Modeling Selectivity In Stock Synthesis (SS)

Richard Methot, Senior Scientist for Assessments

Webinar

March 7, 2019

Resources

- Stock Synthesis – VLAB Link:
  https://vlab.ncep.noaa.gov/web/stock-synthesis/home
- 2013 CAPAM Workshop on Selectivity: http://www.capamresearch.org/selectivity/workshop
- Special Issue on Selectivity (Fisheries Research) resulting from CAPAM's Selectivity Workshop. Fisheries Research, Volume 158, Pages 1-204 (October 2014). Edited by M.N. Maunder, P.R. Crone, J.L. Valero and B.X. Semmens. https://www.sciencedirect.com/journal/fisheries-research/vol/158/suppl/C

CAPAM Workshop - 2013

- A-comparison-of-parametric-semi-parametric-and-non-parametric-approaches-to-selectivity-in-age-structured-assessment-models_2014_Fisheries-Research.pdf
- A-historical-review-of-selectivity-approaches-and-retrospective-patterns-in-the-Pacific-halibut-stock-assessment_2014_Fisheries-Research.pdf
- A-proposed-tested-and-applied-adjustment-to-account-for-bias-in-growth-parameter-estimates-due-to-selectivity_2014_Fisheries-Research.pdf
- Data-conflict-caused-by-model-mis-specification-of-selectivity-in-an-integrated-stock-assessment-model-and-its-potential-effects-on-stock-status-estim.pdf
- Direct-calculation-of-relative-fishery-and-survey-selectivities_2014_Fisheries-Research.pdf
- Estimation-of-time-varying-selectivity-in-stock-assessments-using-state-space-models_2014_Fisheries-Research.pdf
- Evaluation-of-length-vs-age-composition-data-and-associated-selectivity-assumptions-used-in-stock-assessments-based-on-robustness-of-derived-managemen.pdf
- Evaluation-of-the-sensitivity-of-biological-reference-points-to-the-spatio-temporal-distribution-of-fishing-effort-when-seasonal-migrations-are-sex-sp.pdf
- Evaluation-of-virgin-recruitment-profiling-as-a-diagnostic-for-selectivity-curve-structure-in-integrated-stock-assessment-models_2014_Fisheries-Resear.pdf
- Fishery-selection-and-its-relevance-to-stock-assessment-and-fishery-management_2014_Fisheries-Research.pdf
- Full-Title-Page-Editorial-Board-_2014_Fisheries-Research.pdf
- Investigating-the-influence-of-length-frequency-data-on-the-stock-assessment-of-Indian-Ocean-bigeye-tuna_2014_Fisheries-Research.pdf
- Model-selection-for-selectivity-in-fisheries-stock-assessments_2014_Fisheries-Research.pdf
- Selectivity-Theory-estimation-and-application-in-fishery-stock-assessment-models_2014_Fisheries-Research.pdf
- The-ability-of-two-age-composition-error-distributions-to-estimate-selectivity-and-spawning-stock-biomass-in-simulated-stock-assessments_2014_Fisherie.pdf
- Towards-defining-good-practices-for-modeling-time-varying-selectivity_2014_Fisheries-Research.pdf
- Tradeoffs-between-bias-robustness-and-common-sense-when-choosing-selectivity-forms_2014_Fisheries-Research.pdf
- Use-of-likelihood-profiling-over-a-global-scaling-parameter-to-structure-the-population-dynamics-model-An-example-using-blue-marlin-in-the-Pacific-Oce.pdf
- Use-of-multiple-selectivity-patterns-as-a-proxy-for-spatial-structure_2014_Fisheries-Research.pdf
- Using-areas-as-fleets-selectivity-to-model-spatial-fishing-Asymptotic-curves-are-unlikely-under-equilibrium-conditions_2014_Fisheries-Research.pdf

http://www.capamresearch.org/selectivity/workshop

Outline

- How selectivity works
- Selectivity parameterizations in SS
- Age vs. length selectivity
- Tradeoff: stiff vs. flexible forms
- Time-varying selectivity options
- Evolution to semi-parametric formulation
- Interaction with data-weighting

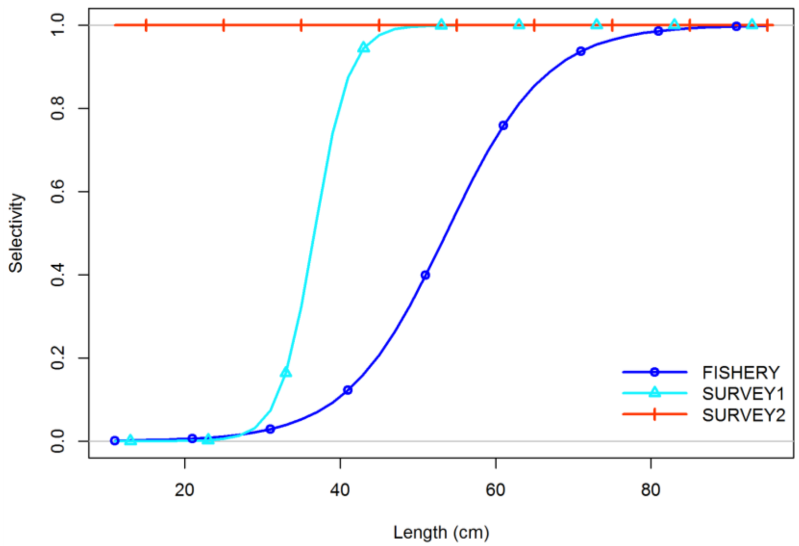
What is selectivity?

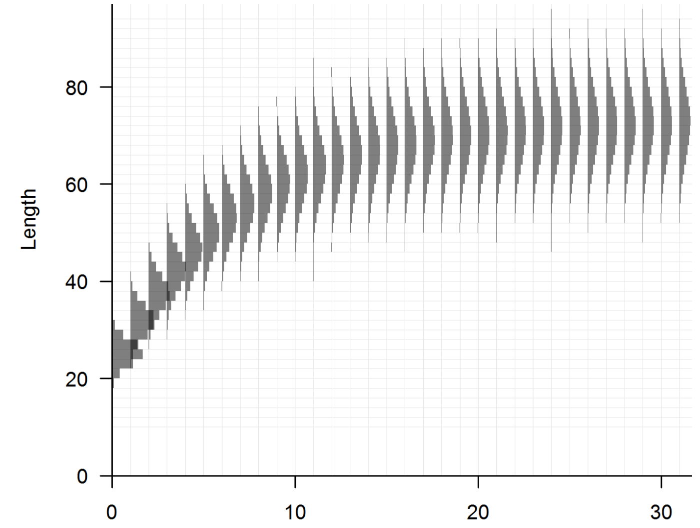
- Selectivity is a component of catchability-at-age
- Catchability is typically decomposed into a scaling factor, q or F, and vectors to distribute the q (F) across ages, sizes and sexes.

