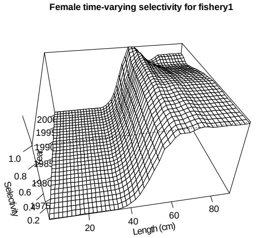
Introduction to Stock Synthesis

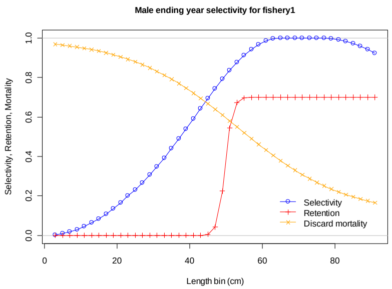
Richard Methot

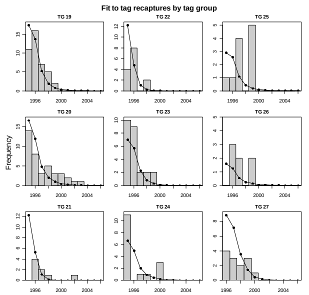
Senior Scientist for Stock Assessments

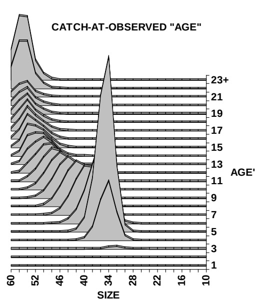
NOAA Fisheries

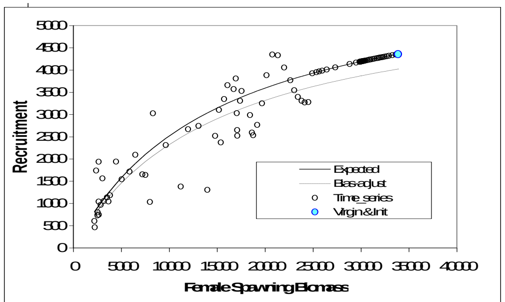
Seattle, WA











Integrated Analysis: Create Model for the Data

- Don't transform data to meet rigid model structure (as in VPA)
- Do add processes to model to develop expected values for diverse, lightly processed data
    - Improves understanding of processes
    - Allows simultaneous use of more types of data
    - Statistical properties of data are preserved and transferred to variance of final model results

Estimation, Benchmarks, Forecasts

Sometimes we use a sequence of separate analyses, each using a separate package and not transferring variance from step to step:

1. Estimate population abundance time series
2. Calculate benchmark quantities: target and limit F rates, sometimes based on first fitting spawner-recruitment curve.
3. Forecast future abundance and catch using the target F

Estimation, Benchmarks, Forecasts

Integrated Analysis can:

- Bring all steps into one analytical package
- Parameter variance propagates from population estimation to reference points to quantities in forecast
- Example output: probability that stock abundance will dip below the overfished threshold 5 years into the future, and the standard error of this probability

Sub-Models

- **<u>Population</u> Model**
  - Recruitment, mortality, growth
  - Age and/or size structured
- **<u>Observation</u> Model**
  - Derive expected values for data
- **Likelihood-based <u>Statistical</u> Model**
  - Quantify goodness-of-fit
- **Algorithm to search for parameter set that <u>maximizes</u> the <u>likelihood</u>**
  - Auto-Differentiation Model Builder (ADMB)
- **Cast results in terms of <u>management quantities</u>**
- **Propagate <u>uncertainty</u> onto confidence interval for management quantities**
- **Parametric bootstrap**

Components of population dynamic models

- State variables: In the future depend on
  - Current state
  - Parameters
  - Forcing (external shocks)
  - Rules of change (equations)

Rules of change

Forces (e.g. catches, env.)

$$S_{t+1} = f(S_t, P, u_t)$$

State variables in next time period

State variables (e.g., numbers)

Parameters (e.g., growth rate)

Stock Synthesis Population Sub-Model

NUMBERS-AT-AGE

Morphs: gender, settlement time, growth pattern;

"platoons" can be nested within morphs to achieve size-survivorship;

RECRUITMENT

Expected recruitment is a function of total female spawning biomass;

Optional environmental input;

apportioned among cohorts and morphs;

Forecast recruitments are estimated, so get variance

AREA

Recruits are distributed among areas

Age-specific movement between areas

FLEET / SURVEY

Length-, age-, gender selectivity

CATCH

F to match observed catch;

Catch partitioned into retained and discarded, with discard mortality

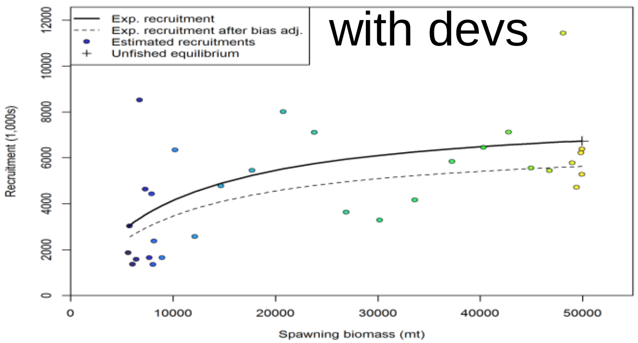
PARAMETERS

Can have prior/penalty;

