Flexibilty in respect of the vectors of movement coefficients in respect of: spatial structure, temporal assumptions

Region structure

Incidence matrix

	2	**3**	**4**	**5**
1	0	0	1	1
2		0	1	1
3			1	1
4				0

nvar = 12 coefficients; $\boldsymbol{m_{ij}}$

i	***j***
1	4
1	5
2	4
2	5
3	4
3	5
4	1
4	2
4	3
5	1
5	2
5	3

Movement processes cont.

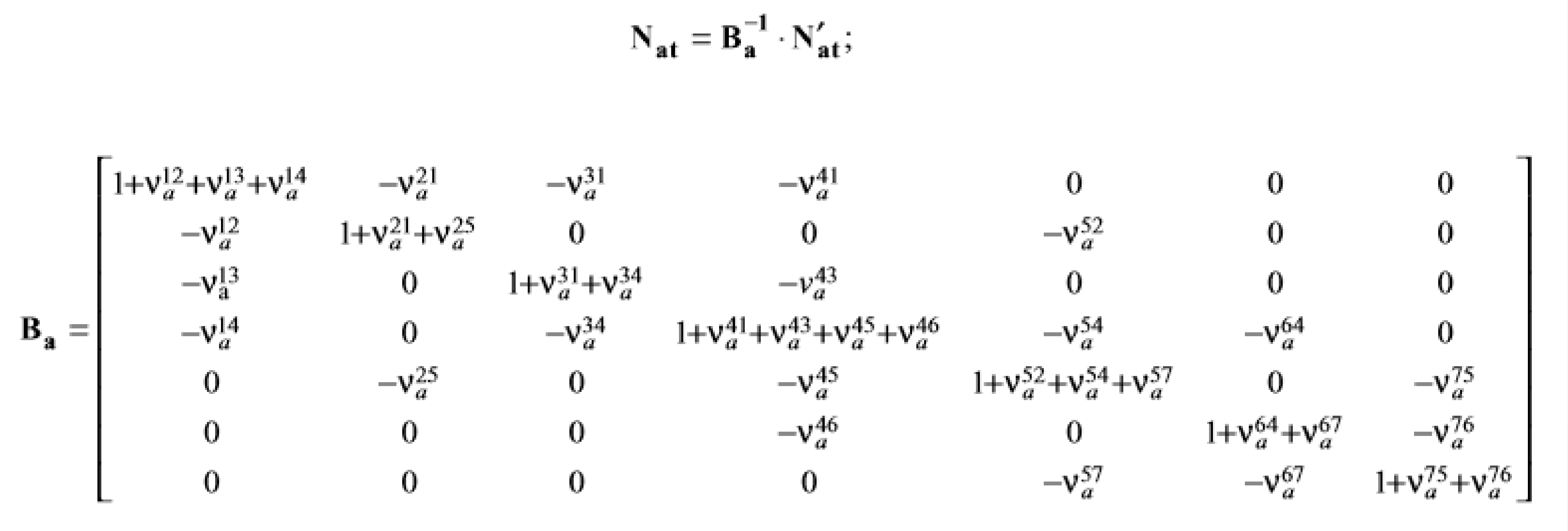
- movement occurs at start of each $\boldsymbol{t}$
- region-specific coefficients $\boldsymbol{v_a^{rs}}$ from region $\boldsymbol{r}$ to region $\boldsymbol{s}$
- abundance after movement is

$$N_{atr} = N'_{atr} - \left(\sum_{s \neq r} v_a^{rs} \right) N_{atr} + \sum_{s \neq r} v_a^{sr} N_{ats}$$

$\boldsymbol{N'_{atr}}$ and $\boldsymbol{N_{atr}}$ are pre- and post-movement

Movement processes cont.

- Fully implicit solution → movement can occur to all regions (including non-adjacent) in a single time step
- guarantees numerical stability



$$\mathbf{N_{at}} = \mathbf{B_a^{-1}} \cdot \mathbf{N'_{at}};$$

$$\mathbf{B_a} = \begin{bmatrix} 1+\nu_a^{12}+\nu_a^{13}+\nu_a^{14} & -\nu_a^{21} & -\nu_a^{31} & -\nu_a^{41} & 0 & 0 & 0 \\ -\nu_a^{12} & 1+\nu_a^{21}+\nu_a^{25} & 0 & 0 & -\nu_a^{52} & 0 & 0 \\ -\nu_\mathrm{a}^{13} & 0 & 1+\nu_a^{31}+\nu_a^{34} & -\nu_a^{43} & 0 & 0 & 0 \\ -\nu_a^{14} & 0 & -\nu_a^{34} & 1+\nu_a^{41}+\nu_a^{43}+\nu_a^{45}+\nu_a^{46} & -\nu_a^{54} & -\nu_a^{64} & 0 \\ 0 & -\nu_a^{25} & 0 & -\nu_a^{45} & 1+\nu_a^{52}+\nu_a^{54}+\nu_a^{57} & 0 & -\nu_a^{75} \\ 0 & 0 & 0 & -\nu_a^{46} & 0 & 1+\nu_a^{64}+\nu_a^{67} & -\nu_a^{76} \\ 0 & 0 & 0 & 0 & -\nu_a^{57} & -\nu_a^{67} & 1+\nu_a^{75}+\nu_a^{76} \end{bmatrix}$$

Movement processes cont.

- Options for age dependency of the movement coefficients:

$$\nu_a^{rs} = \phi_0^{rs} \exp\left(\phi_1^{rs} (\kappa_{\mathrm{a}})^{\phi_2^{rs}}\right)$$

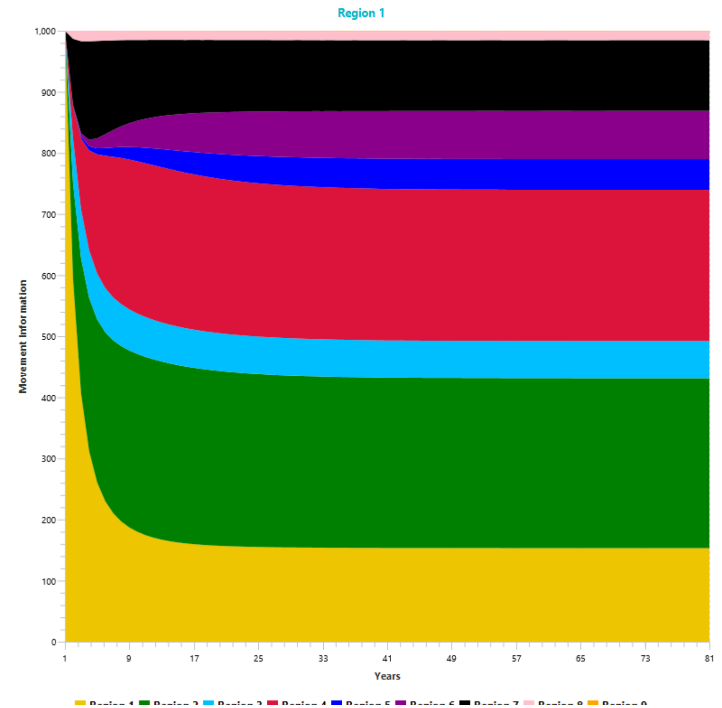
  - linear-dependence ϕ_1^{rs}
  - non-linear-dependence ϕ_2^{rs}

Example: **Bigeye tuna** – selected regions only

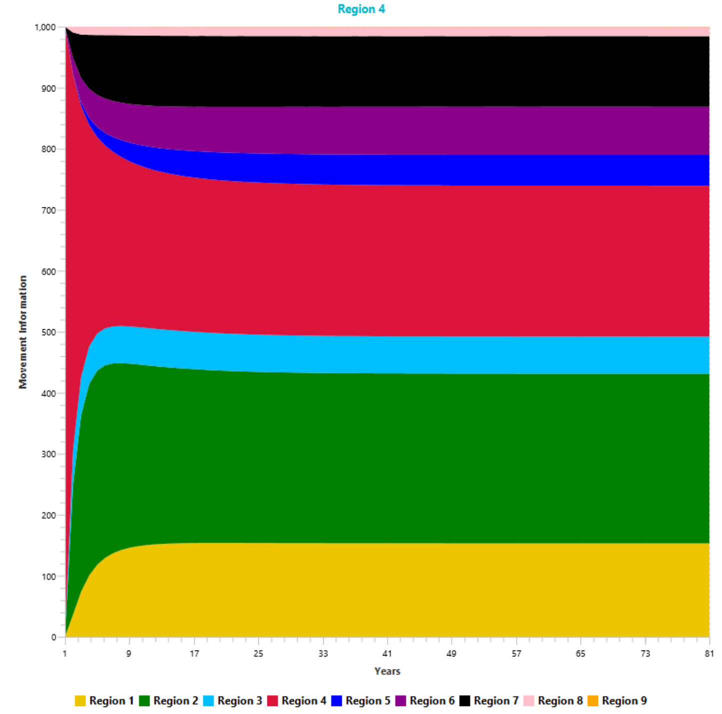
nregions = 9

n_coffs = 26

4 movement time periods



Region 1



Region 4

3. Tag partition

Tagged population model

- Each release event represents a “parallel” **tagged** population

tagnum_fish(**1,tag_events**,1,nregions,1,nperiods, 1,nage)

- shares most of the dynamic processes with the model (**un-tagged**) population
- **Identical movement and growth processes**
- The exception process is recruitment: where for the tagged population a **“cohort”** is a release event comprising a “recruitment” in numbers at length

Tagged population model

- MULTIFAN-CL is age-structured. Tag releases are **length-specific** – transformed to be **age-specific** via the estimated growth function
- A **mixing period** is specified for assumed random mixing of tagged population
- Grouping of recaptures: specified fisheries for which recaptures can be aggregated

Tag population dynamics – post-release

- Mixing phase of specified time periods
  - not included in likelihood
- Mixing phase fishing mortality = corrected tag returns during period

$$\mathrm{fn}_3(R^{\mathrm{Tobs}}_{ctf}, X_{tf}) \ ; \quad t^{\mathrm{rel}}_c \le t < t^{\mathrm{rel}}_c + n^{\mathrm{mix}}$$

