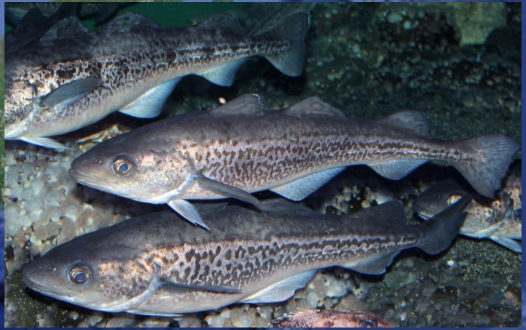
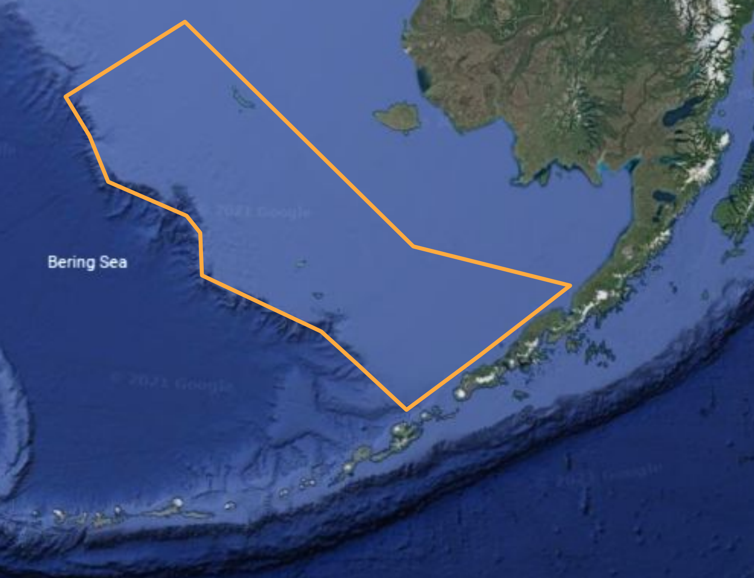
Acoustic-Trawl Survey Uncertainty

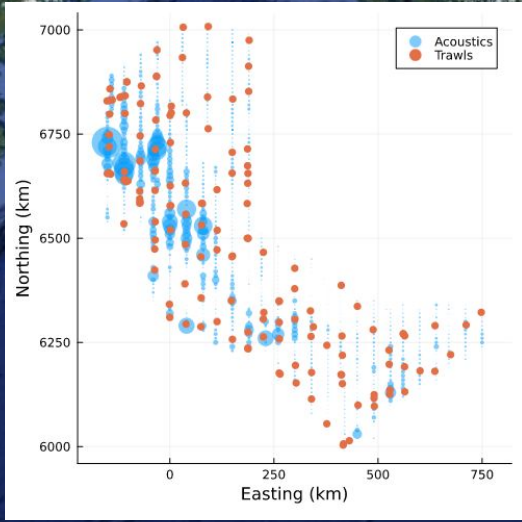
NPFMC Groundfish Plan Team Meeting, 22 September 2023

Samuel S. Urmy, Alex De Robertis, and Patrick Ressler
Midwater Assessment and Conservation Engineering Group
NOAA Alaska Fisheries Science Center

Acoustic-trawl surveys of Alaska pollock in EBS





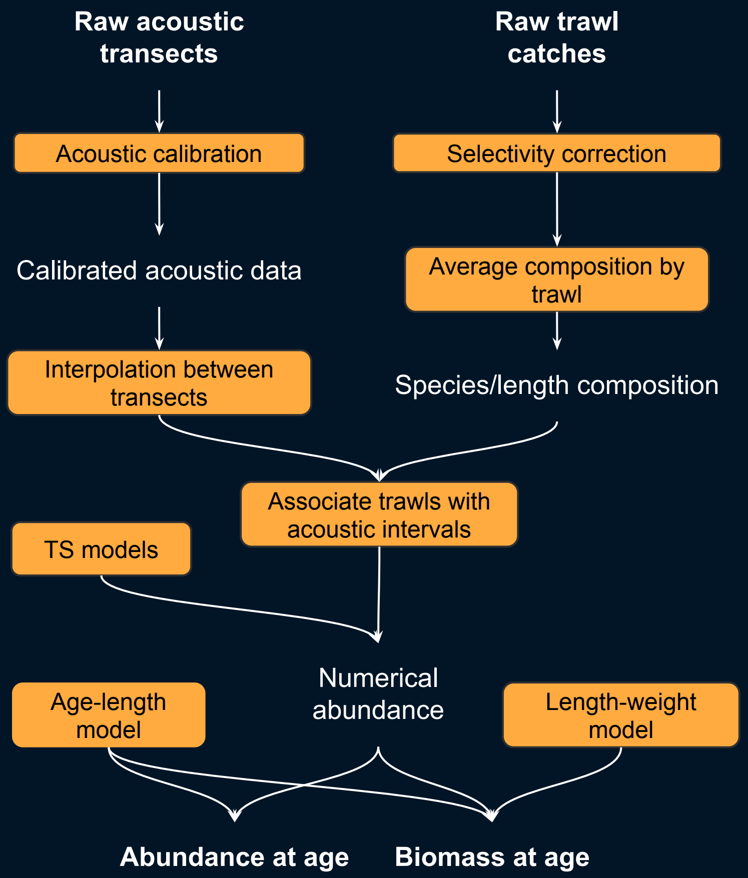


- Currently, estimate relative error using 1D geostatistics
- Typical coefficient of variation: 4-8%
- Seems too precise, so inflated to 20% for stock assessment
- What is it really?