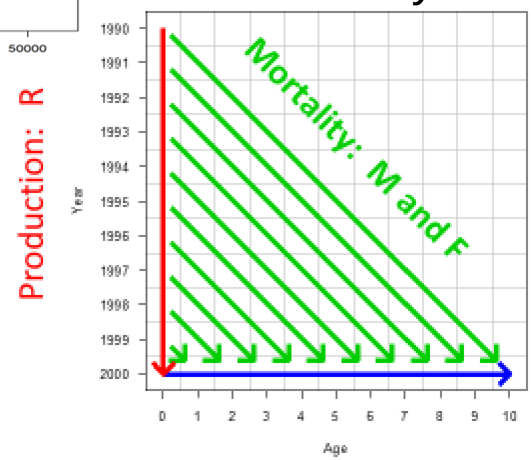
Time-vary as time blocks, random annual deviations, or a function of input environmental data

Modeling Population Numbers

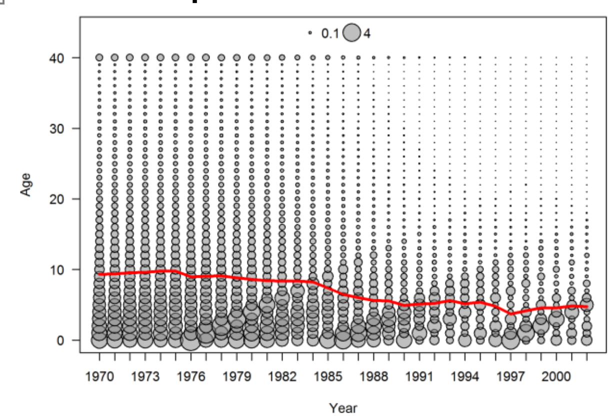
Spawn_Recruit
with devs



Mortality



Population Numbers



Length Compositions

PRO:

- Fish are easy and cheap to measure
- Usually a lot more have been measured, than have been aged
- Once you admit that ageing error and bias exist, then is it that much worse to just measure the fish?
- Length modes, especially for young fish can be fairly distinct, so easily translated to age
- Once one admits the potential for lengths to be informative, logical to consider broad size categories, say 1 kg weight intervals, or a really broad measure like mean length.

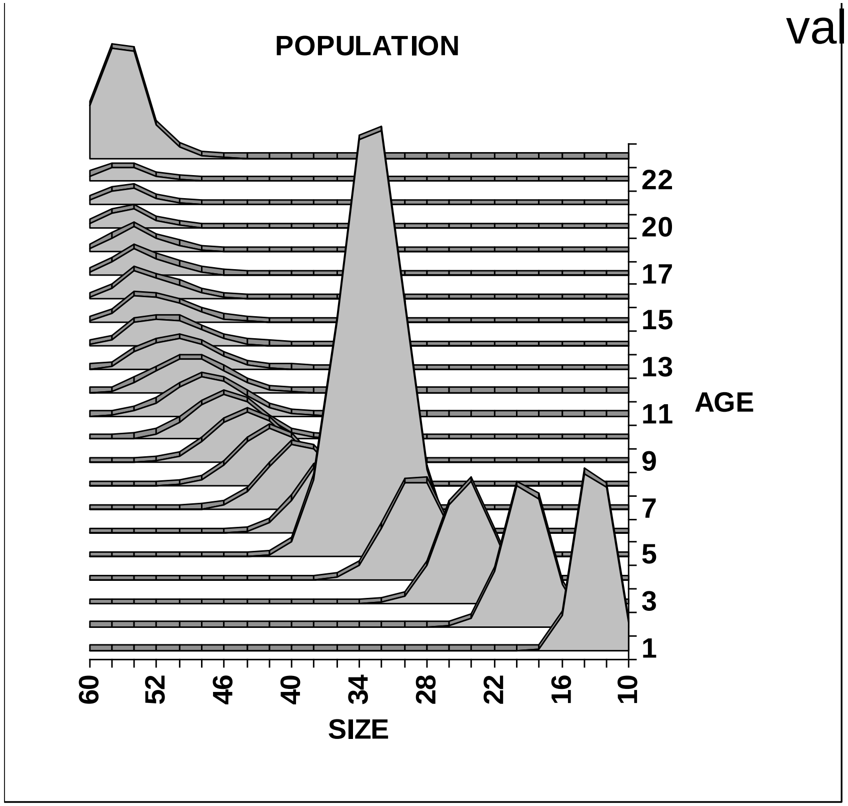
CON:

- Fish grow within the year and don't grow at same rate every year
- Fish don't grow linearly, so ability to distinguish ages degrades as fish approach Lmax
- Not all fish have the same size-at-age, but that is similar to fact that not all otoliths of age 8 fish will be aged to be "8"