- Fishing mortality on tag cohort (post-mix)

$$F_{a(c,t),t,f} \ ; \qquad t \ge t^{\mathrm{rel}}_c + n^{\mathrm{mix}}$$

Tag population dynamics - release

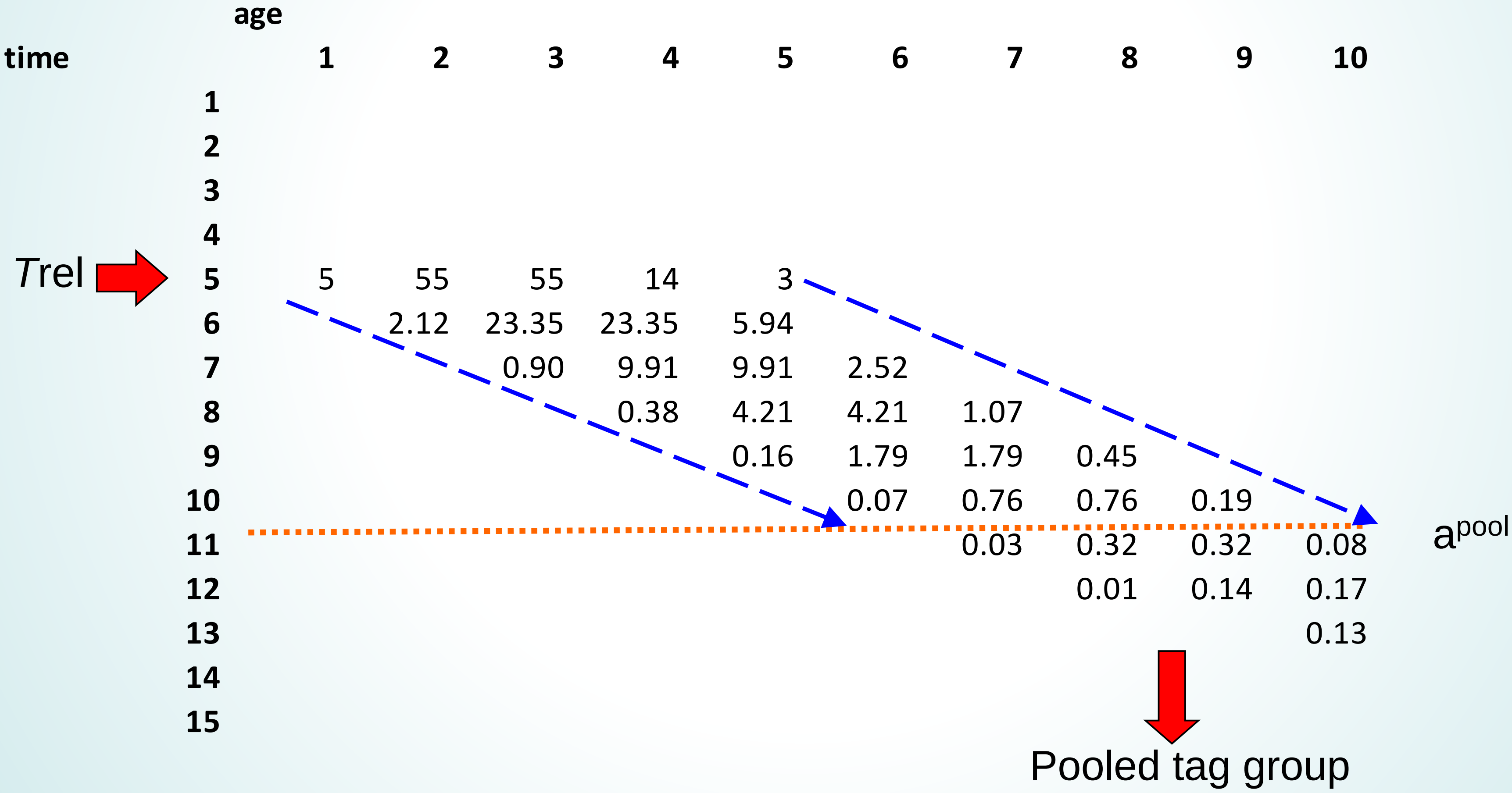
- Tags released are assigned to a ***tag cohort*** (c) – being a tag release **group**

- **Pooled group** – aggregate cohorts (c*) from tag groups in a single group when attain a^{pool}. Age structure is maintained

Tag cohorts - releases

Tag group X N_c^{rel} $N^{\mathrm{T}}_{c,t-1,r} e^{-Z^{\mathrm{T}}_{c,t-1,r}}$ $Z^{\mathrm{T}}_{ctr} = \sum_{f \in f_r} F^{\mathrm{T}}_{ctf} + M_{a(c,t)}$

Region 1



	age									
time	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1										
2										
3										
4										
5	5	55	55	14	3					
6		2.12	23.35	23.35	5.94					
7			0.90	9.91	9.91	2.52				
8				0.38	4.21	4.21	1.07			
9					0.16	1.79	1.79	0.45		
10						0.07	0.76	0.76	0.19	
11							0.03	0.32	0.32	0.08
12								0.01	0.14	0.17
13										0.13
14										
15										

Trel

a^{pool}

Pooled tag group

Predicted tag recaptures

- Fishery specific reporting rate $\boldsymbol{X_{tf}}$

Predicted recoveries by tag cohort (*c*), age (a), time (*t*), and fishery (*f*). From each tag group and the pooled tag group.

$$R^{\mathrm{Tpred}}_{ctf} = \frac{F^{\mathrm{T}}_{ctf} X_{tf}}{Z^{\mathrm{T}}_{ctr}} \left[1 - e^{-Z^{\mathrm{T}}_{ctr}}\right] N^{\mathrm{T}}_{ctr} \;\; ; \;\; t \geq t^{\mathrm{rel}}_{c} + n^{\mathrm{mix}}$$

Recapture grouping

- Option for aggregating predicted recoveries by fishery groupings (*g*) – aggregated tag returns, e.g. PS fishery.

Observed tag recaptures

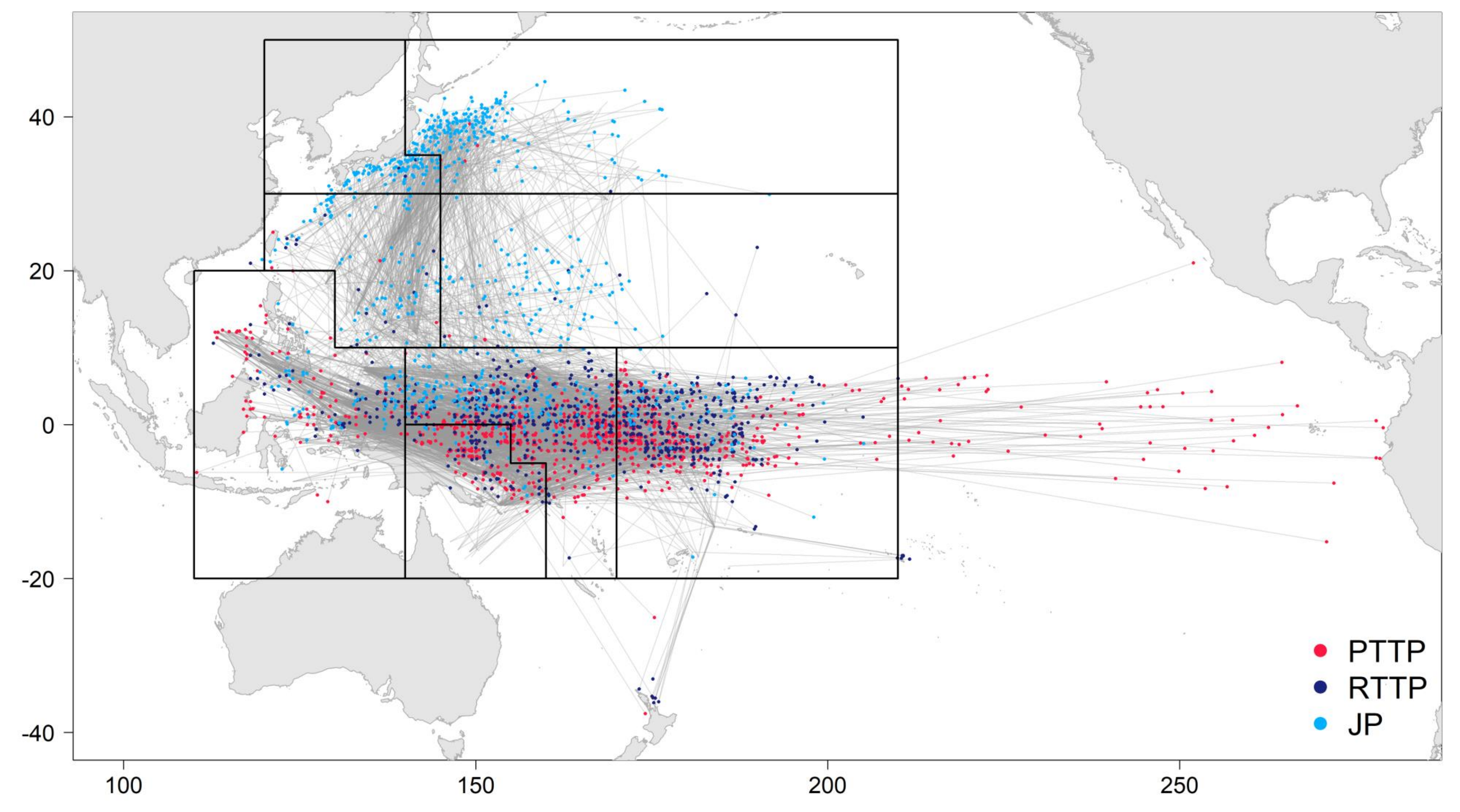
- Observed details: release group, length (at release), recapture period, fishery at recovery.
- Release cohort’s ages at recapture is calculated using: length at release, and time at liberty.
- Assignment to **regions of release and recapture** based on tag group and fishery of recapture.

Example: WCPO skipjack tuna 2019

- Tagging data input
- Fitted recaptures
- Movement estimates

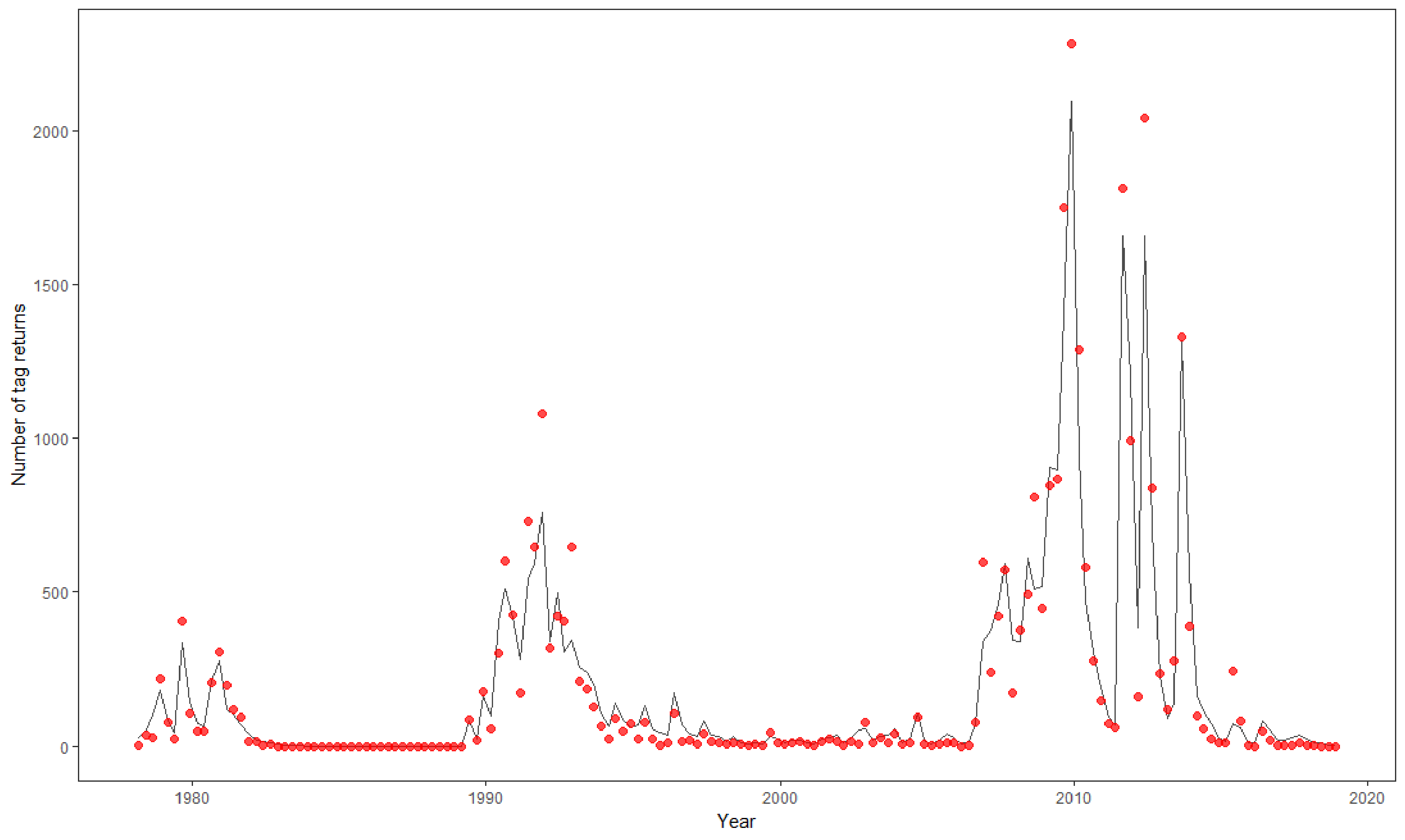
Vincent et al. 2019. WCPFC-SC15-2019/SA-WP-05-Rev2