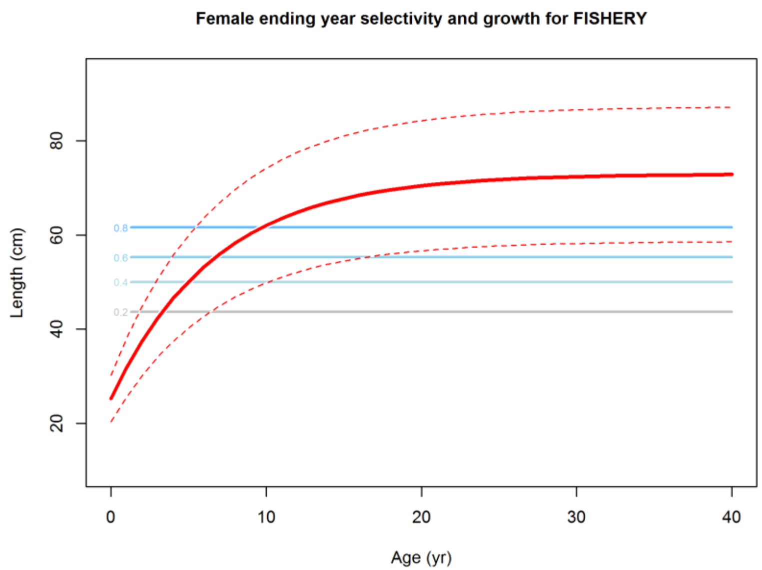
$$Z_{t,\gamma,a} = M_{\gamma,a} + \sum_{f=1}^{A_f} \left(S_{y,f,\gamma,a} F_{t,f} \right)$$

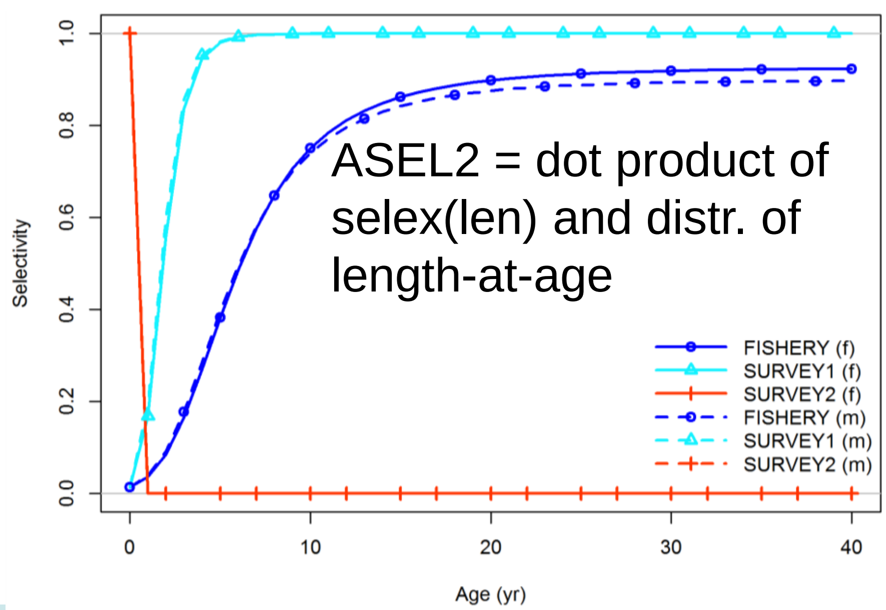
- Typically the age-specific factor is scaled to have a maximum value of 1.0, but it need not. All that matters is the product of selectivity and catchability.