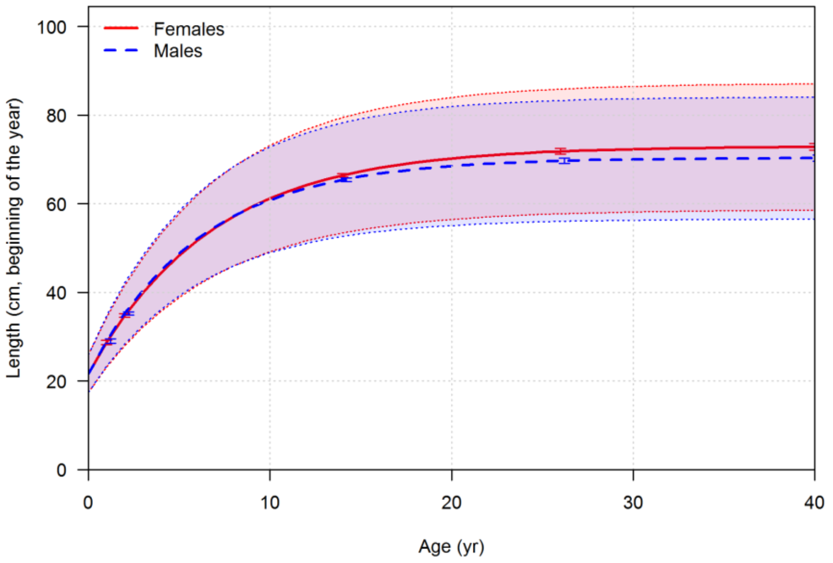
Age-Length Structure

GOALS:

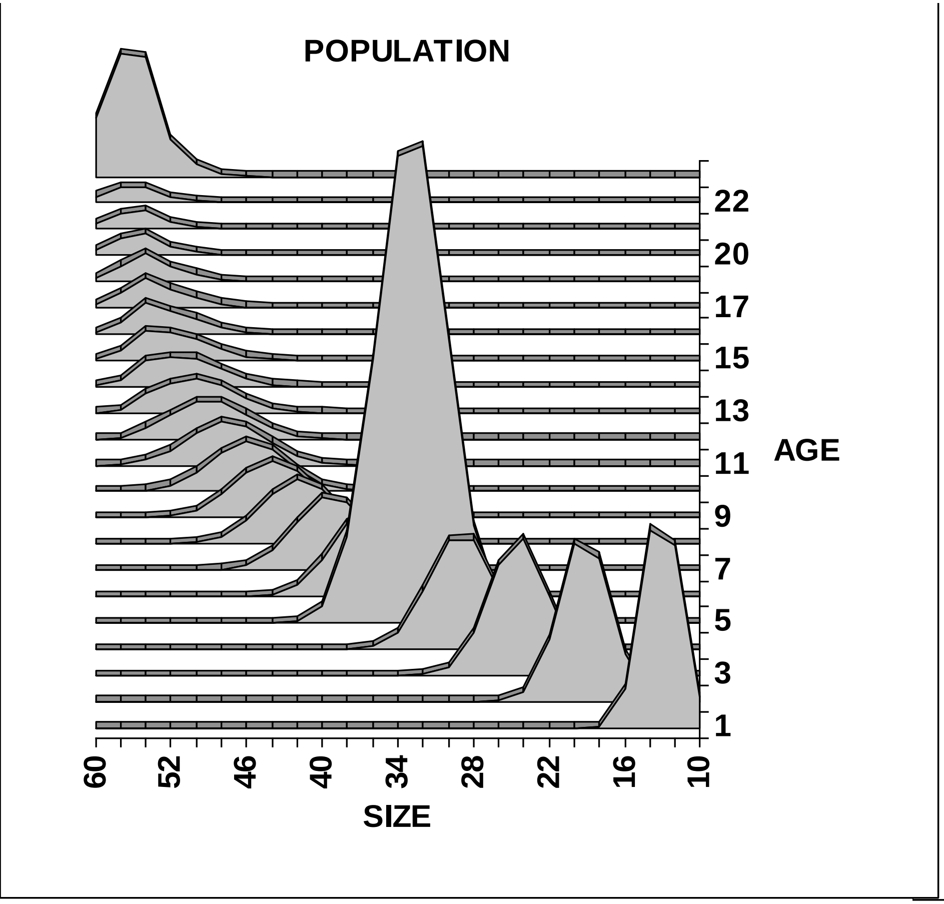
- Derive body-wt-at-age when empirical observations not available
- Use growth curve estimated from available size-based data by deriving expected values for such data