Tag movements - SKJ



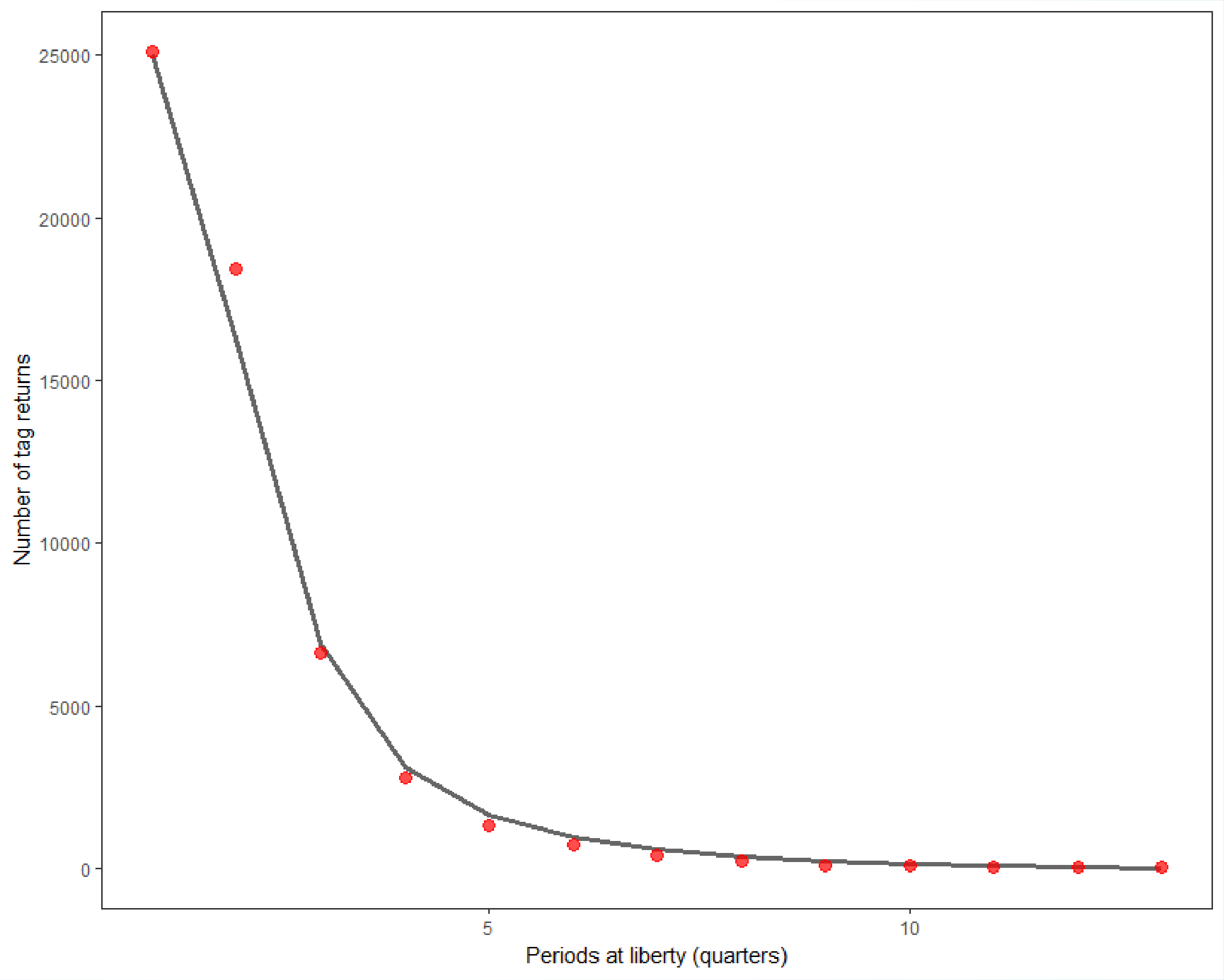
269 release groups 329,811 releases 56,092 recaptures

Fits to the tagging data - SKJ

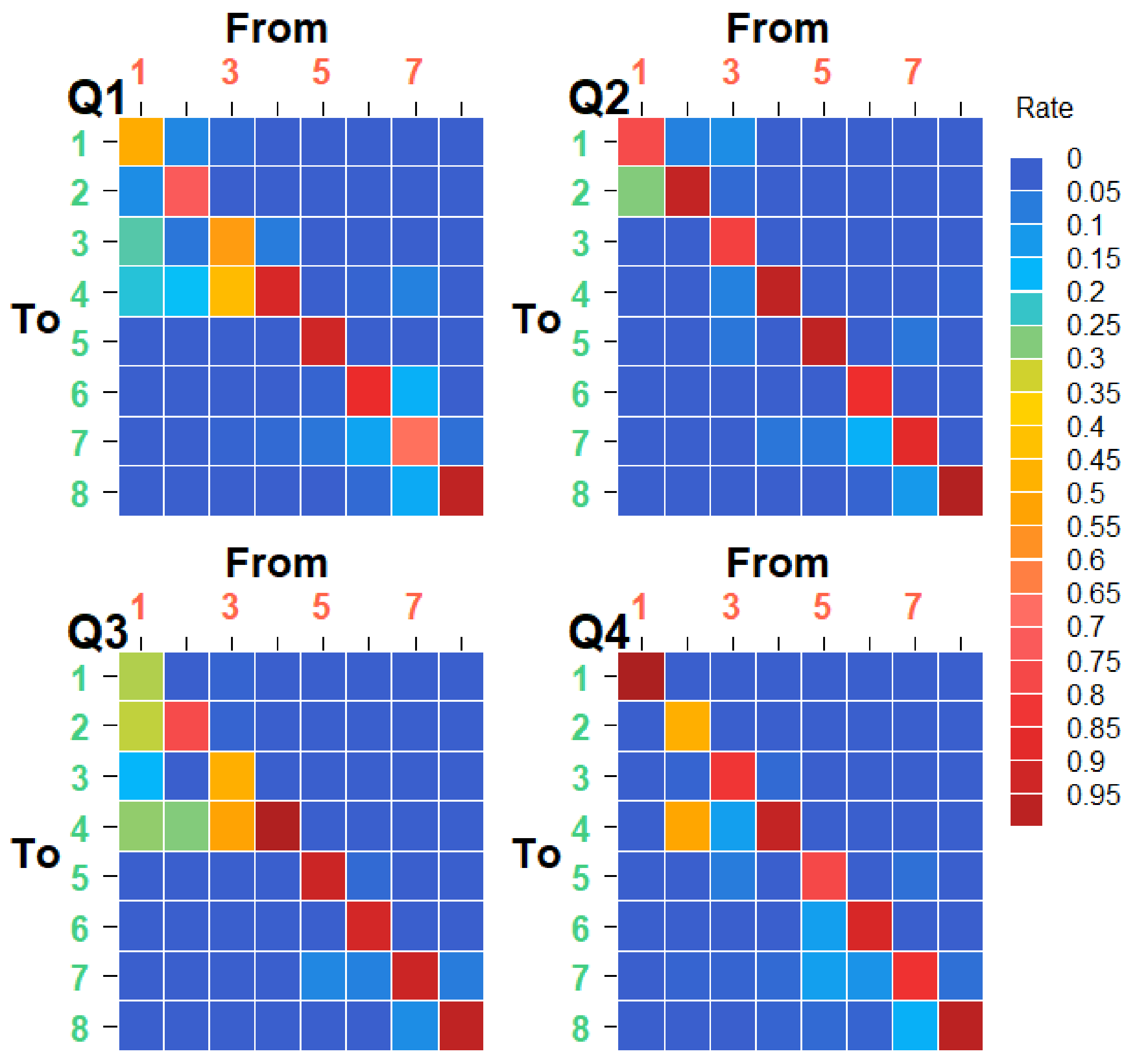


Fits to the tagging data - SKJ

Fit to observed tag recaptures in respect of time at liberty



Movement estimates - SKJ



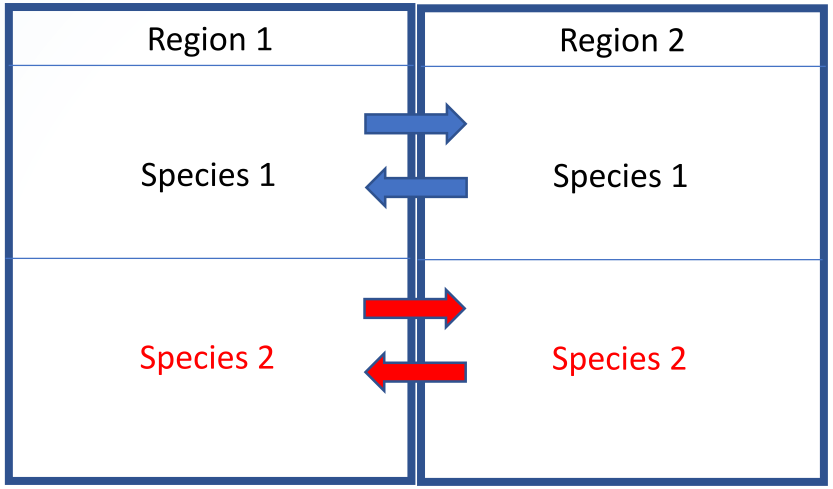
4. Partitions for: species, stocks, and gender

- A <u>substantial</u> development – so why?
- Rationale:
  - Sharing of parameters among species/stock/sex
    - reduces the number of parameters among assessments
    - shared catchability for mixed-species fisheries
  - Better describe stock-, sex-specific biological processes
  - Capability to fit to data aggregated among species/stocks