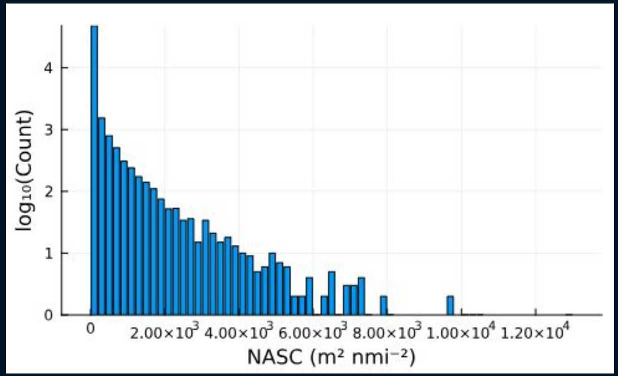
Estimating uncertainty is a challenge for acoustic-trawl surveys

- A few more-comprehensive estimates elsewhere:
    - Newfoundland cod (Rose et al. 1999)
    - Antarctic krill (Demer 2004)
    - New Zealand hoki (O’Driscoll 2004)
    - Norwegian herring (Løland et al. 2007)
- …And for pollock at AFSC:
    - 2007: Walline, “Geostatistical simulations of EBS walleye pollock…”
    - 2016: Woillez et al., “Evaluating total uncertainty…”
    - 2020-present: myself

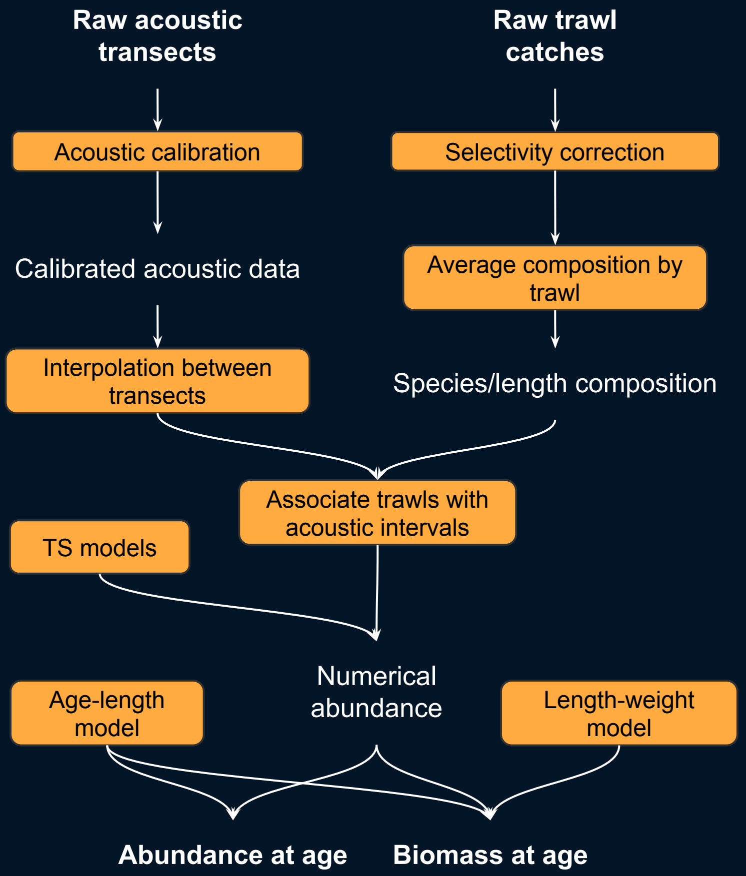
Why this is a hard problem



- Multiple steps between acoustics/ trawls and numbers/biomass
- Combining two datasets with:
  - Unique uncertainties/biases
  - Very different spatial scales
- Acoustic data are *extremely* non-Gaussian