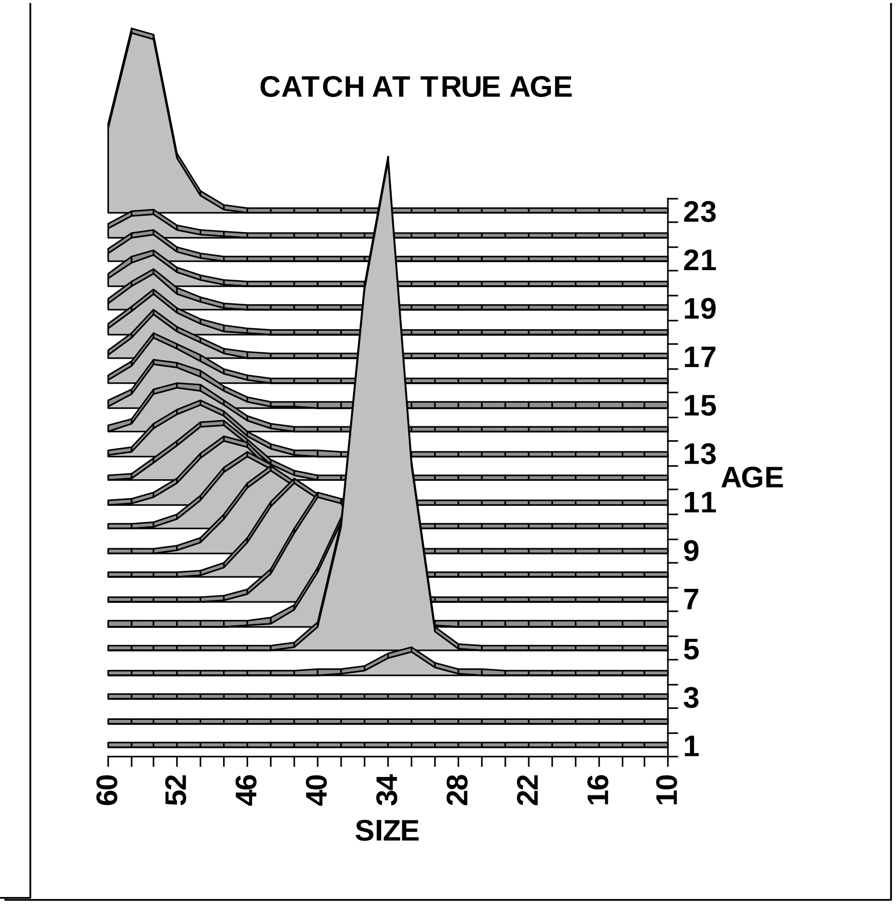


Estimated Growth Curve

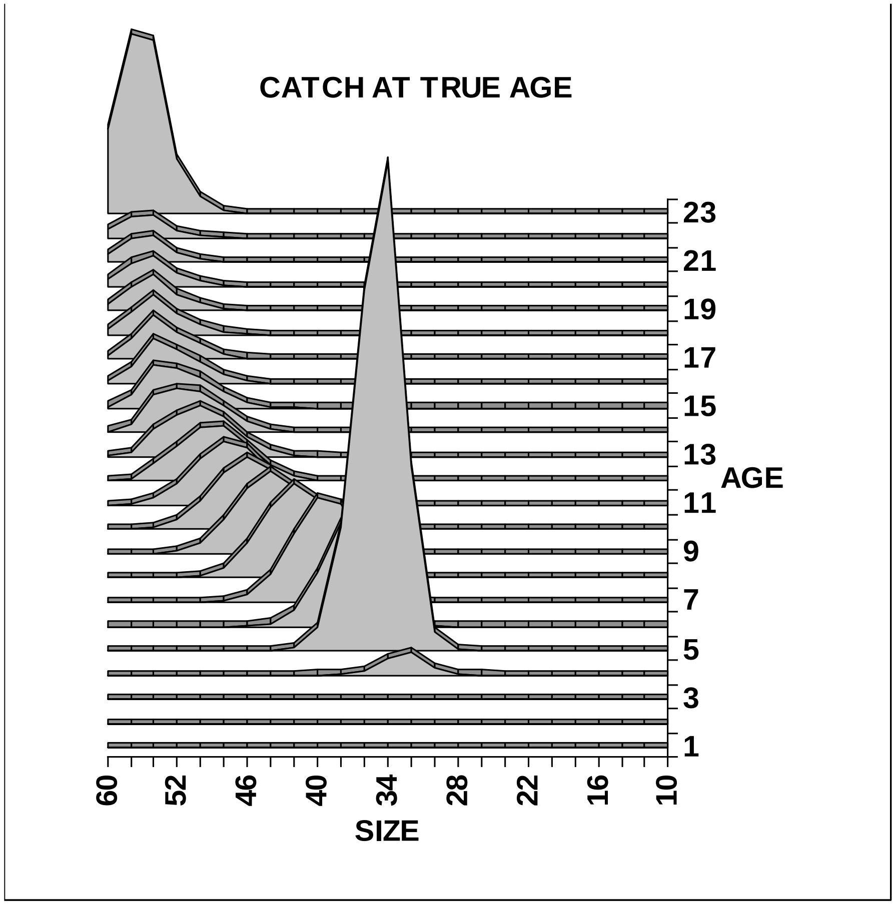


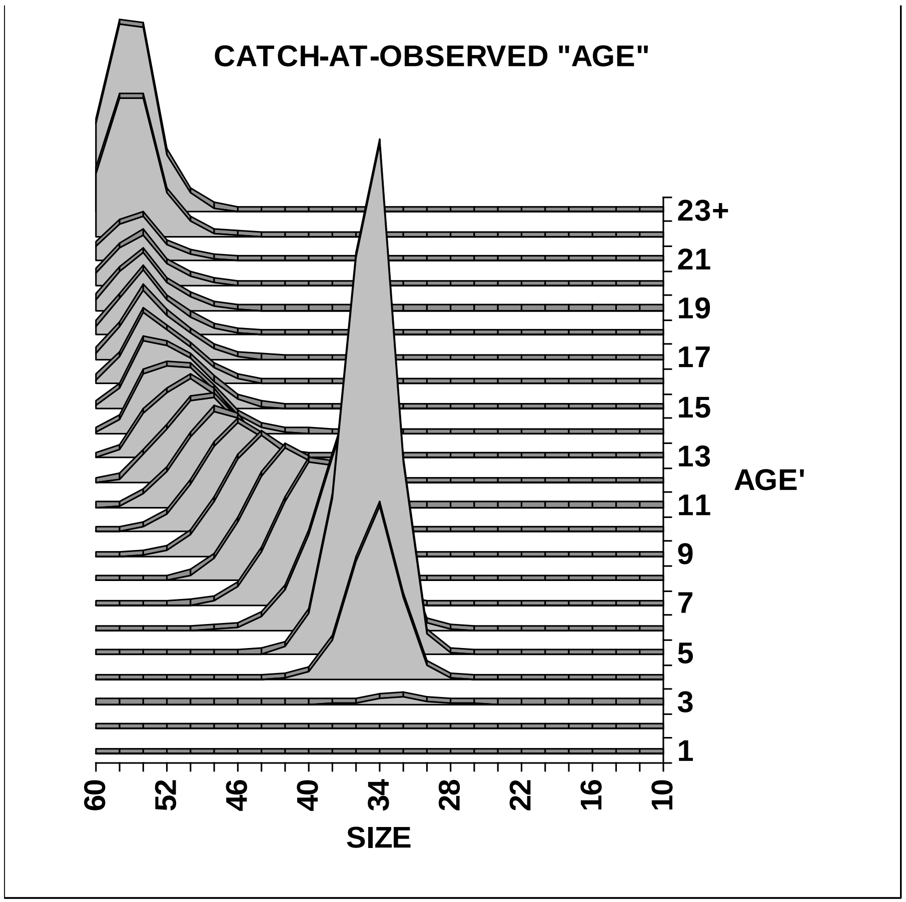
Selectivity Produces Sample





Ageing Error adds Blur



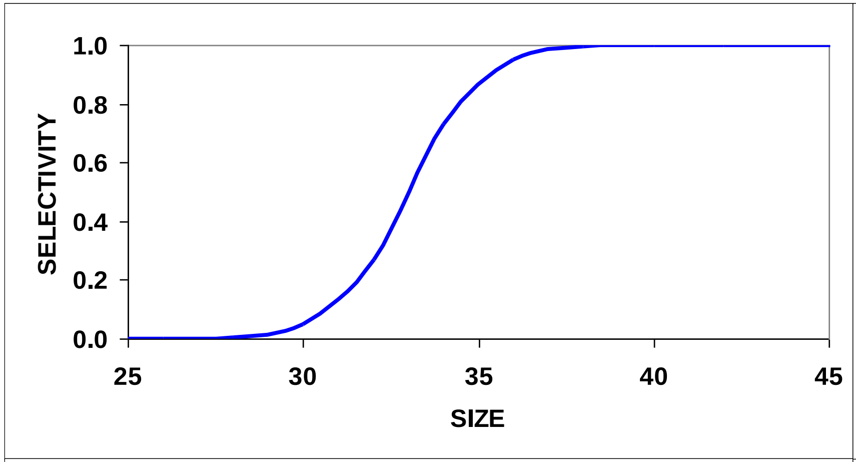