Adding the species partition

Region adaptation for multi-species mode in MULTIFAN-CL

Reality model



Region 1

Species 1

Species 2

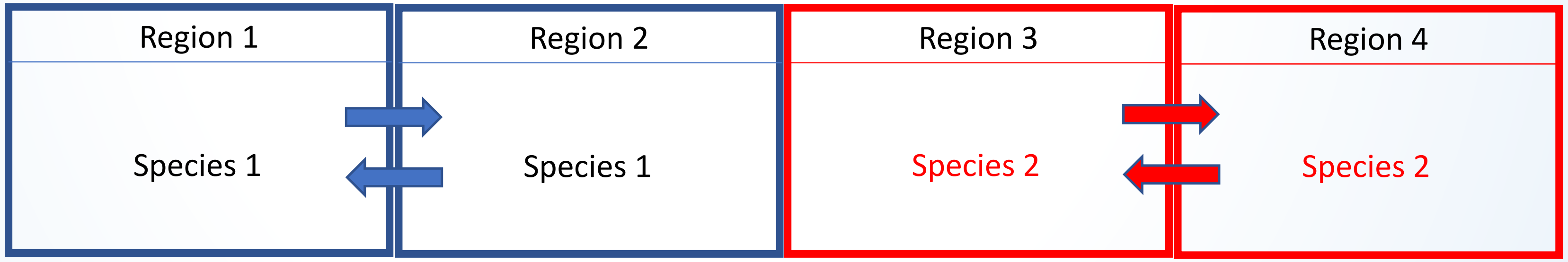
Region 2

Species 1

Species 2

- Specific recruitment
- Specific growth
- Specific movement

Region adaptation



Region 1

Species 1

Region 2

Species 1

Region 3

Species 2

Region 4

Species 2

- Specific recruitment
- Specific growth
- Specific movement

Adding the species partition

Regional fisheries adaptation for multi-species mode in MULTIFAN-CL

Reality model

Region 1	Region 2
Fishery 1 Fishery 2	Fishery 3 Fishery 4

Region adaptation

Region 1	Region 2	Region 3	Region 4
Fishery 1 Fishery 2	Fishery 3 Fishery 4	Fishery 5 Fishery 6	Fishery 7 Fishery 8

- Fisheries mirrored from regions 1 and 2 into regions 3 and 4
- Data mirrored for regions 3 and 4
- Data is aggregated or dis-aggregated

Code structure for the region adaptation

- A multi-species **class** (`pmsd`) was created to contain the species > 1 members for parameters, pointers, etc.
- Operations were then performed if `pmsd` is allocated

```
if (pmsd && pmsd->num_species>1)
{
  int ns=pmsd->num_species;
  for (int is=2;is<=ns;is++)
  {
    if (parest_flags(400)==0)
    {
      set_value(pmsd->recr(is)(max(2,first_unfixed_year),
        last_real_year),x,ii,-20.,20,pen,
        100,year_flags(1),-10.0);
    }
    else
    …
```

Code structure for the region adaptation

- Region-specific operations performed in respect of the mirrored regions for each species

```
for (int is=1;is<=fsh.pmsd->num_species;is++)
  {
    int rmin=fsh.pmsd->region_bounds(is,1);
    int rmax=fsh.pmsd->region_bounds(is,2);
    dmatrix tmp(1,fsh.nyears,1,fsh.nage);
    tmp.initialize();
    for (int ir=rmin;ir<=rmax;ir++)
    {
      tmp+=value(fsh.F_by_age_by_year_by_region(ir));
    }
    …
```

Fisheries and research data –

dis-aggregated/aggregated

Fishery observations of catch and size compositions (LF and WF)

- any combination of each data type may be either aggregated or disaggregated among species in any specified fishing incident **using input flags**
- If aggregated:
  - Scaling of the species-specific predicted proportions at size for each fishing incident by the species ratio in the predicted catches for the incident
  - predictions are aggregated in the likelihood

Fisheries and research data –

dis-aggregated/aggregated

Tag partition – readily accommodated multi-species – data structures are region-specific

Tagging data:

- any release event may be disaggregated or aggregated
- **using species flags** in the input tag data header record
- Disaggregrated events assigned to regions$_{\text{species}}$
- Aggregated observations:
  - replicated for species > 1
  - initial release tagged population is apportioned over species according to the predicted region-specific ratios
  - the predicted recaptures for each species in the group are aggregated before fitting to the observed aggregated tag recaptures

Method is generic: species / stocks / sexes

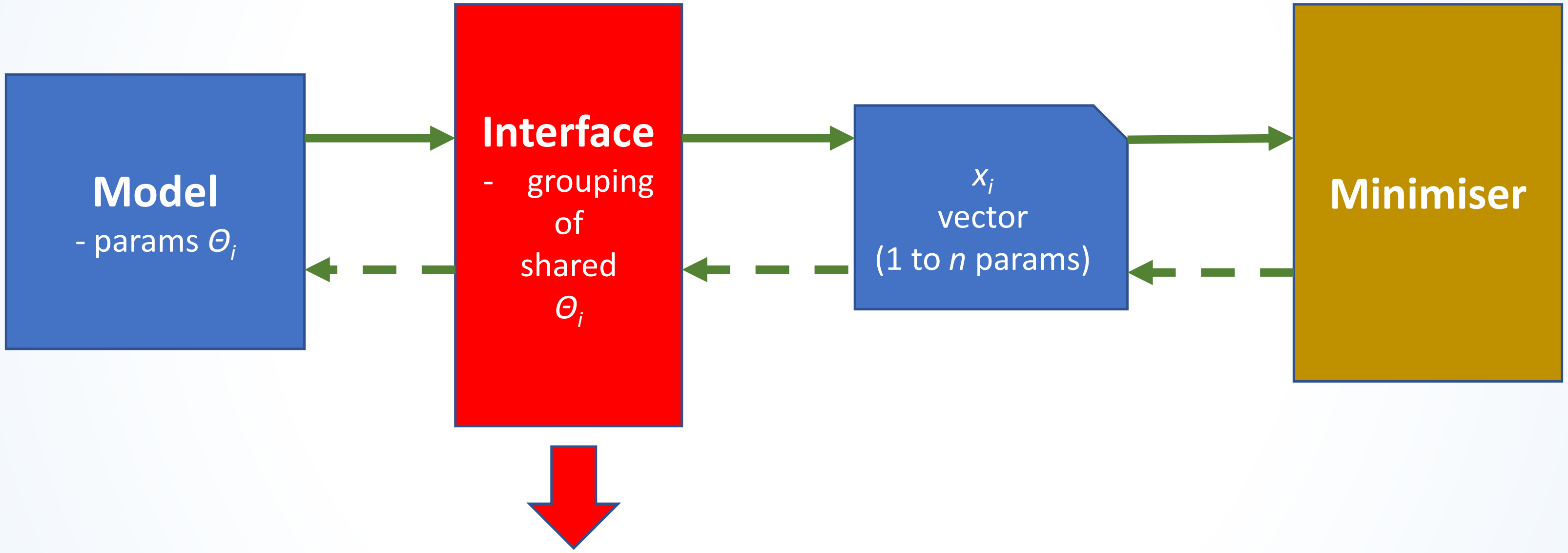
- Region adaption facilitates partitioning by species
- Mirrored regions may act as generic “structures” for partitions for all of: **species**, **stocks**, or **sexes**
- Stocks and sexes can be treated as ***special cases*** of species – where some biological and fishery processes are shared
- Differences in the processes for partitioning by species, stocks, or sexes are managed via alternative parameterizations for each, achieved by sharing or not …

Parameterisation differences – species/stock/sex

Configuration	Species	Stock	Sex
Selectivity at length	specific	shared	shared
Catchability – average and deviates	specific	shared	shared
Recruitment region of origin	all or specific	specific	all or specific
Recruitment + deviates	specific	specific	shared
BH-SRR	specific	specific	female only

Other biological parameters may be specific in all cases:
natural mortality, growth, maturity, etc.

Parameter management for minimisation



Interface - routines to perform various groupings for parameter vectors or matrices

Interface - method for parameter sharing

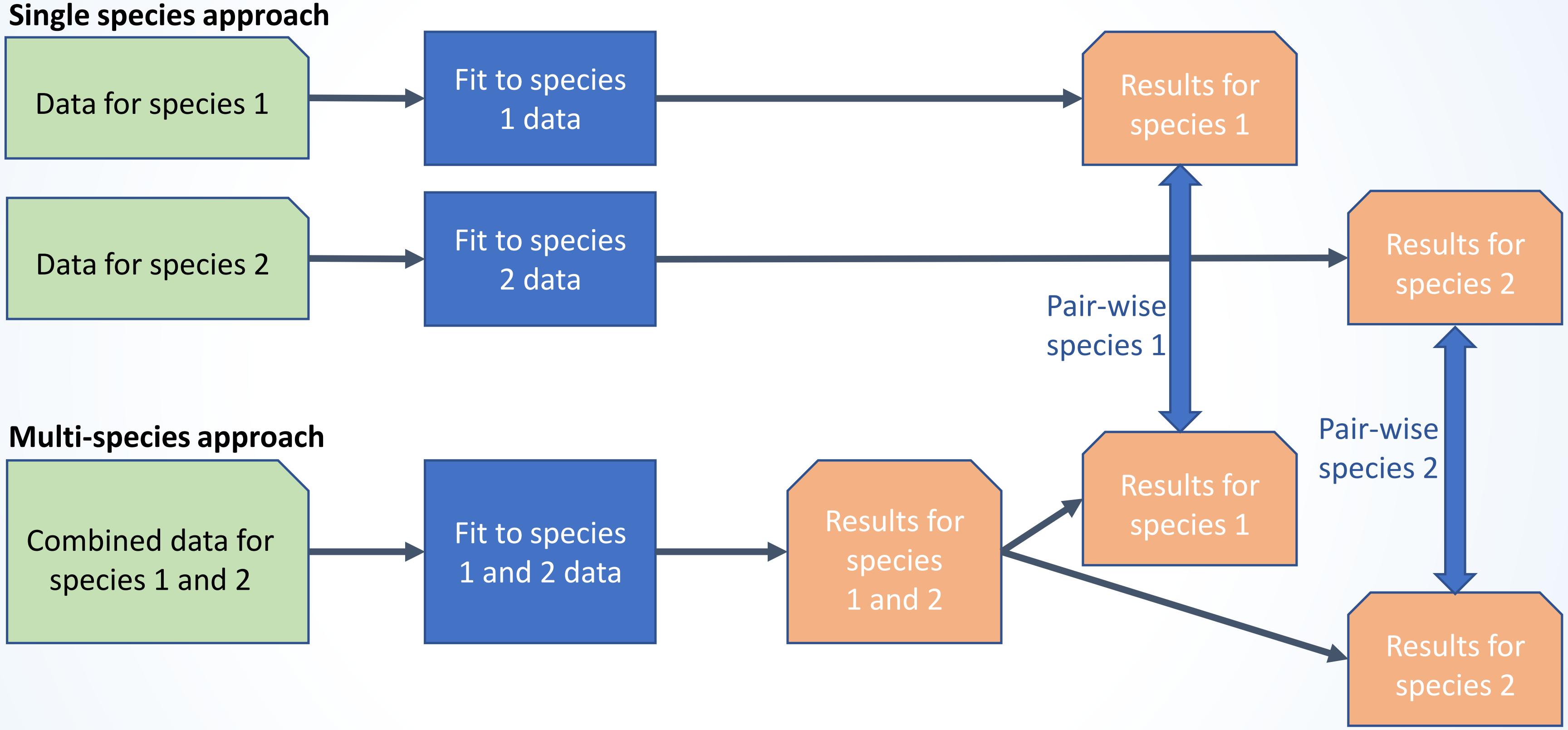
Groups		**Fisheries**		**Mirrored fisheries** (offset = number of fisheries)	
1	… … … …	$\{i_1, i_2 \dots , i_k\}$	;	$\{(i_1+\text{offset}), \dots , (i_k+\text{offset})\}$	;
2	… … … …	$\{j_1, j_2 \dots , j_l\}$	;	$\{(j_1+\text{offset}), \dots , (j_l+\text{offset})\}$	;
.		.		.	
.		.		.	
.		.		.	
n	… … … …	$\{m_1, m_2, \dots , m_p\}$	;	$\{(m_1+\text{offset}), \dots , (m_p+\text{offset})\}$	;
		Species 1		**Species 2**	

… ***n*** species

Developing and testing the multi-species code

- A <u>substantial</u> development requiring rigorous testing so as not to break the “single species” code
- Generated test model data sets of reduced size (BET/YFT), in a new input format to integrate multi-species data
- A testing framework was created:
  - Deterministic evaluations using input parameters in multi-species format (but generated from the single-species solutions) - generated identical model quantities to the SSM
  - Fitted multi-species models compared to the single-species solutions
- Routine benchmark tests of single species data using code versions before and after the development.