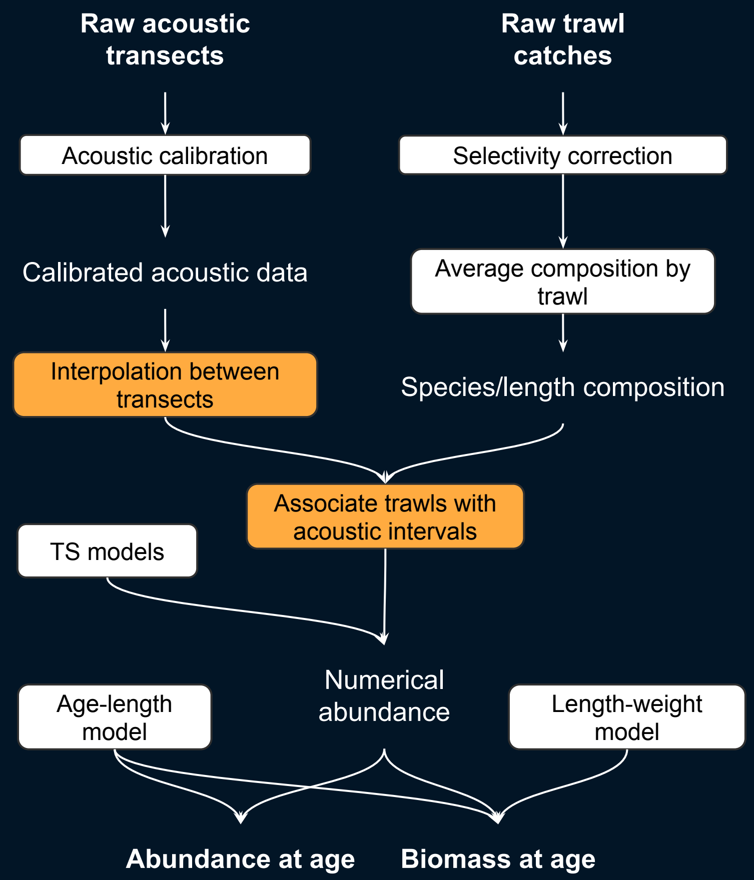


Parametric and non-parametric bootstrapping



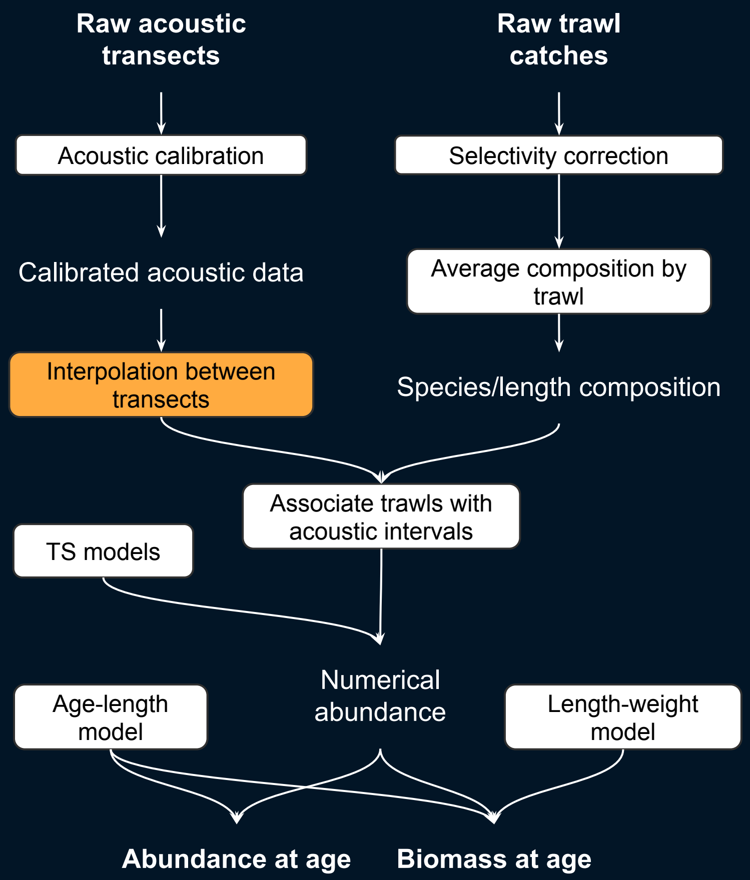
- Follows attempts at integrated Bayesian modeling
- Resampling/simulation for each step of calculations
- Mirrors standard MACE survey analysis
- Computationally fast and stable
  - Can investigate contributions of individual sources of uncertainty

Parametric and non-parametric bootstrapping



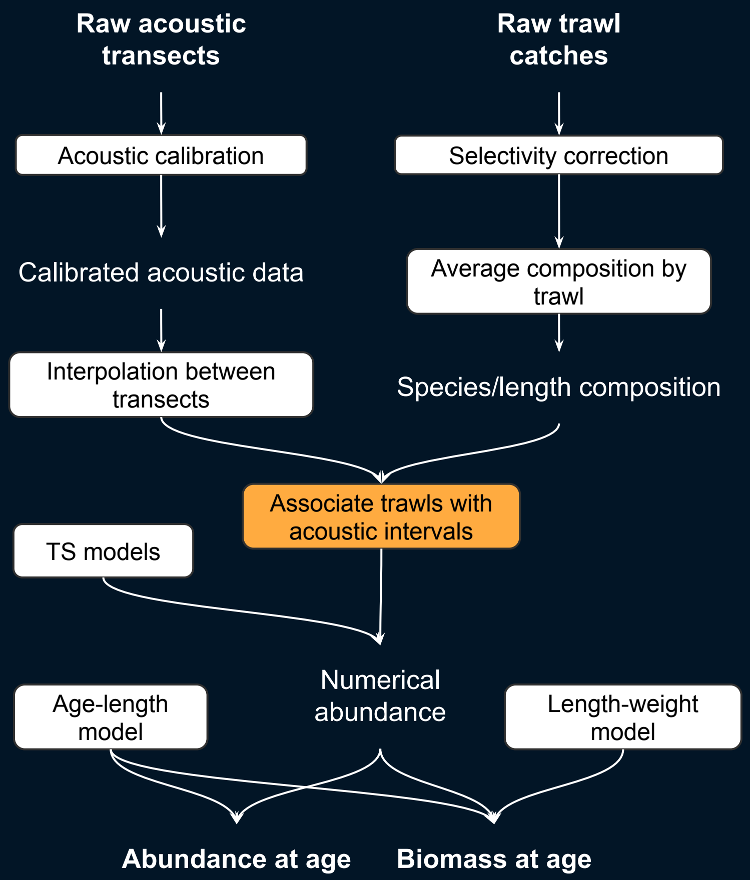
- Follows attempts at integrated Bayesian modeling
- Resampling/simulation for each step of calculations
- Mirrors standard MACE survey analysis
- Computationally fast and stable
  - Can investigate contributions of individual sources of uncertainty
- Key: interpolate acoustics and trawls to common spatial grid