Length Selectivity



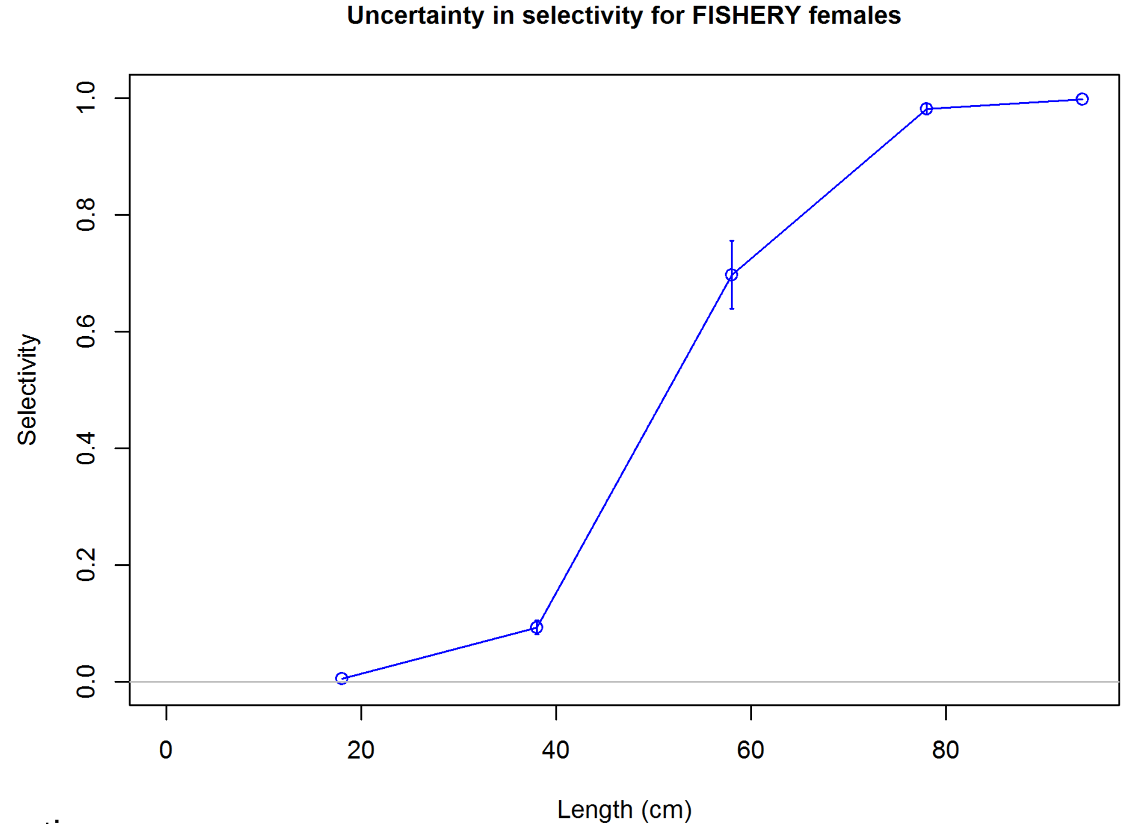






ASEL2 = dot product of selex(len) and distr. of length-at-age

Selectivity is a Derived Quantity

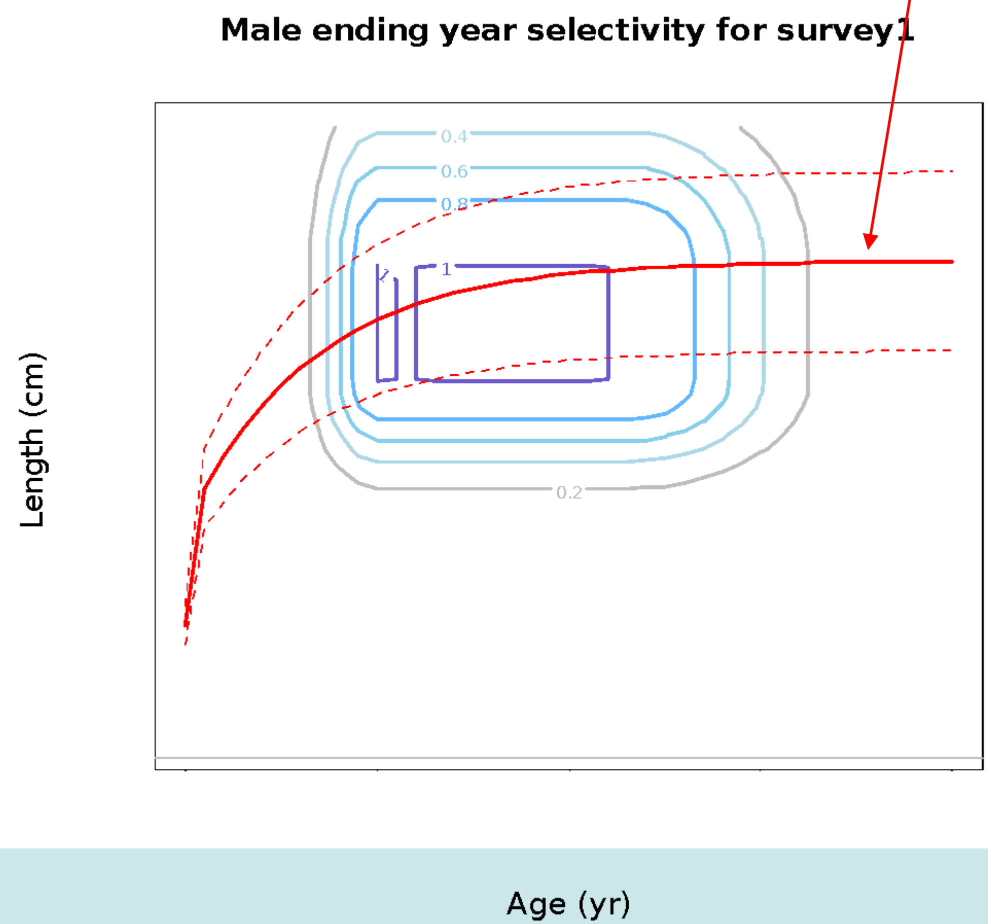


1 # (0/1) read specs for more stddev reporting
1 1 -1 5 1 5 1 -1 5 # selex_fleet, 1=len/2=age/3=both, year, N selex bins, [delete]
5 15 25 35 43 # vector with selex std bins (-1 in first bin to self-generate)
1 2 14 26 40 # vector with growth std ages picks (-1 in first bin to self-generate)
1 2 14 26 40 # vector with NatAge std ages (-1 in first bin to self-generate)

Combining selectivity at length at age

- Independent functions
- Example with two double normals:

growth curve



Compound Age Selectivity in SS – ASEL2

- Length selectivity calculated for mid point of each population length bin. So, broad pop length bins give coarse resolution for selectivity and for distribution of length-at-age. See Monnahan et al 2015. Fewer, broader bins is faster; but can have ragged convergence.
- Average age-selectivity due to length selectivity is calculated as dot product of selectivity-at-length and the distribution of length at each age. This can result in no age having selectivity equal to 1.0.
- Then multiple that preliminary age selectivity by any direct age selectivity to get the compound age selectivity, called ASEL2 in the SS output
- In calculation of catch, SS uses the numbers-at-age, the ASEL2, and the selected body weight at age to calculate age-specific catch numbers for each fleet and total catch in numbers and weight.

Internal SS: Using selectivity to generate expected composition

- In calculation of the expected value for any non-catch observation, a different but equivalent process is used to apply the selectivity
- Numbers-at-age, for each bio entity, g, (sex, growth morph, settlement, and platoon), are decayed by $e^{-Z*time}$ to the exact time of the observation;
- If length selectivity is involved, get length-at-age distribution for each entity at the subseason that most closely corresponds to the exact time of the observation. Call this ALK_g.
  - If growth is time-varying, this step is what slows SS down
  - If fish are growing rapidly through a time step in your model, you should set the number of subseasons to be greater than 2 to achieve more granularity in timing of ALK updating, and enter month values that match the timing of the observation (not just the typical mid-season timing).
- For each entity, multiple numbers-at-age by the ALK to get the set of age-length populations at time of the sample
- Apply length, sex, and age selectivity to each age-length pop, to get the sampled age-length for each fleet/survey with observations at the current subseas. If there is discard/retention partition, then get retained, total and discarded age-length samples at this step.
- Sum across ages to get the overall expected length composition for each entity
- For each sex, sum these compositions across settlements, morphs and platoons
- Similarly sum across lengths to get overall age composition, the apply age-age’ transition to convert it to an age’ composition. For conditional age’-at-length, do this for each observation-specified length range, rather than all lengths.
- For mean length-at-age’, do the needed calculations.

Overall Evolution of Selectivity in SS

- Logistic
- Domed
  - Double normal
  - Double normal with defined peak and plateau to better decouple ascending side from descending side. With various experimental permutations to control behavior in tails.
  - Logistic exponential
- Flexible
  - Cubic spline
  - Random walk across ages; also each age
  - With mean scaling, not max scaling
- Semi-parametric with autoregressive devs

Length Selectivity Options

- 0; parm=0; selex=1.0 for all sizes
- 1; parm=2; logistic; with 95% width specification
- 5; parm=2; mirror another size selex; PARMS pick the min-max bin to mirror
- 15; parm=0; mirror another age or length selex
- 6; parm=2+special; non-parm len selex
- 43; parm=2+special+2; like 6, with 2 additional param for scaling (average over bin range)
- 8; parm=8; New double logistic with smooth transitions and constant above Linf option
- 9; parm=6; simple 4-parm double logistic with starting length; parm 5 is first length; parm 6=1 does desc as offset
- 21; parm=2+special; non-parm len selex, read as pairs of size, then selex
- 22; parm=4; double_normal as in CASAL
- 23; parm=6; double_normal where final value is directly equal to sp(6) so can be >1.0
- 24; parm=6; double_normal with sel(minL) and sel(maxL), using joiners
- 25; parm=3; exponential-logistic in size
- 27; parm=3+special; cubic spline
- 42; parm=2+special+3; // like 27, with 2 additional param for scaling (average over bin range)

Age Selectivity Options

- 0; parm=0; selex=1.0 for ages 0 to maxage
- 10; parm=0; selex=1.0 for ages 1 to maxage
- 11; parm=2; selex=1.0 for specified min-max age
- 12; parm=2; age logistic
- 13; parm=8; age double logistic
- 14; parm=nages+1; age empirical
- 15; parm=0; mirror another age or length selex
- 16; parm=2; Coleraine - Gaussian
- 17; parm=nages+1; empirical as random walk; N parameters to read can be overridden by setting special to non-zero
- 41; parm=2+nages+1; // like 17, with 2 additional param for scaling (average over bin range)
- 18; parm=8; double logistic - smooth transition
- 19; parm=6; simple 4-parm double logistic with starting age
- 20; parm=6; double_normal,using joiners
- 26; parm=3; exponential-logistic in age
- 27; parm=3+special; cubic spline in age
- 42; parm=2+special+3; // cubic spline; with 2 additional param for scaling (average over bin range)

Age Availability

- Age composition of the stock is not uniform geographically, so where fishing happens affects the overall selectivity
- Sampson, D.B., Scott, R.D., 2011. A spatial model for fishery age-selection at the population level. Can. J. Fish. Aquat. Sci. 68, 1077–1086.
- Easy to generate dome-selectivity from this phenomenon
- Also expect time-varying degree of dome due to fishing shifting location from historical areas with large fish to other areas with higher CPUE of smaller fish as large fish decline.