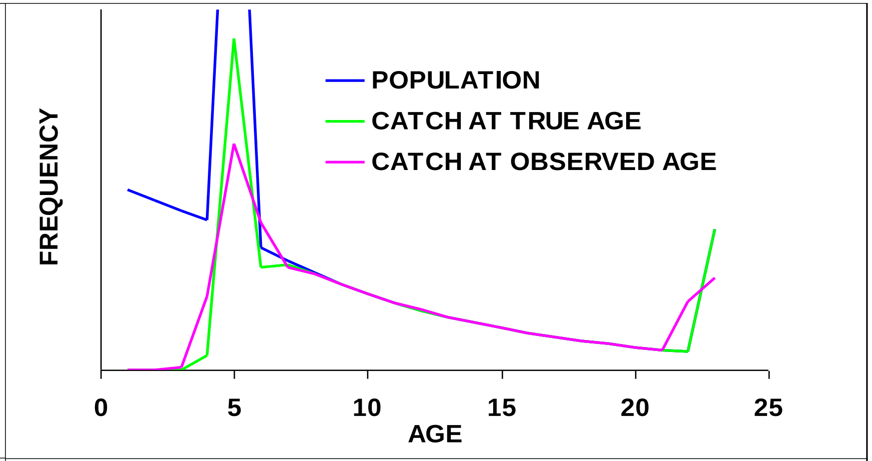
Blur cannot be removed from actual samples, so blur model's expected value to same degree that you expect data are blurred!

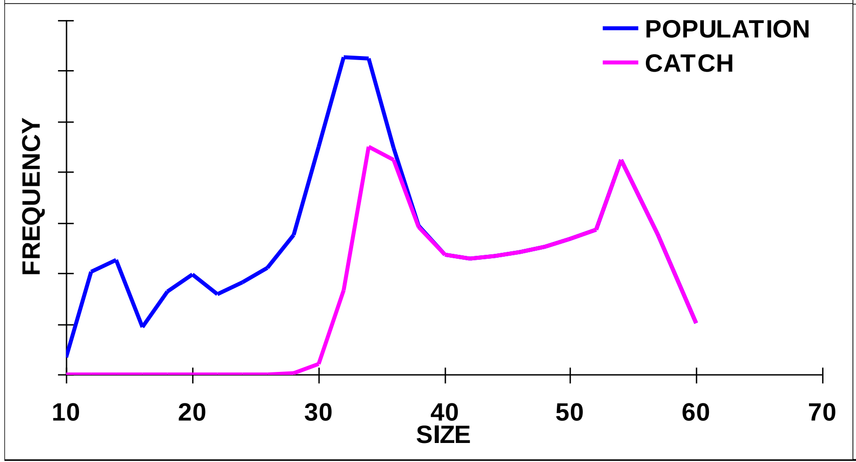
Expected Value for a Length Composition Obs