Flow diagram – multi-species model testing



Single species approach

Data for species 1

Fit to species 1 data

Results for species 1

Data for species 2

Fit to species 2 data

Results for species 2

Pair-wise species 1

Pair-wise species 2

Multi-species approach

Combined data for species 1 and 2

Fit to species 1 and 2 data

Results for species 1 and 2

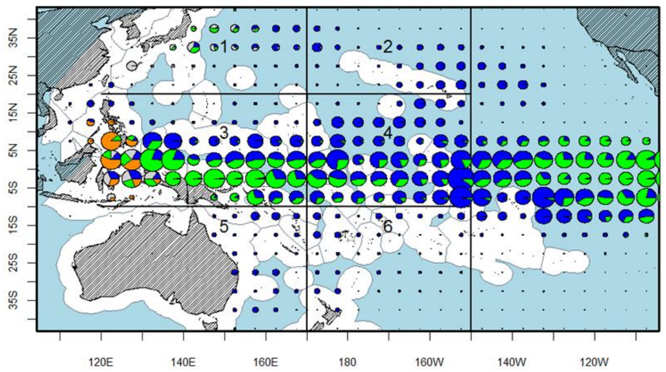
Results for species 1

Results for species 2

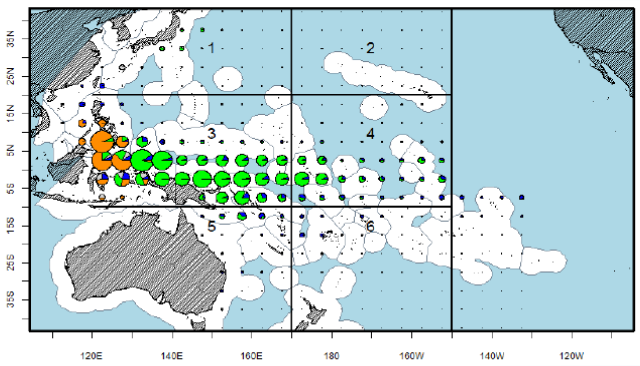
Working example

- Cut-down data sets for bigeye and yellowfin for the 6 region model, a reduced model time period (1990-2009)

Bigeye



Yellowfin



2 species
6 (12) regions
28 age classes
24 (48) fisheries
Quarterly time steps
72 time periods

- No production multi-species example (i.e. for a stock assessment) has yet been developed so the following test using the cut-down example is also the demonstration of its implementation in MULTIFAN-CL.

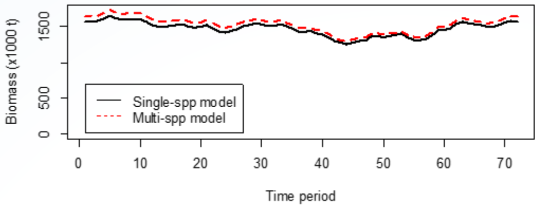
Comparisons

Deterministic – identical

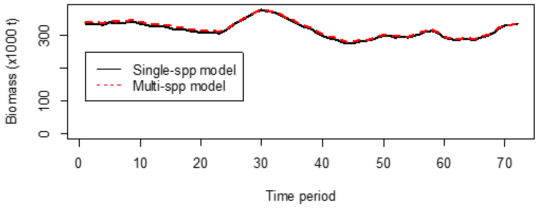
Fitted solutions – negligible differences

Biomass

Single species vs Multi-species Total Biomass: species 1

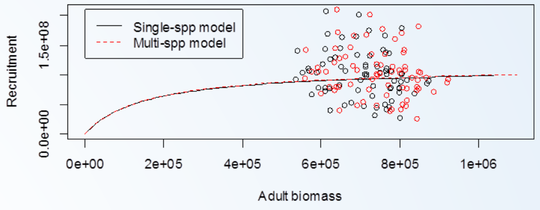


Single species vs Multi-species Total Biomass: species 2

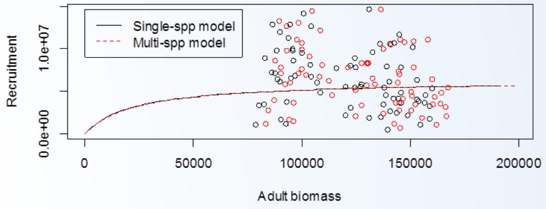


BH-SRR

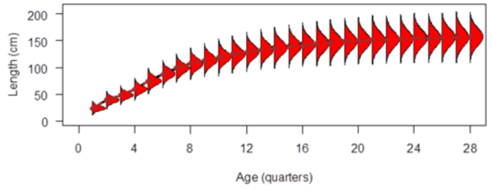
SRR curve and pred.Recrs for single/multi-spp models: species 1

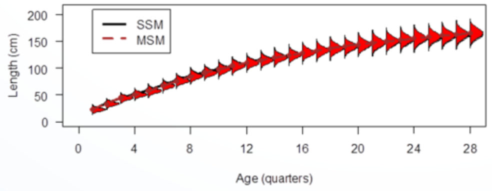


SRR curve and pred.Recrs for single/multi-spp models: species 2



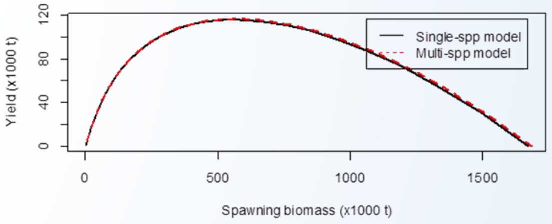
Growth



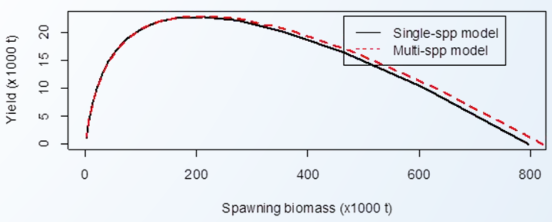


Yield

Single species vs Multi-species Equilibrium Yield: species 1



Single species vs Multi-species Equilibrium Yield: species 2



Stock partition - example

- Example using real tuna fishery data
- Work in progress, results are only preliminary – the fishery definitions and parameterization; the use of tagging data are still in development
- Proof-of-concept demonstration only
- Bigeye tuna data – WCPO and EPO (based on 2015 Pacific-wide model)
- Configuration is a vast simplification of the original single-stock models

Multi-stock example

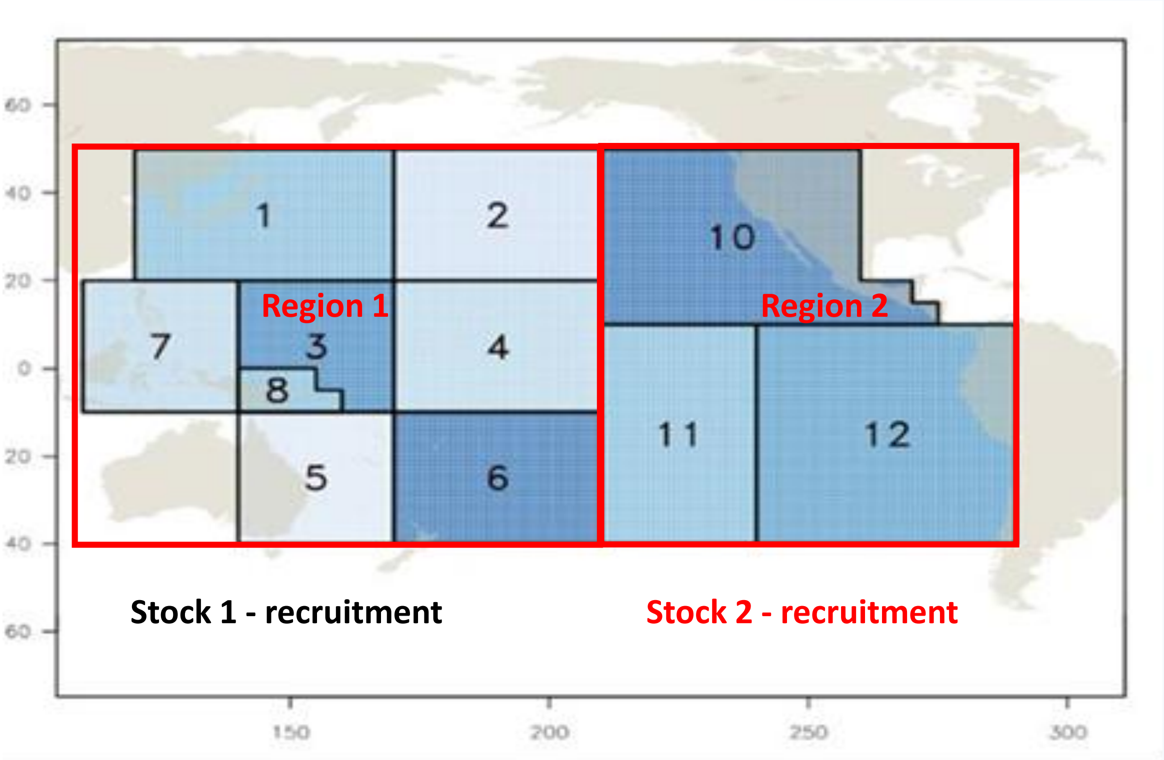
Bigeye tuna – WCPO and EPO

Simplified to two regions:
- Region 1 – 1 to 9
- Region 2 – 10 to 12

Single recruitment regions for each stock
- Stock 1 – region 1
- Stock 2 – region 2

Fisheries:
- region 1 – 33; data aggregated for 6
- region 2- 7; data aggregated for 4

Japanese longline CPUE indices
Tagging data:
- Stock 1 – 58 release groups
- Stock 2 – 12 release groups
- Assumed to be disaggregated



Parameters – specific or shared

Biological:

- **stock-specific** growth (retained with movement among regions)
- recruitment of each stock occurs in its **defined region**
- movement occurs between the regions (populations in a region are a mix of both stocks)

Fisheries:

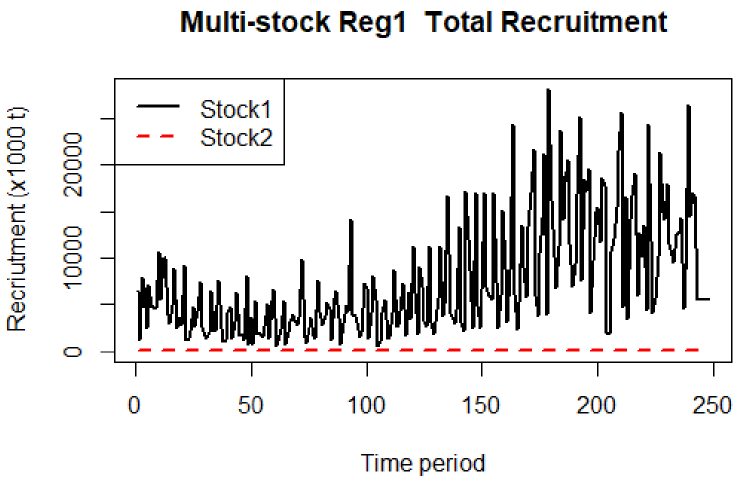
- Fisheries act upon the mix of both stocks within each region subject to:
- shared selectivity and catchability