Age vs. Length Selectivity Considerations

- Generally contact selectivity is considered length-based (think gill nets) and
- availability is considered age-based either due to age-based movement between areas or as effect of regionally localized recruitment
- Both can be active simultaneously in reality and in SS

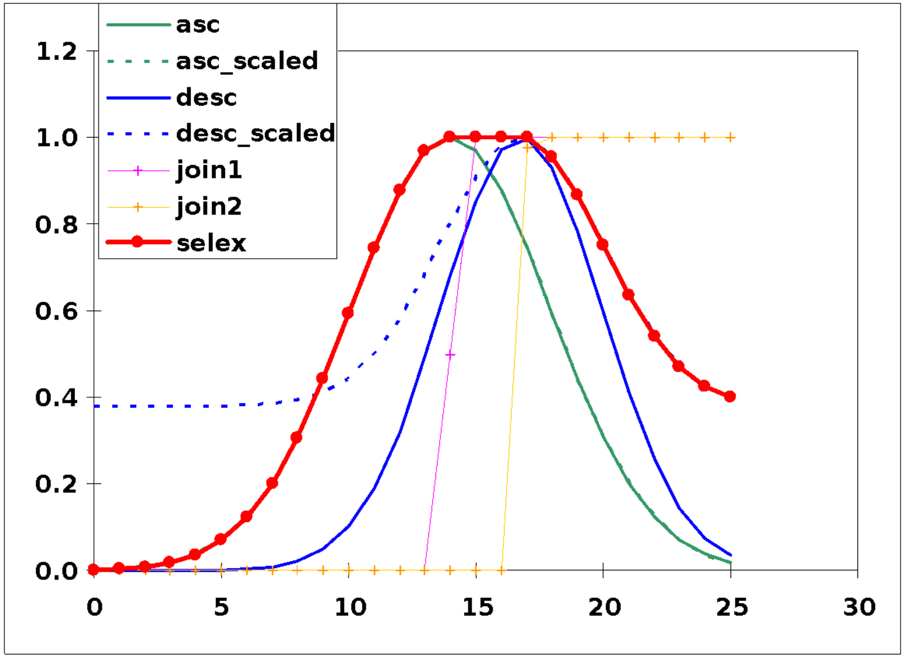
Age vs. Length: Situation

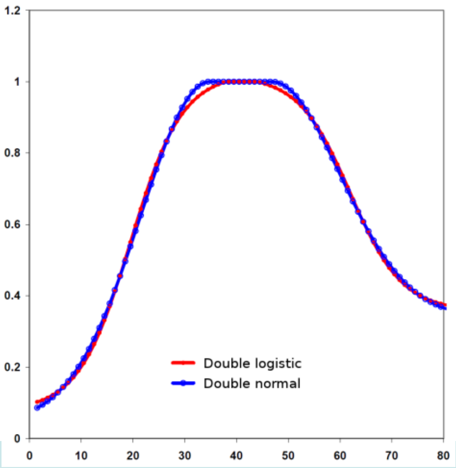
- Careful with using only length selectivity when doing fleets as areas, or generally ignoring possibility that fish of a given length are not sampled randomly for age.
  - Fishery 1 operates where small fish predominate; fishery 2 operates where large fish predominate
  - Movement of fish from area 1 to area 2 is predominately age based
  - Assessment model analyzes using areas-as-fleets, not explicit areas, and uses length-selectivity.
  - But delta between fleets is due to age availability, not length contact selectivity, so estimated growth is biased
- Bias in estimates of growth when selectivity in models includes effects of gear and availability of fish. Kevin R. Piner, Hui-Hua Lee, Lennon R. Thomas. 2017.
- Effects of age-based movement on the estimation of growth assuming random-at-age or random-at-length data
- H. H. Lee, L. R. Thomas, K. R. Piner and M. N. Maunder. 2017.

Double normal

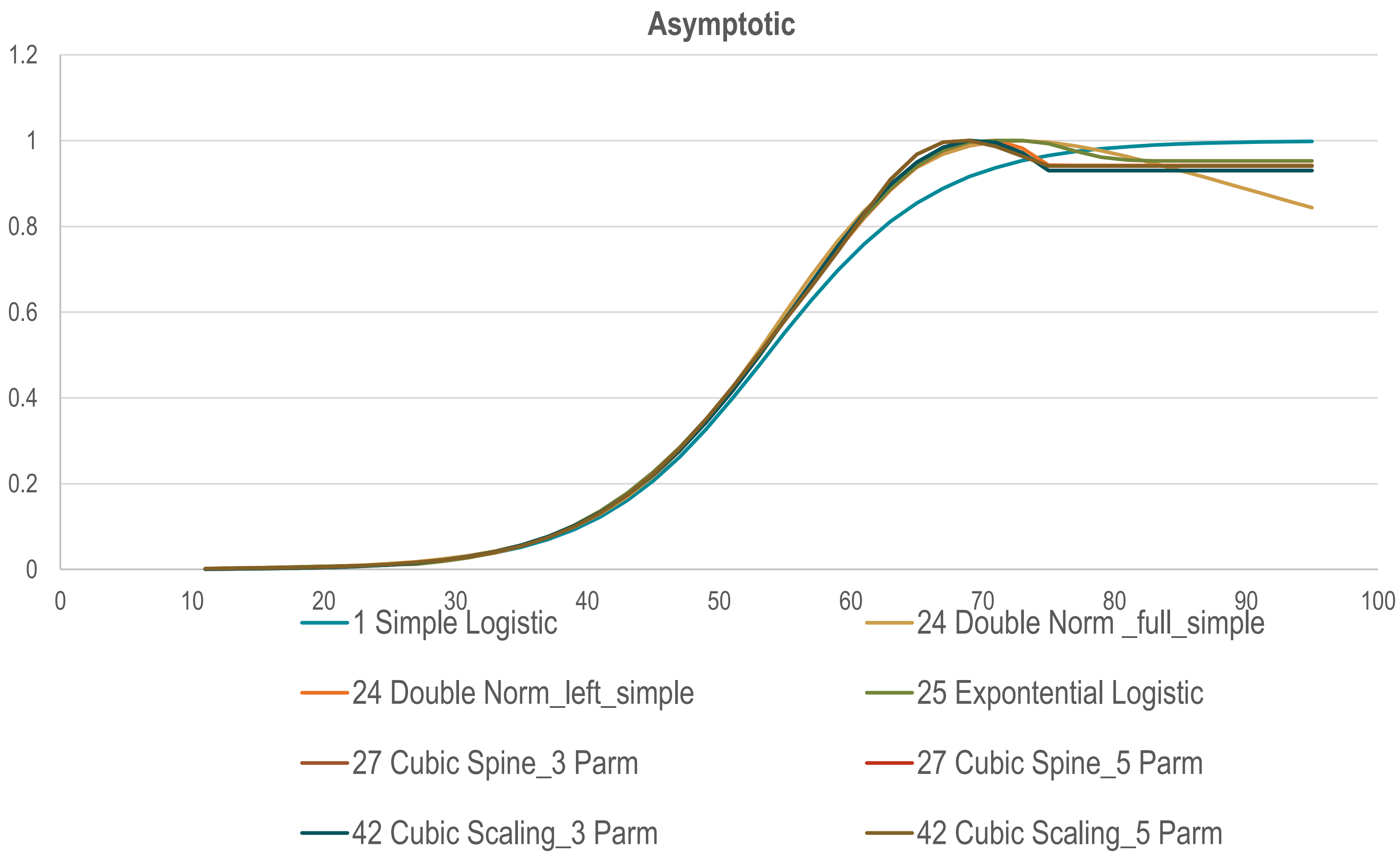
Composed of:

- left-side ascending normal scaled from the initial value to 1.0;
- flat plateau with 2 parameters controlling its location and width;
- right-side descending normal scaled 1.0 to the final value;
- Adjoining segments joined by very steep logistic functions to avoid IF statements and maintain differentiability
- Can be made asymptotic by wide plateau and not estimating desc, or by narrow plateau and nil desc. NOT BOTH!
- Options to control behavior for smallest and largest ages/sizes