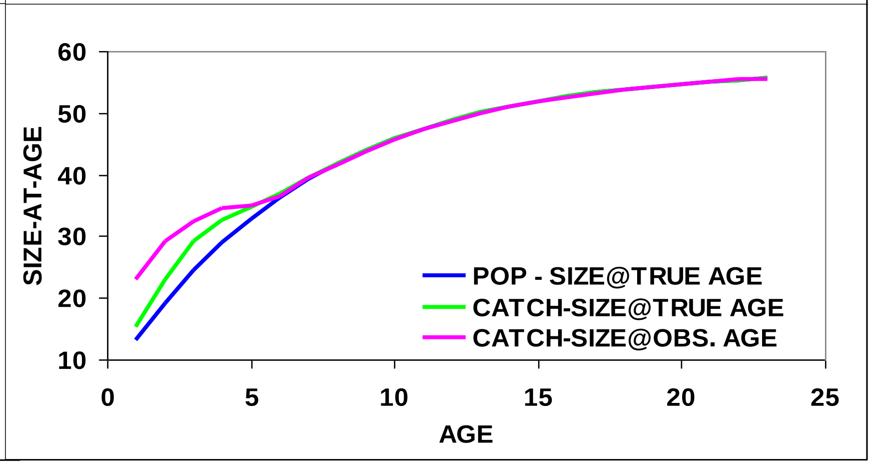
- In each time step, SS calcs the ALK at the beginning and middle of the time step
- The ALK distributes the population numbers-at-age across the length bins for each age
- Then age and size selectivity are applied directly to this pop ALK to get the sampled ALK for each fleet that has data in that time step
- If there is a retention function, then partition the sampled ALK into retained ALK and discarded ALK
- Summing across ages gives e(length comp)
- Summing across lengths gives e(true age comp), then apply ageing error to get e(age’ comp)
- Sum the whole sampled ALK and multiply by Q to the e(survey)
- Do dot product between the age-length’ matrix and the mean length in each length bin to get the e(mean length at age’) and the standard deviation of that mean length. Convert to a s.e. based on the number of fish measured at that age’

Expected Values for Observations









Deriving Age-Based Processes

- Survivorship is age-based in SS
- So, how does length selectivity translate into F-at-Age
- Dot product of (pdf of length-at-age) and Length Selectivity produces mean selectivity-at-age due to size selectivity
- Then times direct age selectivity gives total selectivity-at-age
- Similarly, we get mean body weight-at-age for each size selective fishery or survey, and mean fecundity-at-age from length maturity curve and length fecundity curve

- Lots of calculations:
    - In seasonal model, need to calc derived age selectivity for each season
    - Whenever growth changes, need to update all these derived quantities
    - SS runs much slower when time-varying growth is turned on

Stock Synthesis Observation Model

- Catches
- Guesstimate on depletion
- Discards
- Effort

- Indices of abundance
- Index of random effect
- Absolute abundance
- Average weight, average length
- Tagging

- Catch-at-age with ageing error
- Catch-at-length
- Generalized weight or length composition
- Age-conditional-on-length
- Average length-at-age, average weight-at-length

Goodness-of-Fit

- Calculated in terms of –log(likelihood)
- Basically a variant of (obs-exp)/se –log(se), but details vary by data type
- Separate value stored by source (fleets and surveys) and kinds of data (CPUE, age comp, size comp, etc.)
- Additional logL components for recruitment deviations, parameter priors and other quantities
- Overall logL is user-weighted composite of all components