Parametric and non-parametric bootstrapping



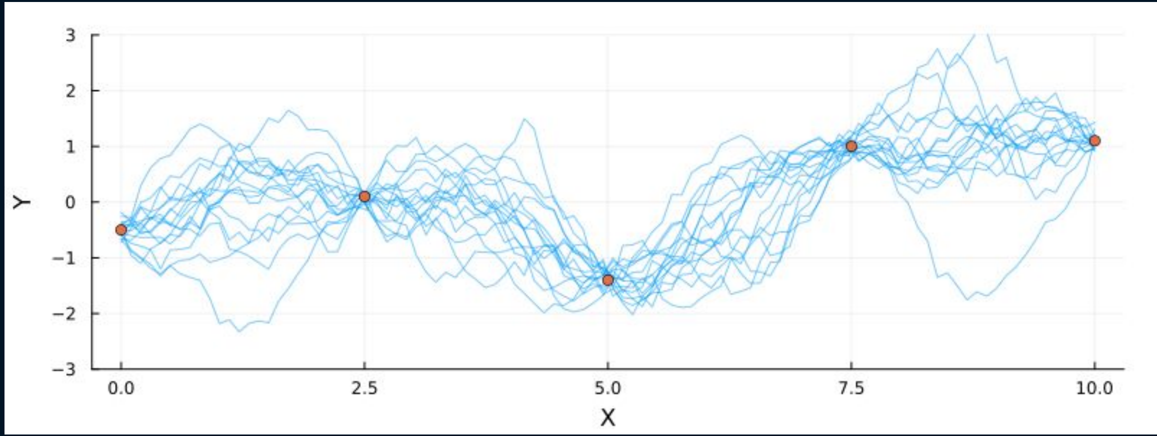
- Follows attempts at integrated Bayesian modeling
- Resampling/simulation for each step of calculations
- Mirrors standard MACE survey analysis
- Computationally fast and stable
  - Can investigate contributions of individual sources of uncertainty
- Key: interpolate acoustics and trawls to common spatial grid
  - Conditional geostatistical simulation

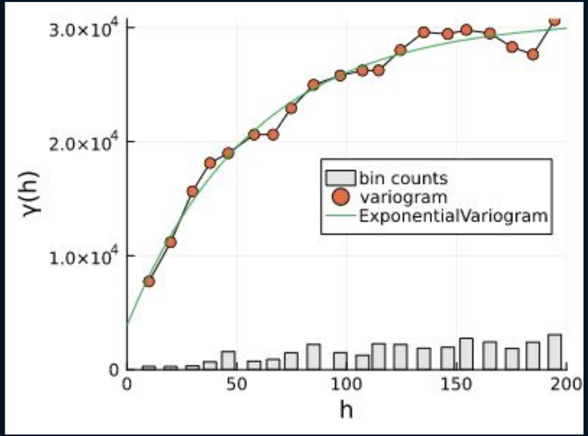
Parametric and non-parametric bootstrapping



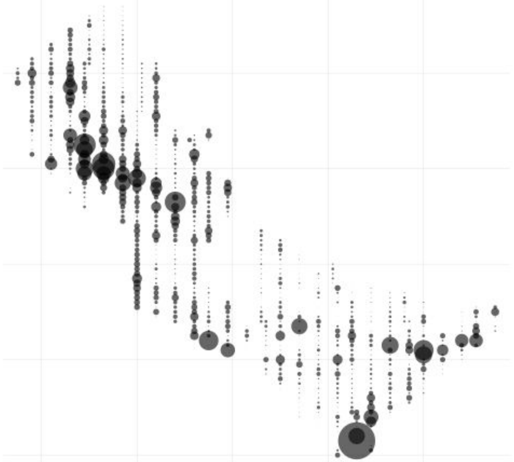
- Follows attempts at integrated Bayesian modeling
- Resampling/simulation for each step of calculations
- Mirrors standard MACE survey analysis
- Computationally fast and stable
  - Can investigate contributions of individual sources of uncertainty
- Key: interpolate acoustics and trawls to common spatial grid
  - Conditional geostatistical simulation
  - Randomized spatial assignment