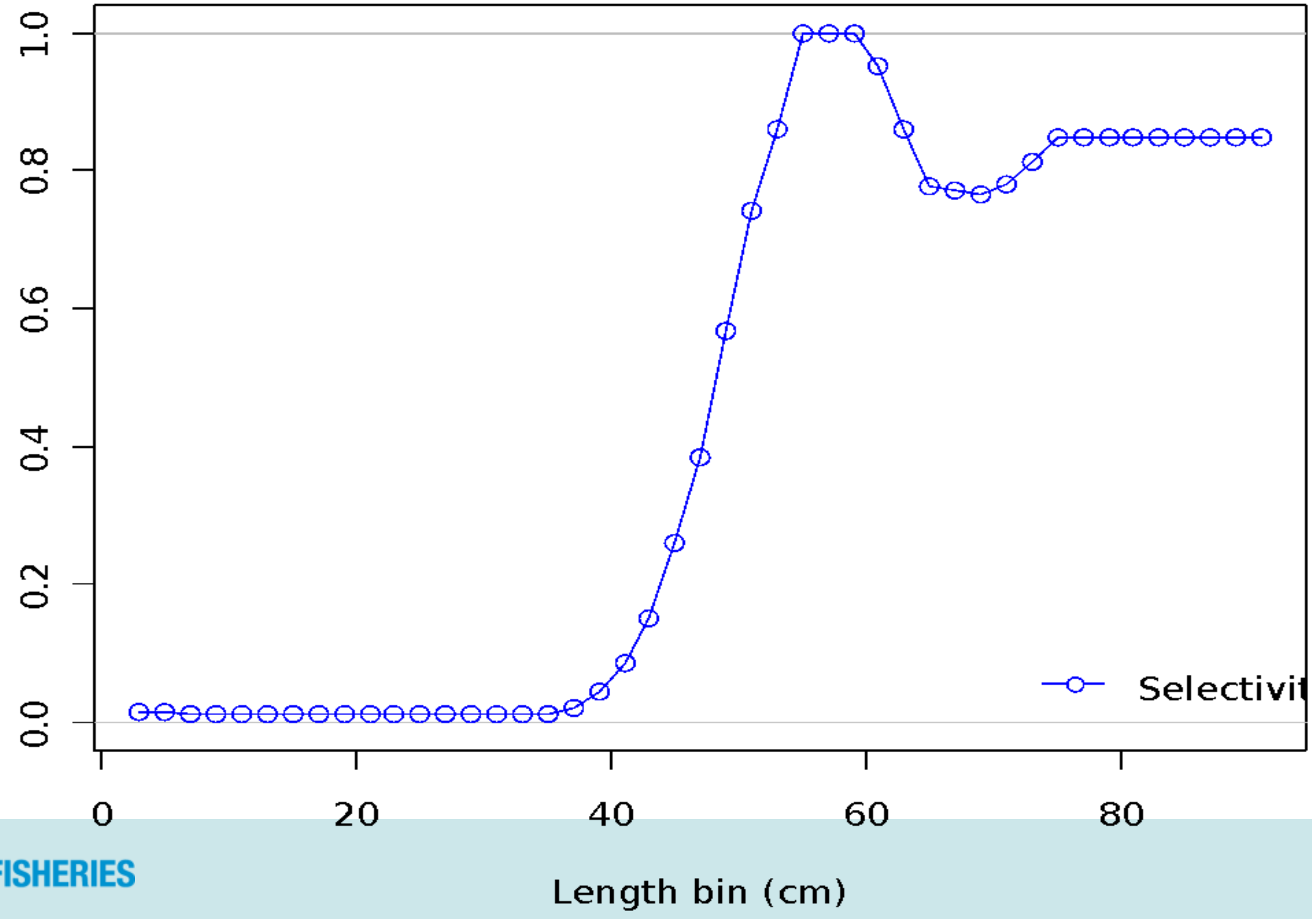


Fits to Data Generated from Asymptotic OM



Piecewise, log-linear length selectivity (#6)

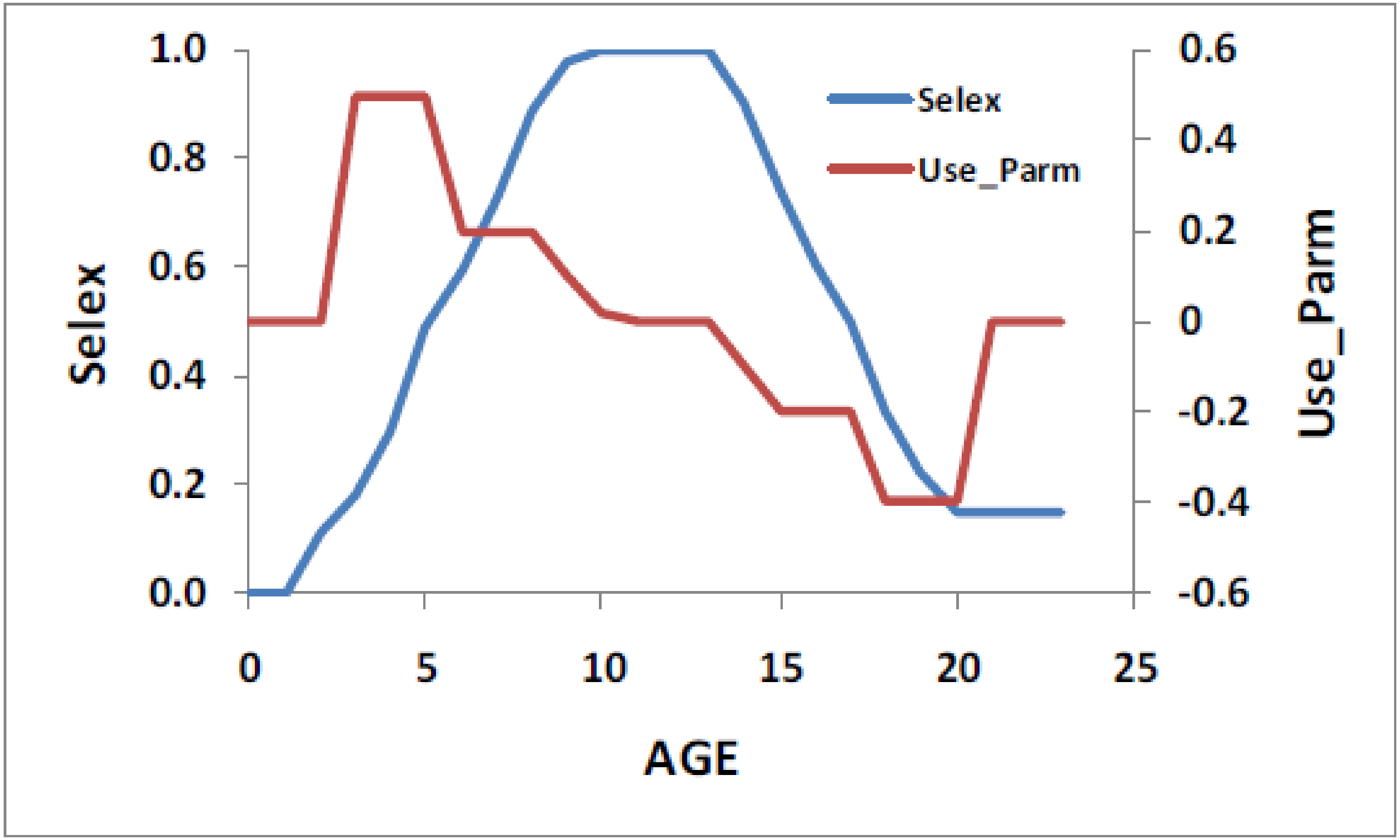
exp(interpolated value$_L$ – max(I.V.))