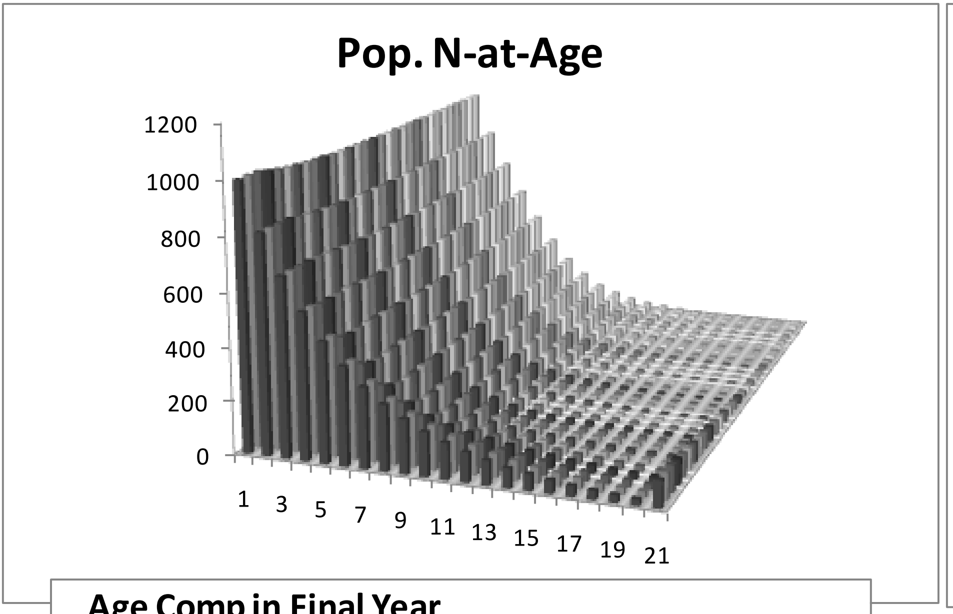
Estimation Procedures

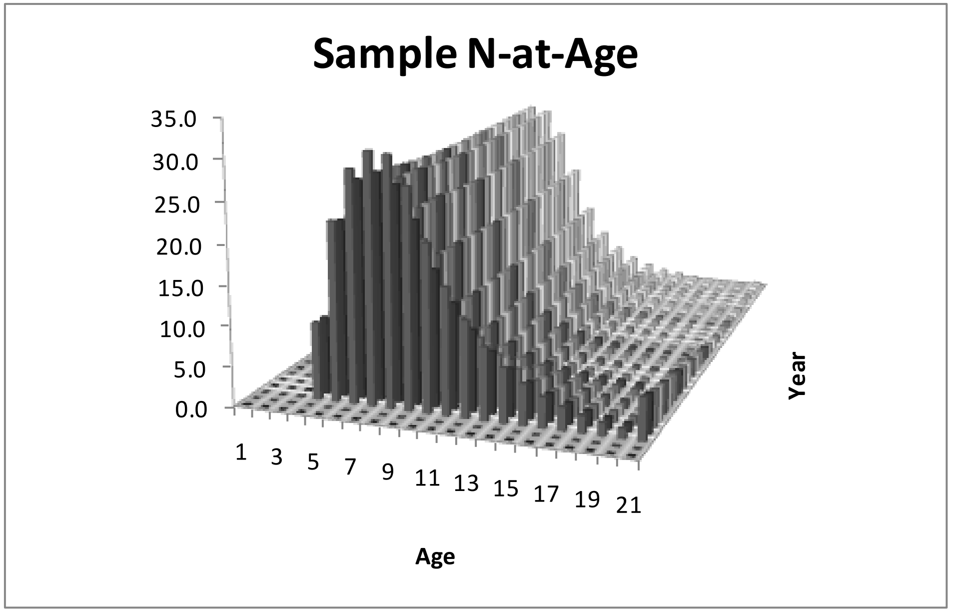
- Maximize negative logL using ADMB
- Parameters have user-controlled: min, max, initial value, phase, and optional prior
- Capable of MLE, MCMC, parametric bootstrap (but no ADMB-RE and no ADMB likelihood profiles)
- Benchmark (MSY, SPR) and forecast done in sd_phase and mceval_phase

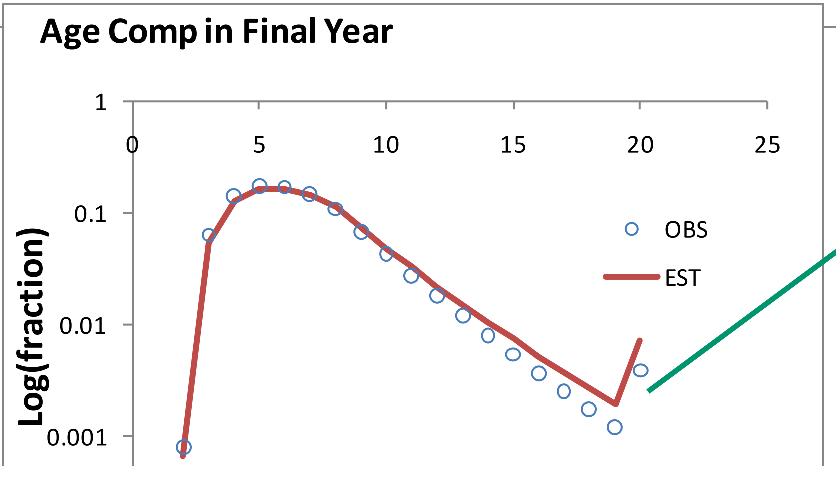
Tuning a Model

- Result will be a complex weighted average of fit to all included data;
  - Type, contrast and precision of data determine its influence
  - Examine residuals and root mean squared error of fit to data
  - Parsimoniously, add enough process to remove pattern to residuals
  - Judicious re-weighting of inputs to make input error assumptions consistent with rmse of fit
  - Some internal options for adjusting data variance according to achieved goodness-of-fit
  - Much more on model diagnostics later in week