Conditional (non-) Gaussian geostatistical simulation





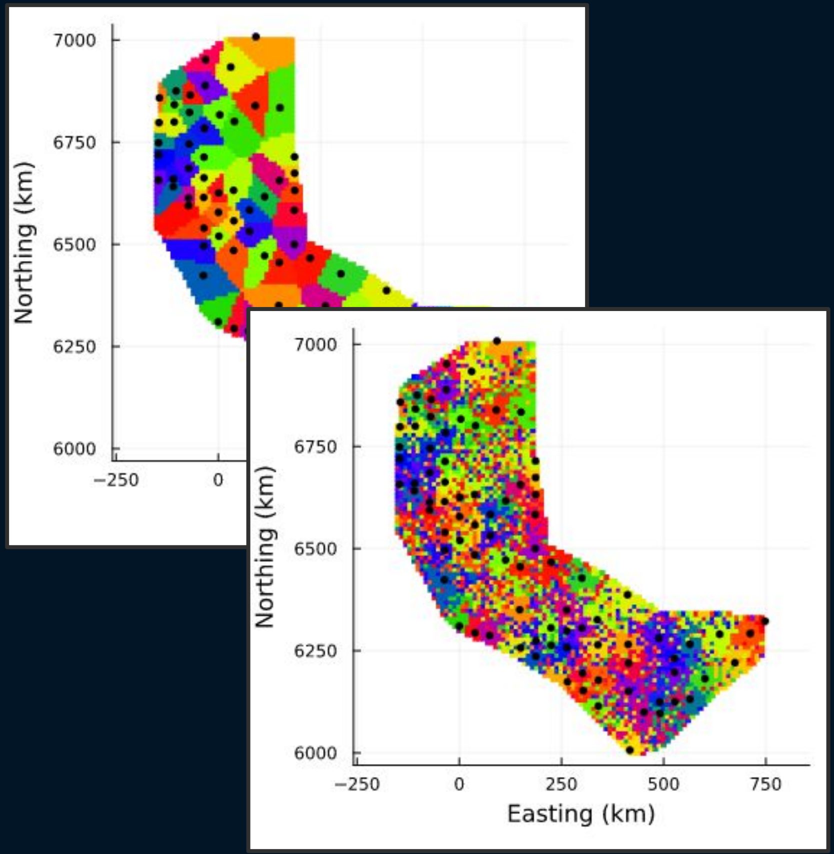
- Standard routine for Gaussian data
  - Variogram defines covariance matrix *Q* of desired simulated data *x*
  - Uses “Cholesky trick,” a.k.a. “Lower-upper Gaussian simulation,” LUGS
  - If $Q = LL'$, then $x = L z$, where $z \sim$ i.i.d. Normal(0, 1)
- Non-Gaussian data: can transform it, but complicated and/or biased
- But...turns out, *z* doesn’t have to be normal! Just need var(*z*) = 1
  - Lower-upper non-Gaussian simulation: LUNGS
  - Choose *z* from {Gamma, Inv. Gamma, Inv. Gaussian, Lognormal} based on KLD of *x* from observed backscatter



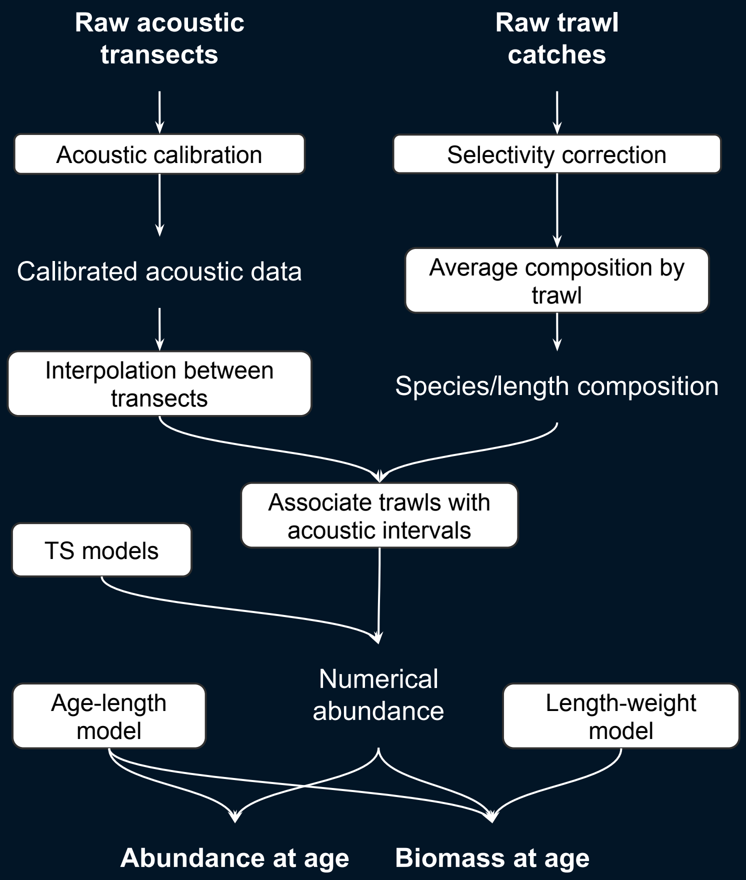
Conditional simulations on 10 x 10 km grid