Pattern 17 (age)

- This selectivity pattern provides for a random walk in ln(selectivity).   In typical usage:
  - First parameter (for age 0) could have a value of -1000 so that the age 0 fish would get a selectivity of 0.0;
  - Second parameter (for age 1) could have a value of 0.0 and not be estimated, so age 1 is the reference age against which subsequent changes occur;
- Next parameters get estimated values. To assure that selectivity increases for the younger ages, the parameter min for these parameters could be set to 0.0 or a slightly negative value.
- If dome-shaped selectivity is expected, then the parameters for older ages could have a range with the parameter max set to 0.0 so they cannot increase further.
- To keep selectivity at a particular age the same as selectivity at the next younger age, set its parameter value to 0.0 and not estimated. This allows for all older ages to have the same selectivity.
- To keep a constant rate of change in selectivity across a range of ages, use the - 999 flag to keep the same rate of change in ln(selectivity) as for the previous age.

Random walk age selectivity



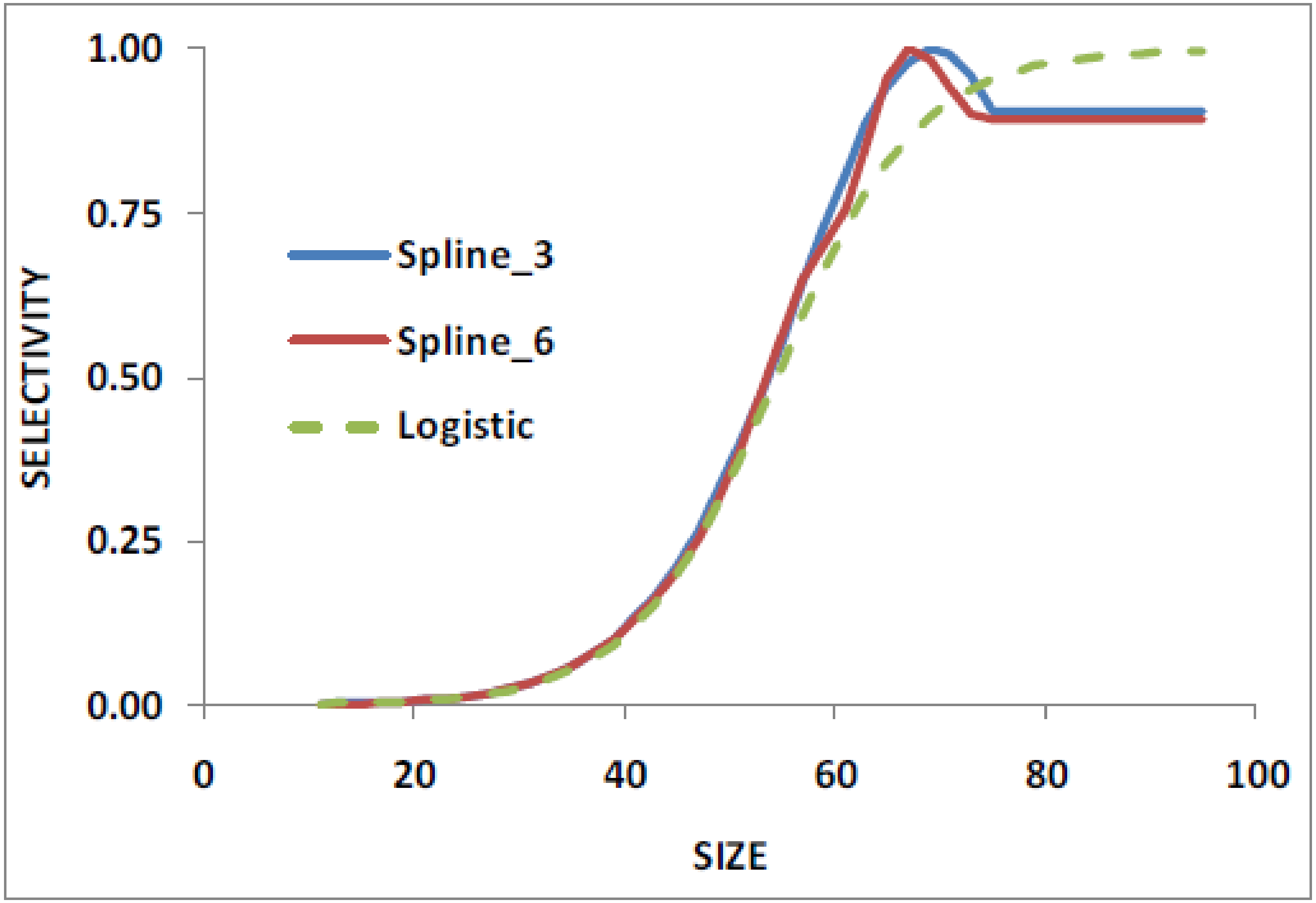
SS Input for Cubic Spline Selectivity, With Specified Scaling

- Specific number of knots (5)
- Request autogeneration (code=2) of knot locations and spline values
  - Autogen based on cumul. size distr. of all obs. for that fleet