Population Scenario

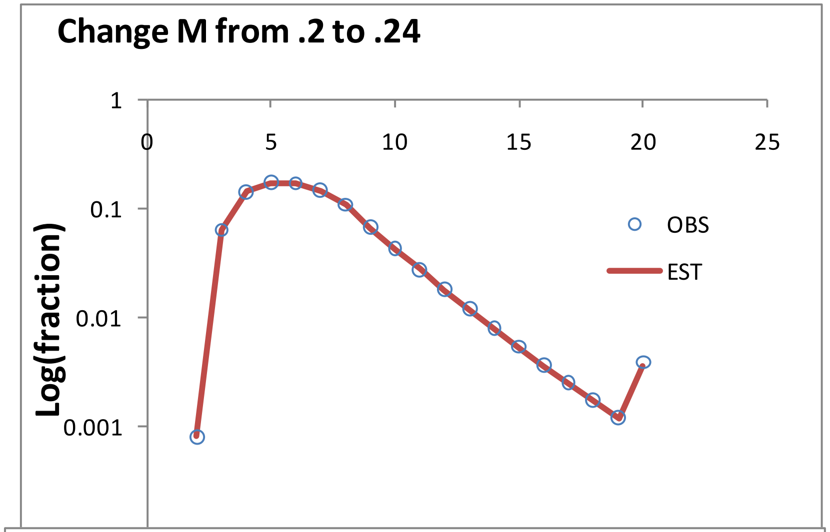


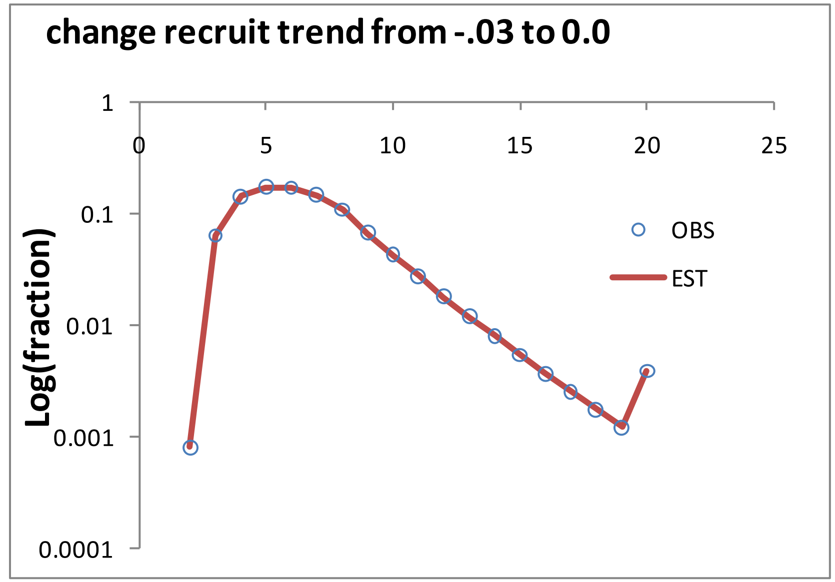


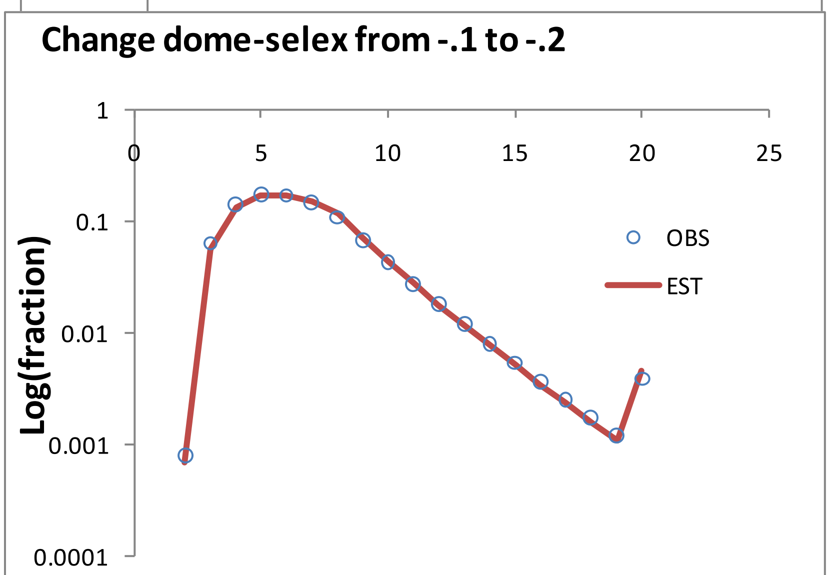


What process' parameter is causing this deviation?

Equally Likely Solutions







- So, productivity, mortality and selectivity are confounded
- Attempt to estimate all 3 parameters with one datum would produce parameter correlation near 1.0
- Unique solutions require more data with sufficient contrast along relevant dimensions

No Magic Bullet

- Allows many kinds of data, but data does not assure contrast that allows accurate estimation
- Allows many processes to be investigated, but cannot magically remove confounding
- Fixing parameter values for some processes (M, asymptotic selectivity) will tighten confidence intervals by excluding some alternative explanations for the data – that’s bad and can be misleading
- Art of assessment modeling is matching model complexity to complexity of the population and sampling processes that created the data
  - Simple models are “stiff” with respect to real world complexity, so will show patterns in residuals when presenting them with rich data
  - Complex models have many parameters, so will tend to over fit simple data

Benefits of Stock Synthesis

- **Flexible range of population processes**
- **Allows diverse data**
- **Propagates uncertainty well**
- **Seamless from data-limited to complex**
- **Widely used**
- **R tools**
- **Evolving based on decades of development and exploration**