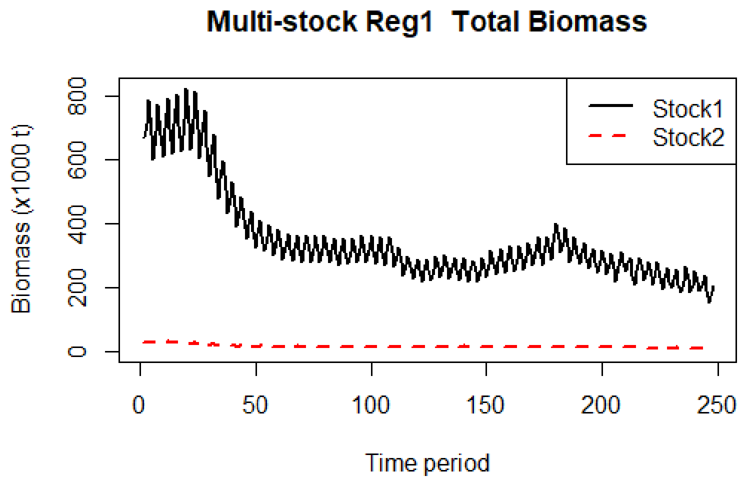
Example: multi-stocks

Recruitment

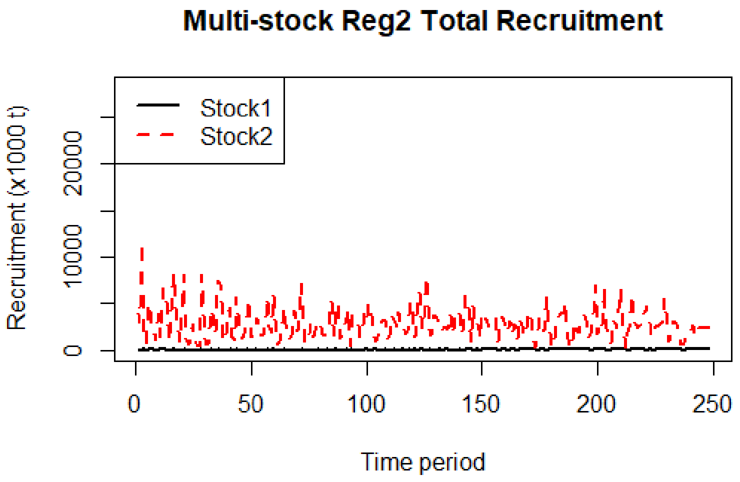
Biomass

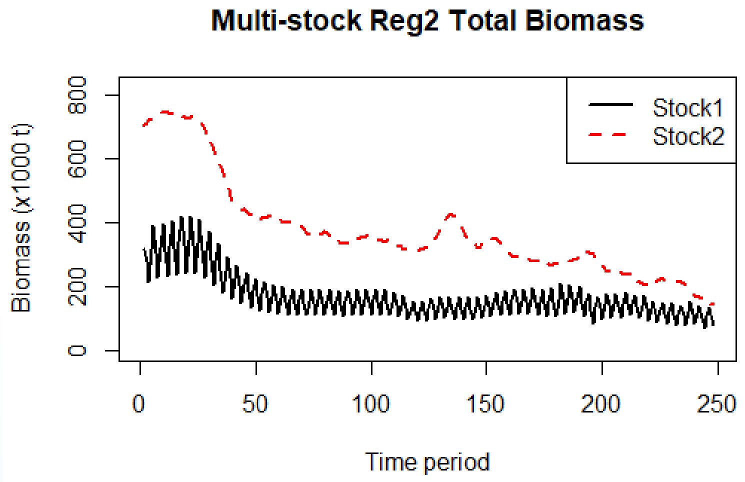
Region 1





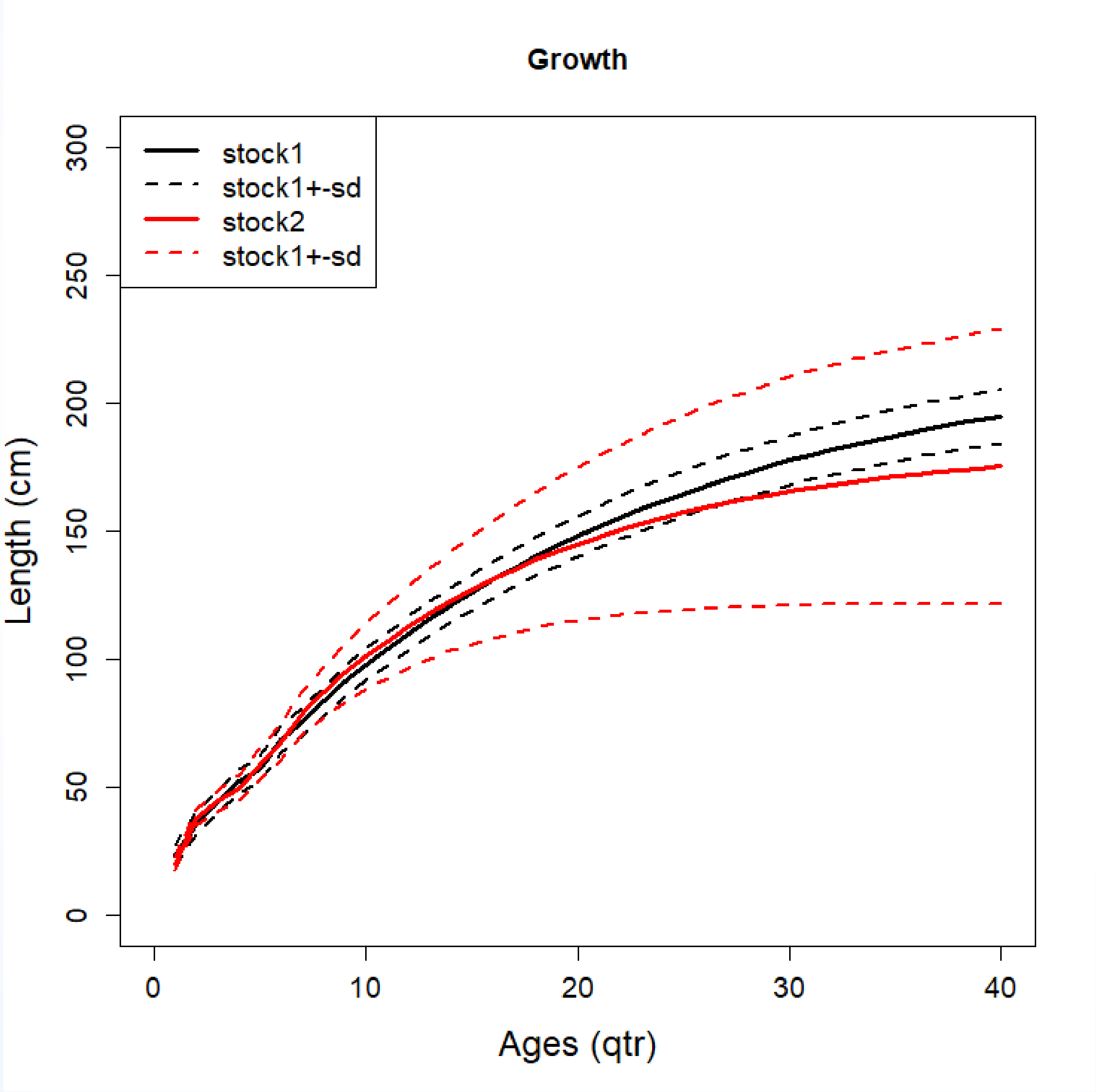
Region 2





Example: multi-stocks

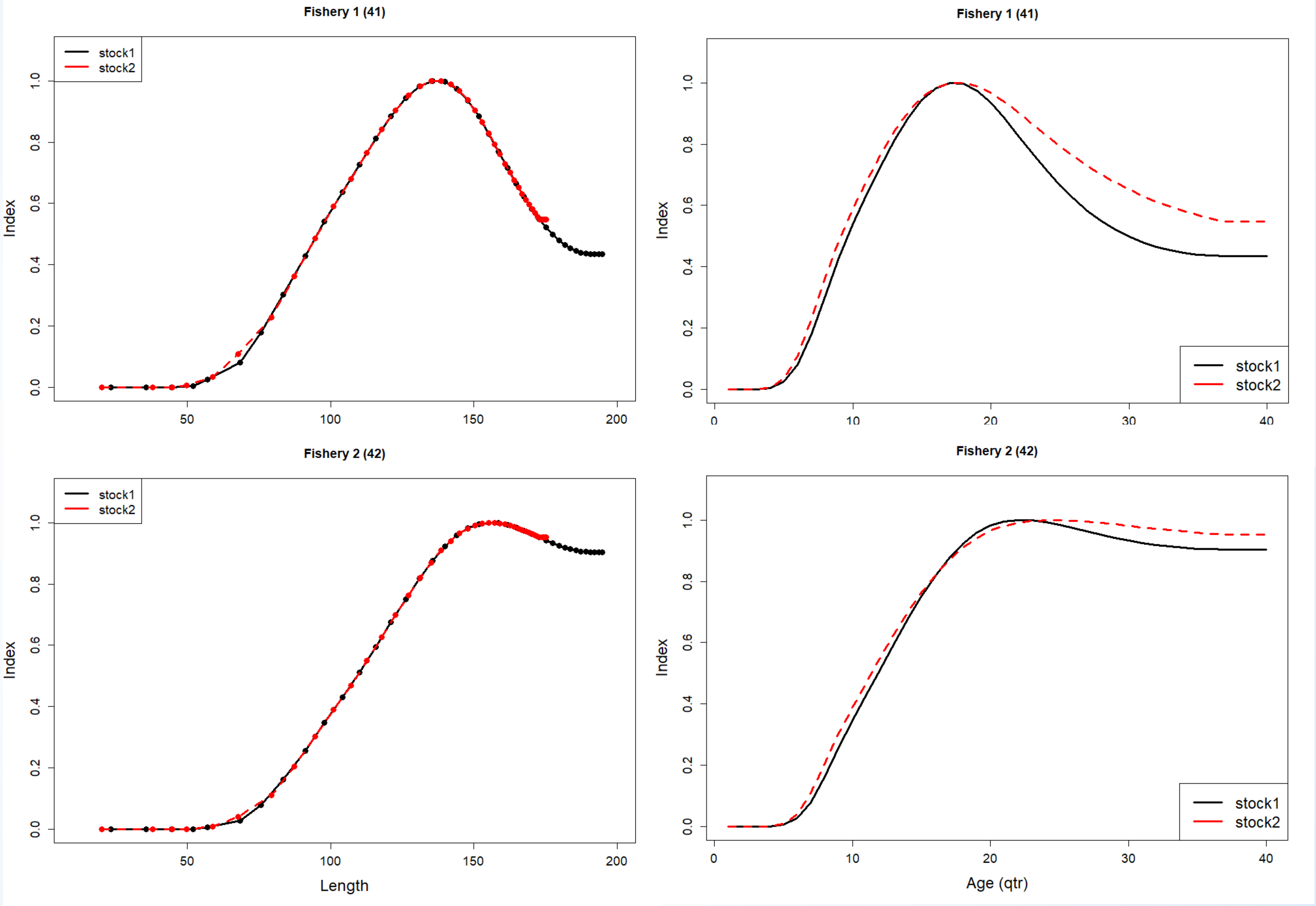
Growth



Example: multi-stocks

Selectivity

Fishery 1 (41)



Fishery 2 (42)