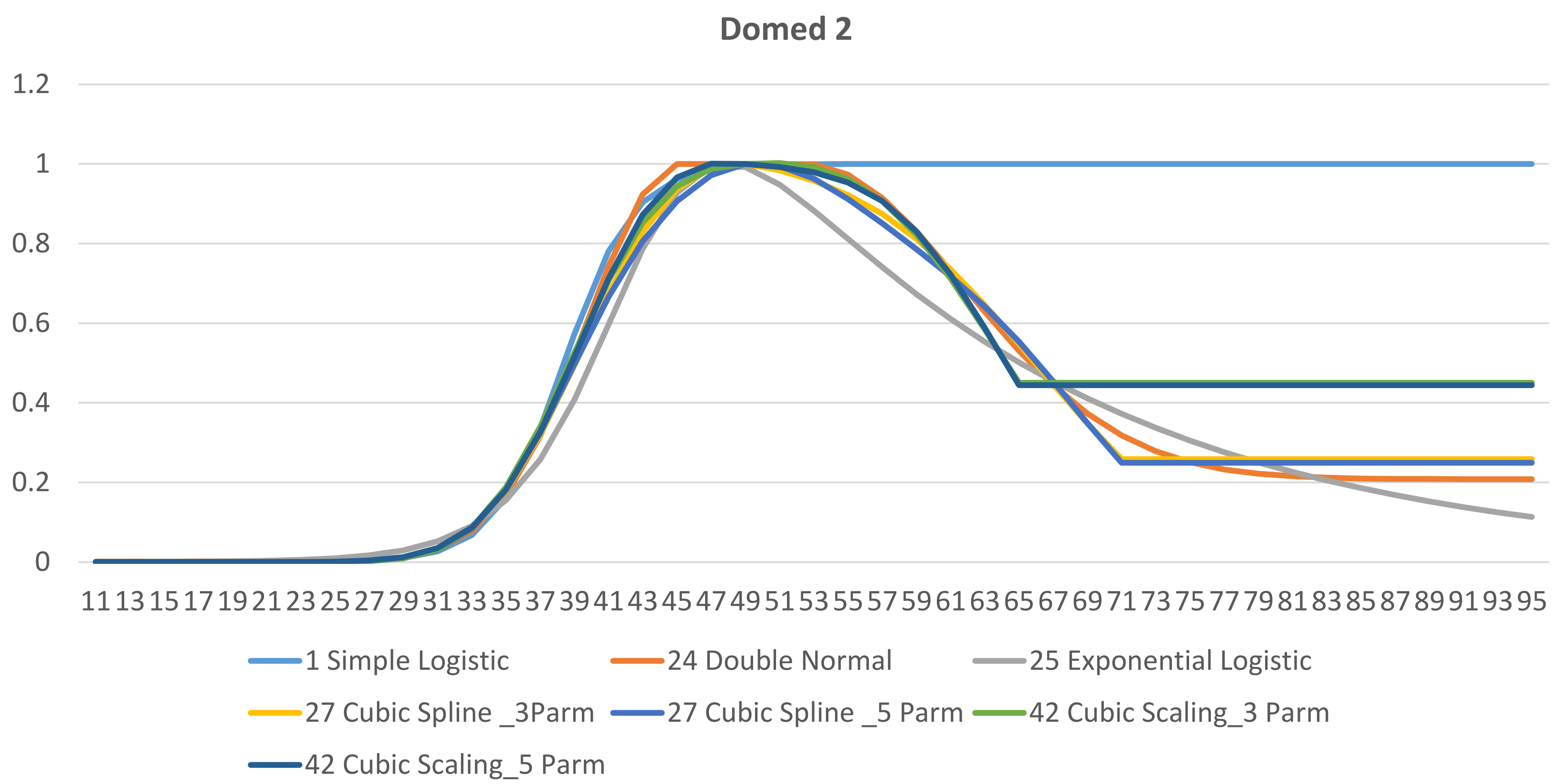
#_Pattern	Discard	Male	Special
42	0	0	5

LO	HI	INIT	# parm_name
0	100	30.00	# SizeSpline_ScaleBinLo_FISHERY(1)
0	100	30.00	# SizeSpline_ScaleBinHi_FISHERY(1)
0	2	2.00	# SizeSpline_Code_FISHERY(1)
-0.001	1	0.15	# SizeSpline_GradLo_FISHERY(1)
-1	0.001	-0.01	# SizeSpline_GradHi_FISHERY(1)
26	90	38.08	# SizeSpline_Knot_1_FISHERY(1)
26	90	52.14	# SizeSpline_Knot_2_FISHERY(1)
26	90	59.16	# SizeSpline_Knot_3_FISHERY(1)
26	90	63.74	# SizeSpline_Knot_4_FISHERY(1)
26	90	74.55	# SizeSpline_Knot_5_FISHERY(1)
-9	7	-3.15	# SizeSpline_Val_1_FISHERY(1)
-9	7	-1.47	# SizeSpline_Val_2_FISHERY(1)
-9	7	-1.00	# SizeSpline_Val_3_FISHERY(1)
-9	7	-0.78	# SizeSpline_Val_4_FISHERY(1)
-9	7	-0.77	# SizeSpline_Val_5_FISHERY(1)

Cubic Splines



Length Selectivity Fit to Domed Data



Estimated Variance of Selectivity

Length Selectivity at Selected Lengths										
	24	32	40	44	48	52	56	60	64	72
Logistic	0.00	0.07	0.78	0.96	0.99	1.00	1.00	1.00	1.00	1.00
Double Normal	0.00	0.08	0.74	1.00	1.00	1.00	0.91	0.73	0.53	0.28
Exponential Logistic	0.01	0.09	0.60	0.93	0.99	0.88	0.74	0.61	0.50	0.34
Cubic_3	0.00	0.09	0.67	0.94	1.00	0.96	0.87	0.74	0.55	0.26
Cubic_5	0.00	0.08	0.67	0.91	1.00	0.96	0.85	0.72	0.55	0.25
Cubic_3_scaled	0.01	0.26	1.59	2.35	2.68	2.54	2.15	1.73	1.35	0.67
Cubic_5_scaled	0.00	0.24	1.87	2.64	2.84	2.69	2.44	2.07	1.53	0.70

s.e. of Length Selectivity at Selected Lengths										
	24	32	40	44	48	52	56	60	64	72
Logistic	0.00	0.02	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Double Normal	0.00	0.02	0.07	0.01	0.00	0.01	0.05	0.06	0.05	0.04
Exponential Logistic	0.00	0.01	0.05	0.04	0.01	0.03	0.04	0.04	0.05	0.05
Cubic_3	0.00	0.02	0.04	0.03	0.00	0.03	0.05	0.06	0.06	0.04
Cubic_5	0.00	0.02	0.05	0.05	0.00	0.04	0.07	0.07	0.06	0.04
Cubic_3_scaled	0.00	0.05	0.22	0.31	0.34	0.31	0.23	0.14	0.06	0.06
Cubic_5_scaled	0.00	0.06	0.30	0.39	0.41	0.35	0.30	0.26	0.16	0.07

Note: Scaled cubic spline runs should have been scaled to 1.0 at 50cm not 70 cm

Time-varying parameter options

- #_Block types:
  - 0: Pblock=Pbase*exp(TVP)
  - 1: Pblock=Pbase+TVP
  - 2: Pblock=TVP
  - 3: Pblock=Pblock(-1) + TVP
- #_Block_trends:
  - -1: trend bounded by base parm min-max and parms in transformed units (beware)
  - -2: endtrend and infl_year direct values
  - -3: end and infl as fraction of base range
- #_EnvLinks:
  - 1: P(y)=Pbase*exp(TVP*env(y))
  - 2: P(y)=Pbase+TVP*env(y)
  - 3: null
  - 4: P(y)=2.0/(1.0+exp(-TVP1*env(y) - TVP2))
- #_DevLinks:
  - 1: P(y)*=exp(dev(y)*dev_se)
  - 2: P(y)+=dev(y)*dev_se
  - 3: random walk
  - 4: zero-reverting random walk with rho

Semi-Parametric Selectivity: 2D Auto-Regressive (2DAR)

A new semi-parametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models.

Haikun Xu, James T. Thorson, Richard D. Methot, and Ian G. Taylor.

Can. J. Fish. Aquat. Sci. **00**: 1–18 (0000) dx.doi.org/10.1139/cjfas-2017-0446

Step 1: Pick any SS selectivity pattern as base

Step 2: Specify range of years and range of ages (or lengths) for deviations

Step 3: Specify std.dev., rho, behavior for earlier and later years

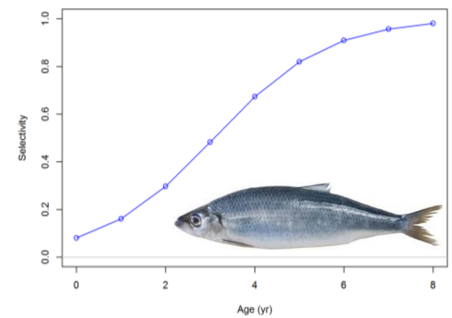
SS Control File Input for 2D Auto-Regressive

```
1   #  use 2D_AR1 selectivity(0/1):  experimental feature
#_specifications for 2D_AR1 and associated parameters
#_specs:  fleet, ymin, ymax, amin, amax, sigma_amax, use_rho, len1/age2, devphase, before_range, after_range
 1 1980 1981 4 6 5 1 1 3 3 0  #  2d_AR specs for fleet: FISHERY1
      0.001 0.3 0.1 0.1 0.5 6 -3  # sigma_sel for fleet, age/size:  1 4
      0.001 0.3 0.2 0.1 0.5 6 -3  # sigma_sel for fleet, age/size:  1 5
      -0.8 0.8 0.3 0.3 0.5 6 -3  # rho_year for fleet:  1
      -0.8 0.8 0.35 0.3 0.5 6 -3  # rho_age for fleet:  1
 1 1980 1990 4 12 8 0 2 4 3 0  #  2d_AR specs for fleet: FISHERY1
      0.001 1.3 0.5 0.1 0.5 6 3  # sigma_sel for fleet, age/size:  1 4
      0.001 1.3 0.5 0.1 0.5 6 3  # sigma_sel for fleet, age/size:  1 5
      0.001 1.3 0.5 0.1 0.5 6 3  # sigma_sel for fleet, age/size:  1 6
      0.001 1.3 0.5 0.1 0.5 6 3  # sigma_sel for fleet, age/size:  1 7
      0.001 1.3 0.5 0.1 0.5 6 3  # sigma_sel for fleet, age/size:  1 8
-9999  0 0 0 0 0 0 0 0 0 0 # terminator
```

SS case study: North Sea herring

Logistic selectivity-at-age curve (Pattern#12)

$$\hat{S}_{a,t} = \frac{1}{1 + e^{-\ln(19)(a-p_1)/p_2}} \times e^{\hat{\varepsilon}_{a,t}}$$



Three SS configurations for $\hat{\varepsilon}_{a,t}$ were compared

- SS-EM1 ("Constant selectivity")

$$\hat{\varepsilon}_{a,t} = 0$$

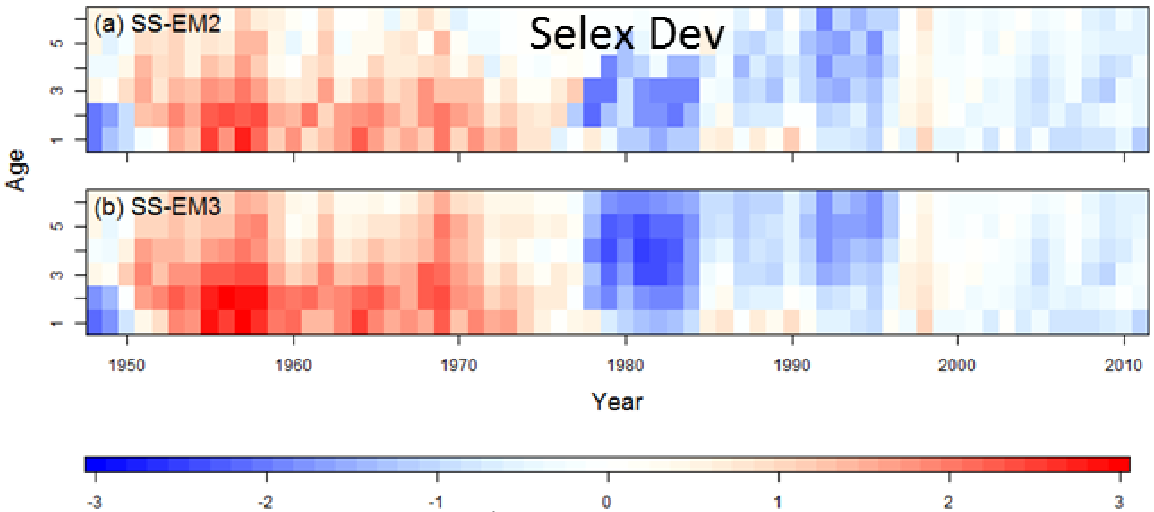
- SS-EM2 ("IID deviations")

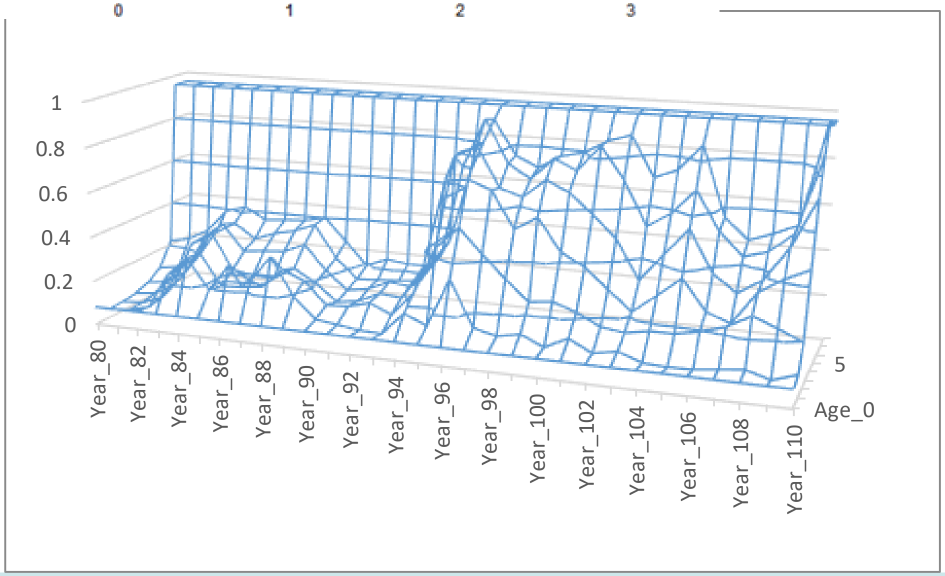
$$\hat{\varepsilon}_{a,t} \sim \mathrm{N}(0, \sigma_S^2)$$

- SS-EM3 ("2D AR deviations")

$$\mathrm{vec}(\hat{\boldsymbol{\varepsilon}}) \sim \mathrm{MVN}(\mathbf{0}, \sigma_S^2 \mathbf{R}_{\boldsymbol{total}})$$

2D_AR Estimated Selectivity





2D AR selectivity

- Why should we use this new selex feature?
  - Highly flexible in its interaction with existing SS features
  - More precise estimates of F and SB**
  - The variance in selectivity can be estimated objectively
  - Autocorrelated selectivity deviations
  - Best with high quality comp data
    - $\left(\mathbf{b} = 1 - \frac{\frac{1}{AT}\sum_{a=1}^{A}\sum_{t=1}^{T}\mathrm{SE}\left(\hat{\varepsilon}_{a,t}\right)^2}{\sigma_S^2}\right)$

Tradeoffs

- Highly flexible selectivity takes more parameters and matches comp. data well. The effective sample size of fit to comp. data will be high, but less information content remains to get precise recruitment and mortality estimates.
- Stiff selectivity means that the model will try to resolve pattern in the comp. data only with recruitment and mortality, or there will be high residual noise, hence low effective N. Forcing asymptotic selectivity is “stiff” and has direct effect on estimated Z.
- So the issue of how much flexibility to allow is intimately related to the issue of weight assigned to the composition data, and to re-weighting.
- When using primarily length data or age data with high ageing error, there is a diffuse connection between selectivity for a particular real age and the expected value for the proportion in a particular age bin or length bin. Hence, selectivity should be fairly smooth in this situation since the effect of selectivity is spread out.

Goals for Selectivity Estimates

- For fisheries, use composition data to accurately scale apical F across ages to get the right F-at-age
  - With noisy data, use stiff selectivity function to smooth across ages
  - With precise fishery composition data, use very flexible selectivity to match catch-at-age to attain the right F-at-age; as with VPA
- With precise survey composition data, use stiff enough selectivity such that composition data is highly informative about recruitment and Z

Weighting Composition Data

- You will not get the variance of derived quantities correct unless you treat each data type with appropriate data variance
- Iterative re-weighting
- Francis weights
- Dirichlet multinomial (D-M) for internal estimation
- Simultaneous D-M and 2D-AR selectivity (in progress)