Multi-sex partition

Rationale

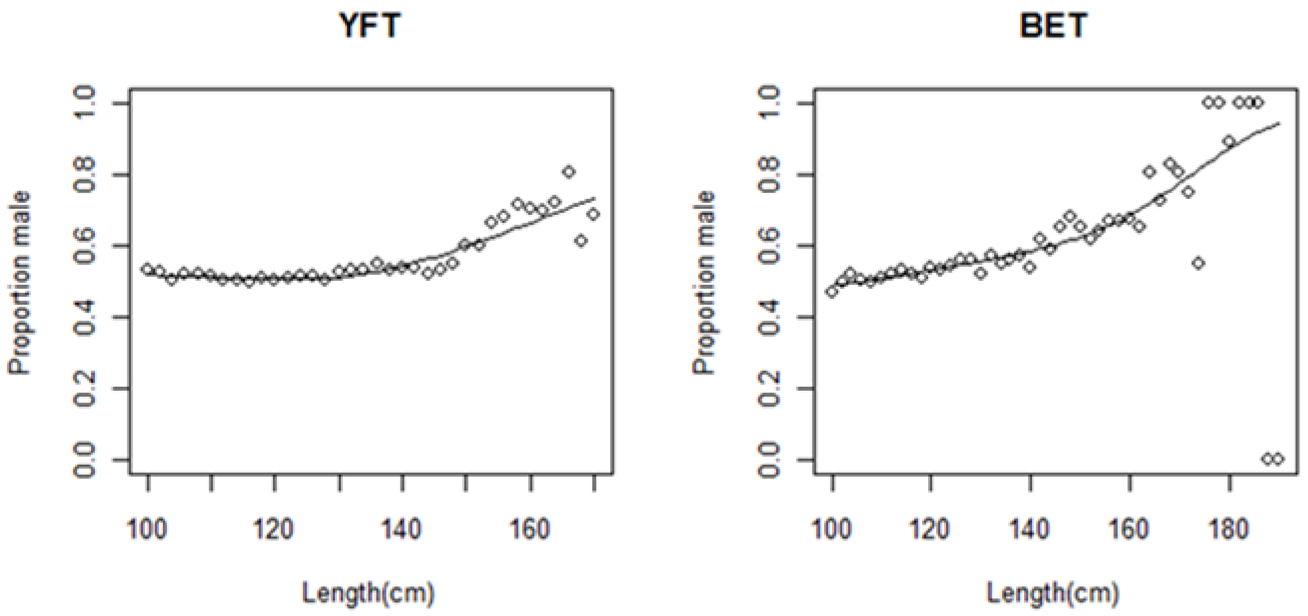
To account for sex-specific population characteristics:

- sex ratio heterogeneity by age/size and region
- growth
- natural mortality
- maturity

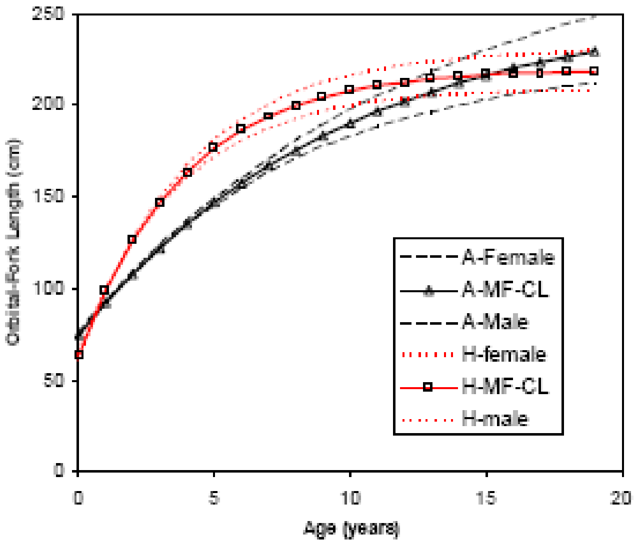
Share fishery mortality parameters (selectivity-at-length, catchability)

Explicitly model dynamics of female spawning potential

Sex ratio



Growth - SWO



Multi-sex parameterisation

- recruitments are shared (50:50%)
- BH-SRR – only female spawning biomass
- increased parameter sharing, e.g. length-based selectivity

When **disaggregated data** is not available…

…provision for shared effort deviate parameters among sexes

Fishing mortality:

$$F_{atf} = s_{af}\, q_{tf}\, B_{rf}^{\beta}\, E_{tf}^{\zeta}\, \boxed{e^{\varepsilon_{tf}}}$$

s_{af}	is the selectivity coefficient of fishery ***f*** for age-class ***a*** fish,
q_{tf}	is the catchability coefficient for fishery ***f*** in time period ***t***,
B	is a biomass index for region ***r*** and time period ***t***,
β_{tr}	is the parameter for effect of biomass on catchability (default= 0),
E_{tf}	is the fishing effort of fishery ***f*** in time period ***t***, and
ζ	is the parameter for effect of effort on catchability (default= 1),
ε_{tf}	represents **transient deviations in effort**

Shared effort deviates for multi-sex model
- disaggregated fisheries data unavailable

The fisheries are mirrored for sex 1 and sex 2

So… multi-sex model has fishing mortalities for each sex:

- without dis-aggregated data, there are no observed sex-ratios in fishery data
- with average catchability $q_{t\,f}$ being shared among sexes
- minimisation instability can occur and strongly positive or negative effort deviates ($\varepsilon_{t\,f}$) may be estimated if unconstrained
- this can be stabilised by assuming the $\varepsilon_{t\,f}$ are also shared

Sex-partition - development and testing

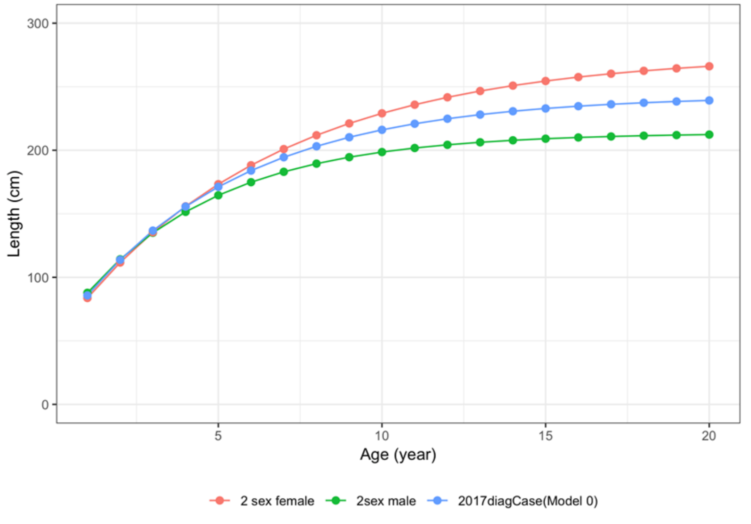
- Applied same approach as for multi-species
- Adapted the example: two identical species; configured for multiple sexes
- Criterion: BH-SRR recruitments, Equilibrium yield, F_{MSY} == double that of the single species
- Was correct in deterministic and fitted comparisons

Examples

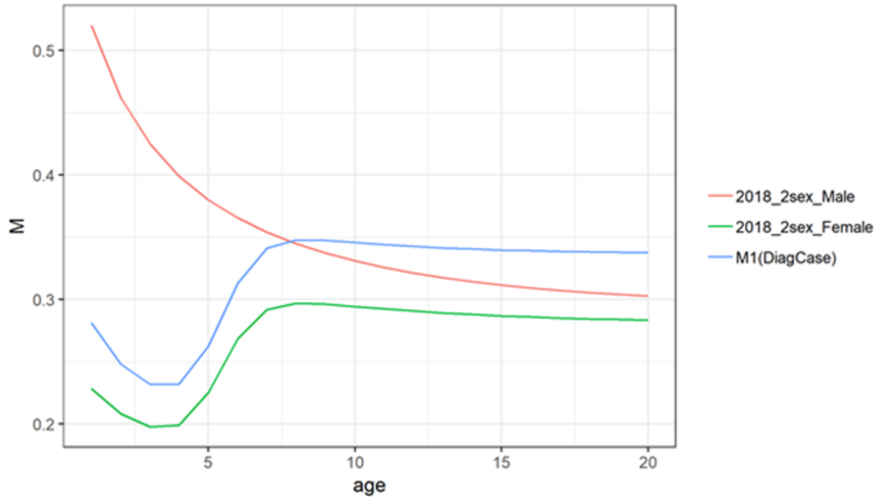
- Takeuchi et al. 2018 – south-west Pacific swordfish, WCPFC-SC14-2018/SA-IP-10 (presented here)
- Ducharme-Barth et al. – south-west Pacific striped marlin, WCPFC-SC15-2019/SA-WP-07

Swordfish

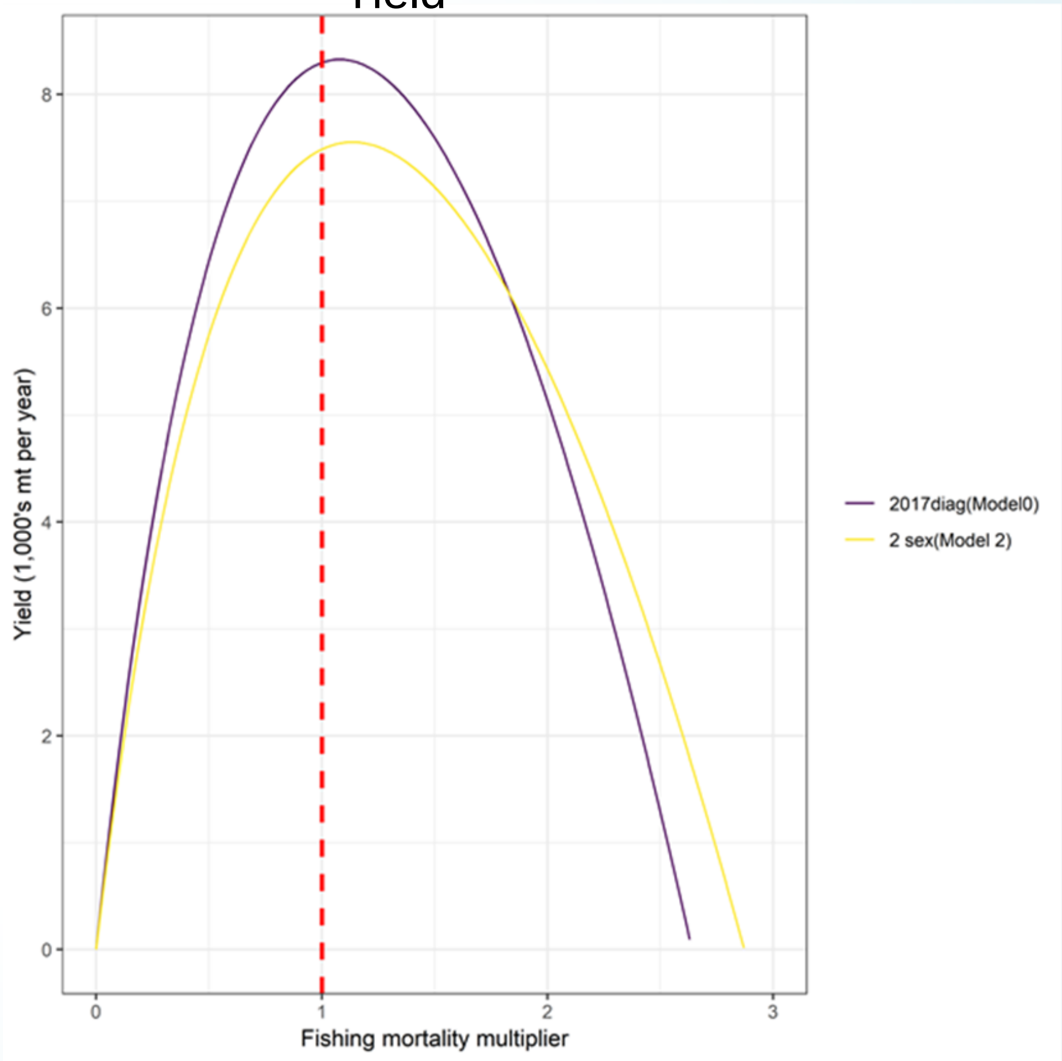
Growth



Natural mortality

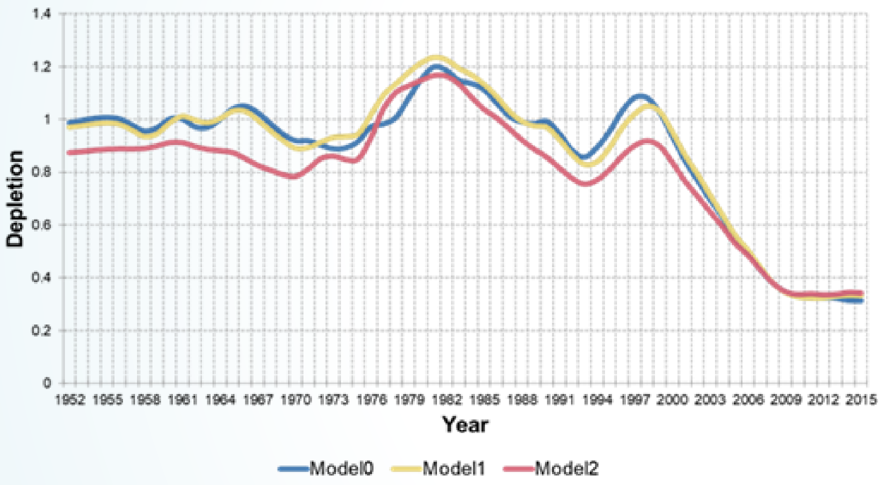


Yield

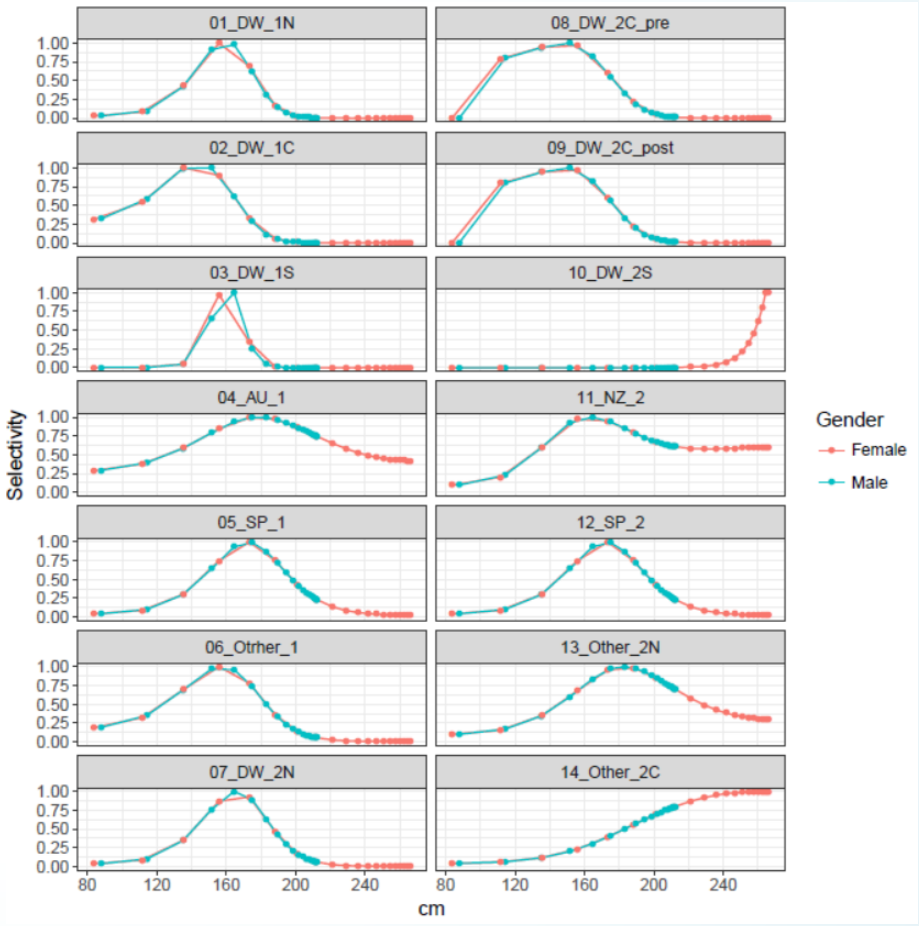


Swordfish

Biomass



Selectivity at length



Model0 (blue line) – single sex (blue)
Model1 (yellow line) – multi-sex; params set equal
Model2 (red line) – multi-sex

Summary of species/stock/sex partitions

- Partitioning has not been achieved using explicit data structures
- Region partition adaptation provided a ready means that was consistent with the existing fisheries and tagging structures
- Development code testing was positive
- Working examples have been developed for each partition, but not yet applied for a production assessment
- While it may, or may not, be feasible, no working example using 2 or 3 of the partitions in a single model has been attempted
- To date, no more that 2 species/stocks/sexes partitions have been attempted
- No interaction processes among partitions have yet been implemented

5. Software development and testing – source code testing and repository

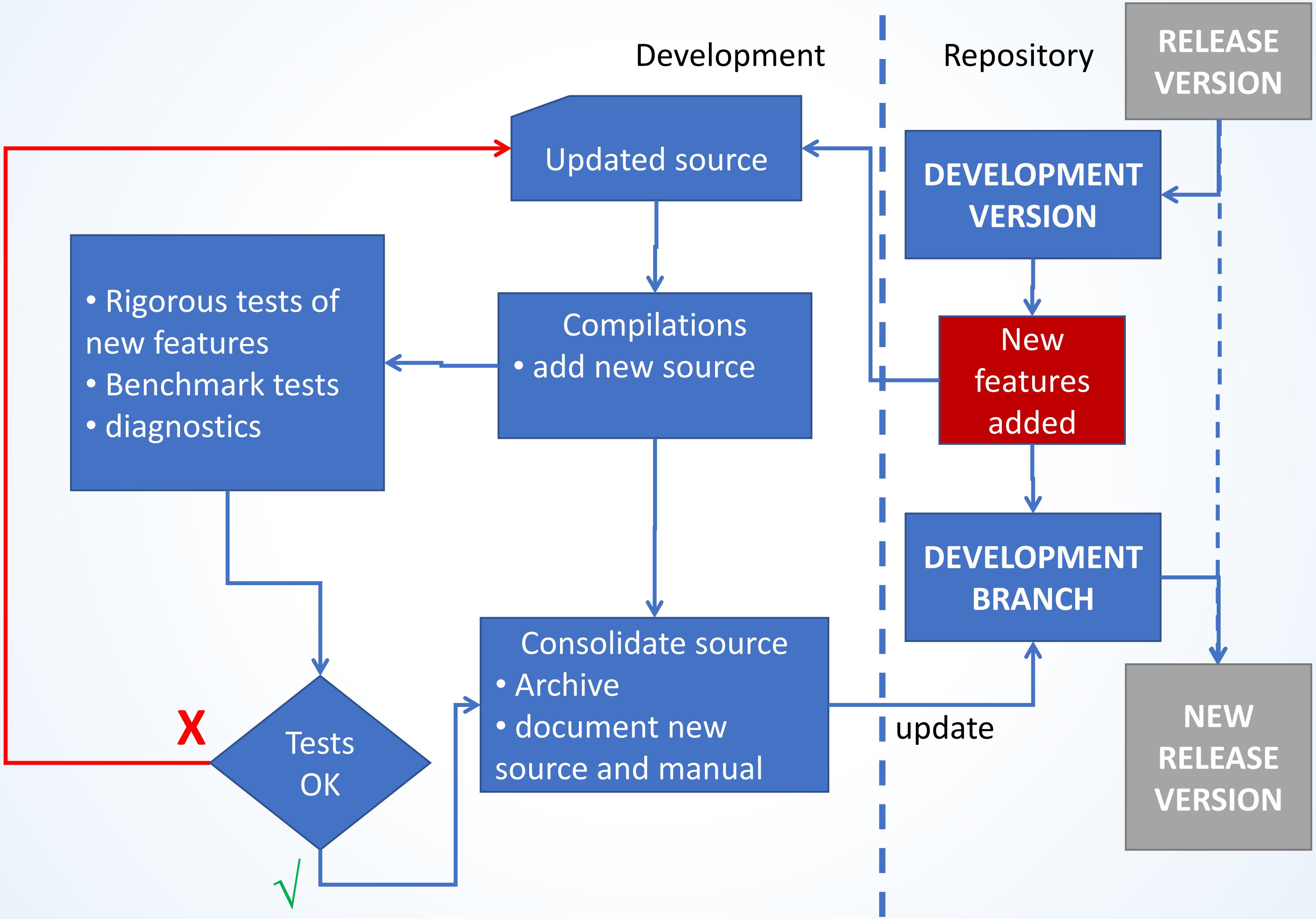
Formal testing procedure for ensuring integrity of source code developments has two stages:

1. Detailed testing of a new feature
    - Specific example used employing the feature

2. “Benchmark” testing
    - Routine tests - Pair-wise comparisons among development and release versions prior to advancing a new release version
    - Uses test data: single-species, multi-species/sex, projections, stochastic simulations

Github repository branches - development and release versions

Schema for MFCL development and testing



6. Improvements possible and development recommendations

1. Data structure

With the benefit of hindsight…

- The first rule: "establish the correct data structures to begin with, and then the code will write itself"
- The region-adaptation appears to work for species/stock/sex partition but…
  - Proved to be a complex coding task
  - Is restricted by the parameterisation requirements using the mirrored fisheries
  - May become difficult when using more than one partition
- Would have been better to use the appropriate data structure

Improvements possible and development recommendations

Fully dimensioned data structure:

N(species, stock, sex, region, growth_morph, time_period, age, length)

Reduces to a current model population data structure:

N(1, 1, 1, region, 1, time-period, age,1)

Coding challenges to avoid inefficiencies in memory management due to a large multi-dimensional array.

Improvements possible and development recommendations

2. Length-structure

N(species, stock, sex, region, growth_morph, time-period, age, length)

Note included additional partitions for:

- **length**
- **growth morphs**

To explicitly model region-specific growth rates in a multi-stock model with movement among regions

To explicitly model individual growth variability

Improvements possible and development recommendations

3. Multi-threading

Motives for increasing model complexity using partitions include:

- Biological processes – the need to better describe real processes that entail partitions, e.g. movement, growth, sub-population processes
- Fishery and management specifications – spatially-based regimes

Increased computational requirements, e.g. multi-region tuna models require up to 24 hours to converge

Multi-threading has been trialed for tagging calculations

TMB offers a ready means to overcome this computational challenge

Improvements possible and development recommendations

4. Development with examples

Development of complex multi-partitioned models is best done with:

- – Example model data expressing all partitions
- – Pair-wise testing versus an existing “accepted” version

Existing multi-partitioned software (e.g. MULTIFAN-CL) should be used to generate these example data sets in simulation mode and for undertaking the development and testing

Thanks!