Randomized acoustic-trawl assignment

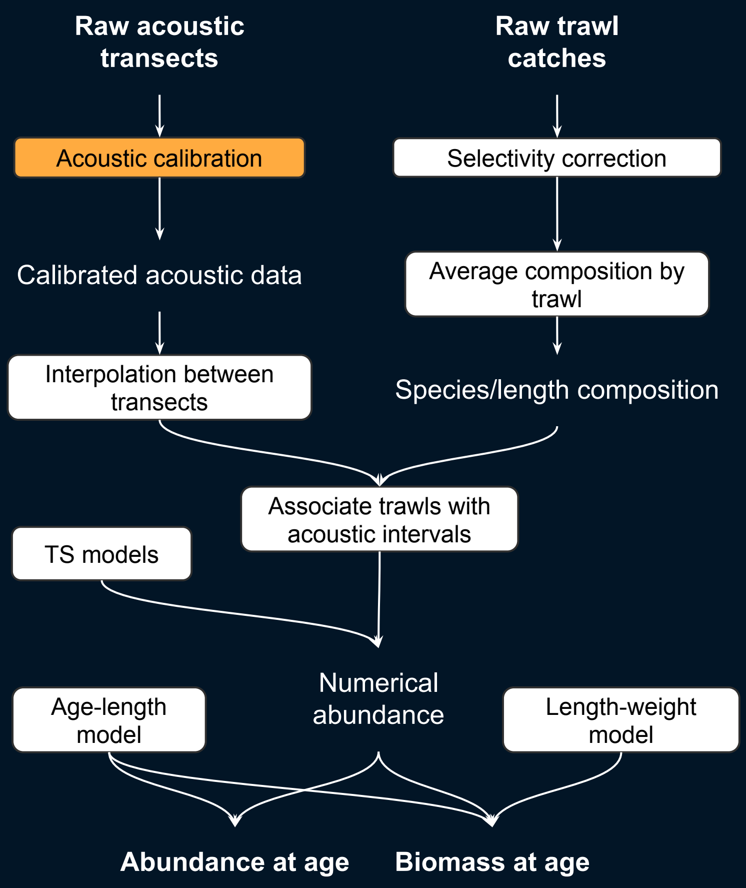
- Trawl composition gives us species, sizes, target strengths
- Each acoustic interval scaled by nearest trawl
- Are these scaling factors representative of area?
- Randomly assign each grid cell to a trawl, probability ~ distance^{-1}



Bootstrapping procedure at each iteration:



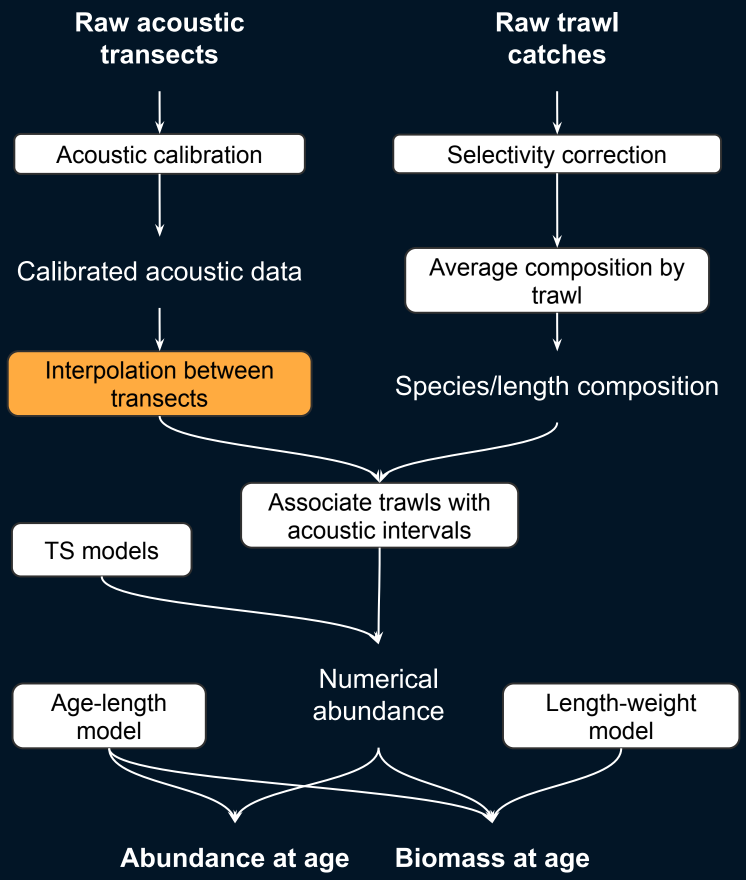
Bootstrapping procedure at each iteration:



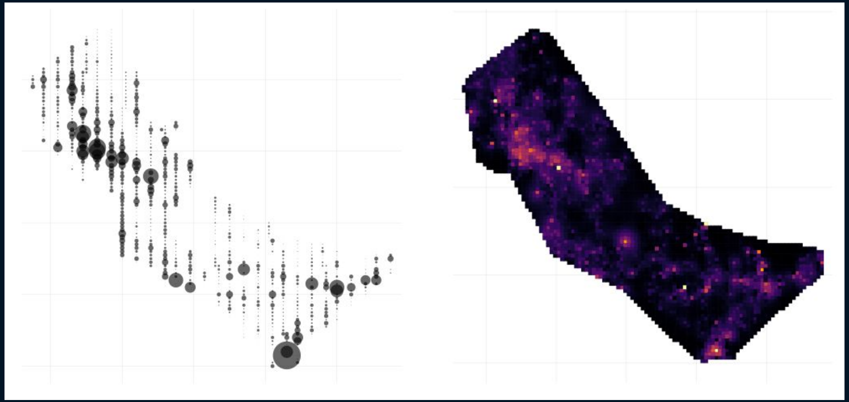
Apply random “calibration error” from normal distribution with standard deviation of 0.1 dB (about 2.4% in linear terms).

Based on variability in past calibrations and acoustic theory.

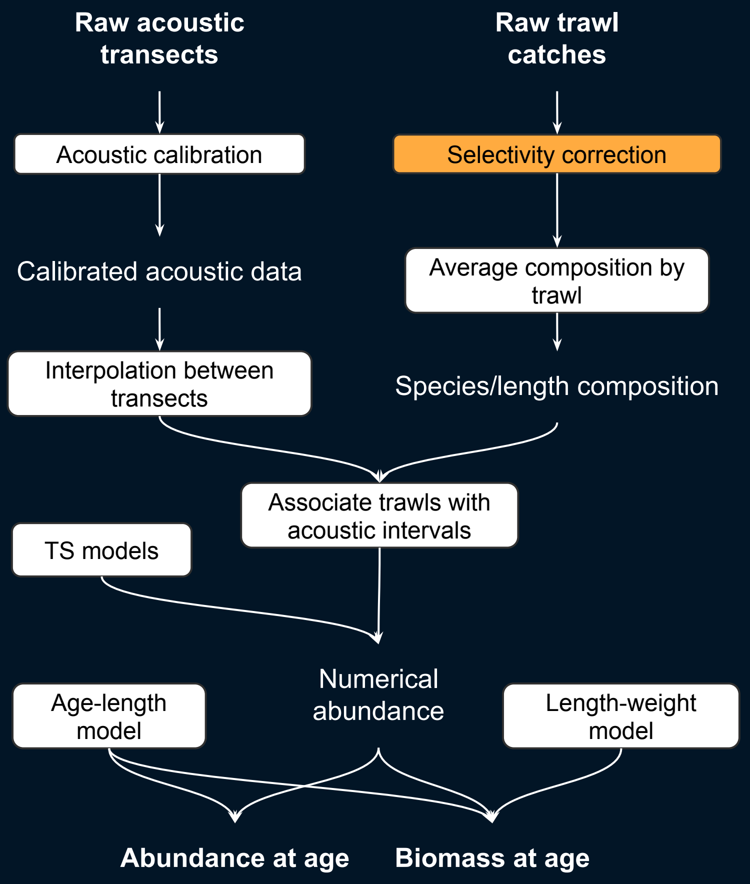
Bootstrapping procedure at each iteration:



Spatial sampling error: simulate backscatter field, conditional on variogram model and data observed along transects.

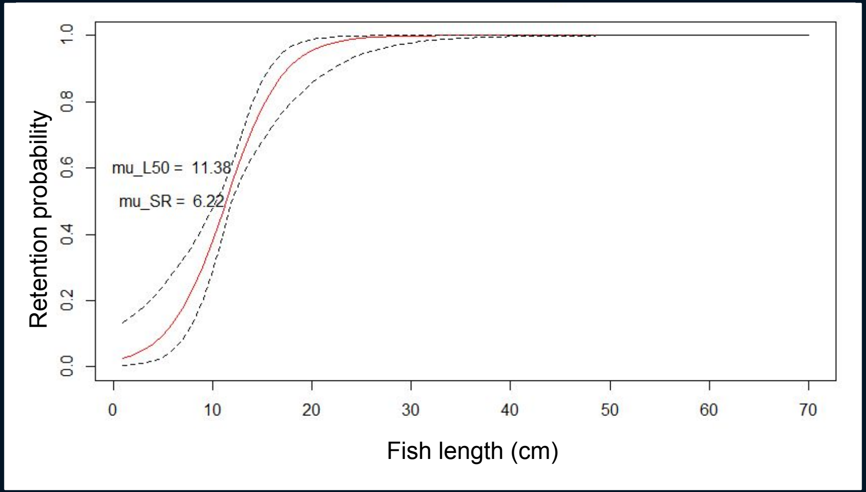


Bootstrapping procedure at each iteration:

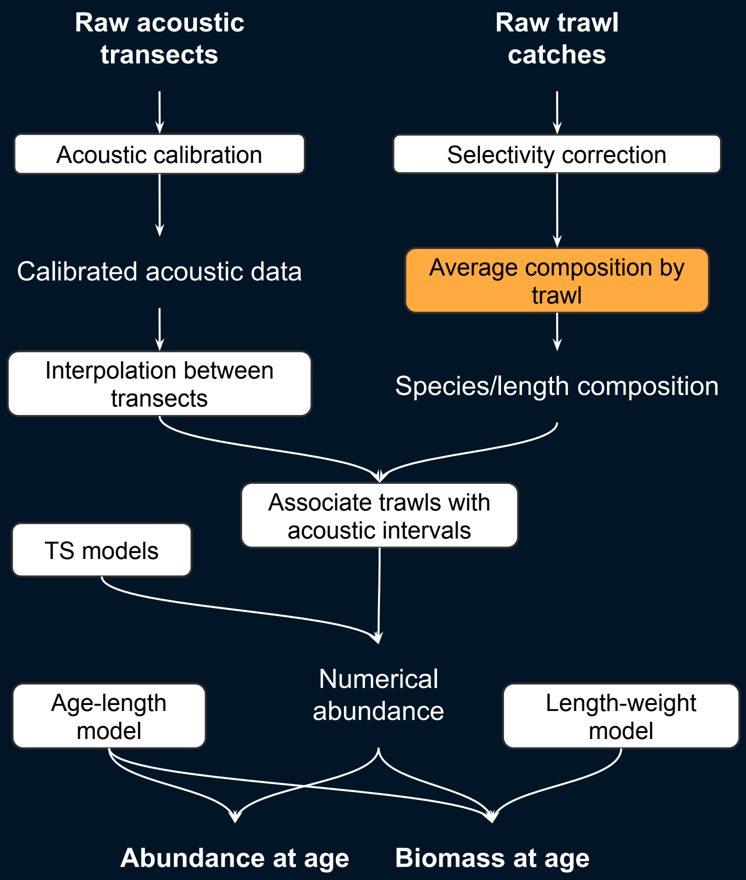


Choose random trawl selectivity curve.

Based on models fit to pocket-net data by Kresimir Williams et al.



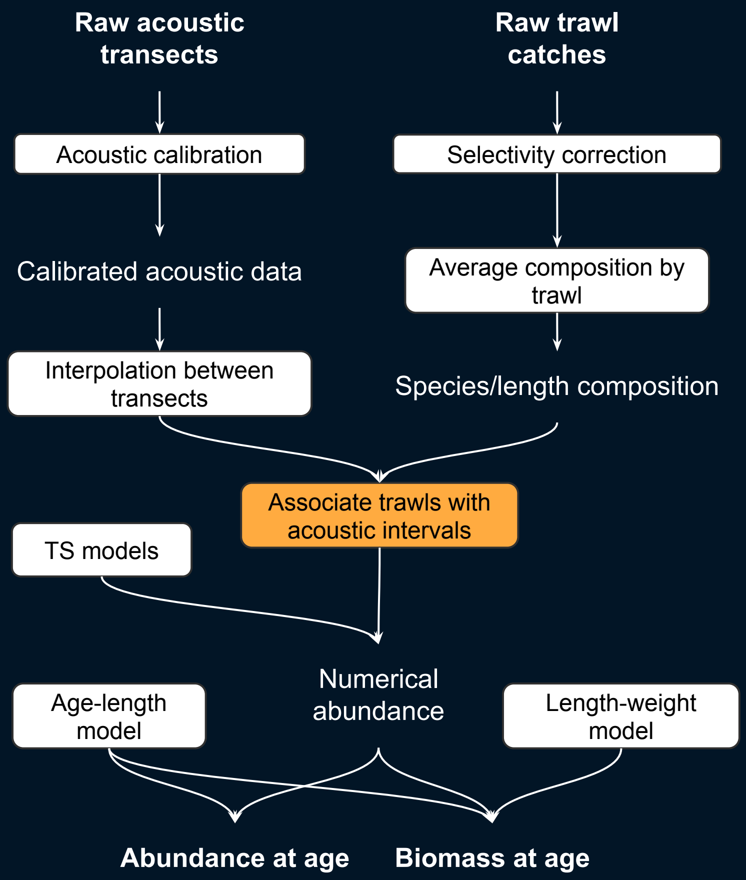
Bootstrapping procedure at each iteration:



Resample catch (individual fish) in each trawl with replacement.

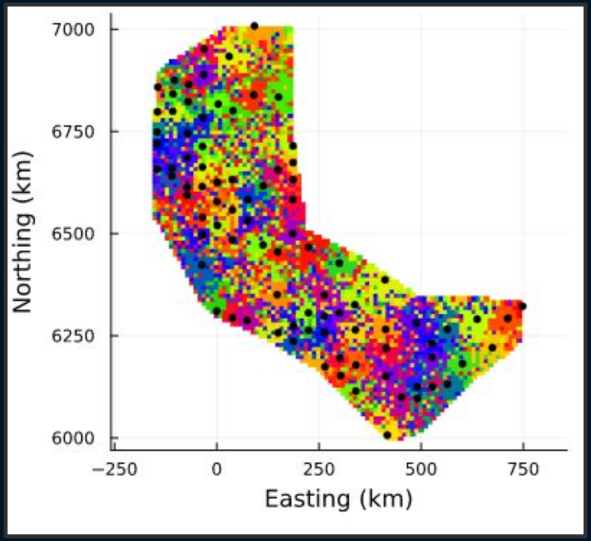
Calculate species/length composition, corrected for selectivity using function from prior step.

Bootstrapping procedure at each iteration:

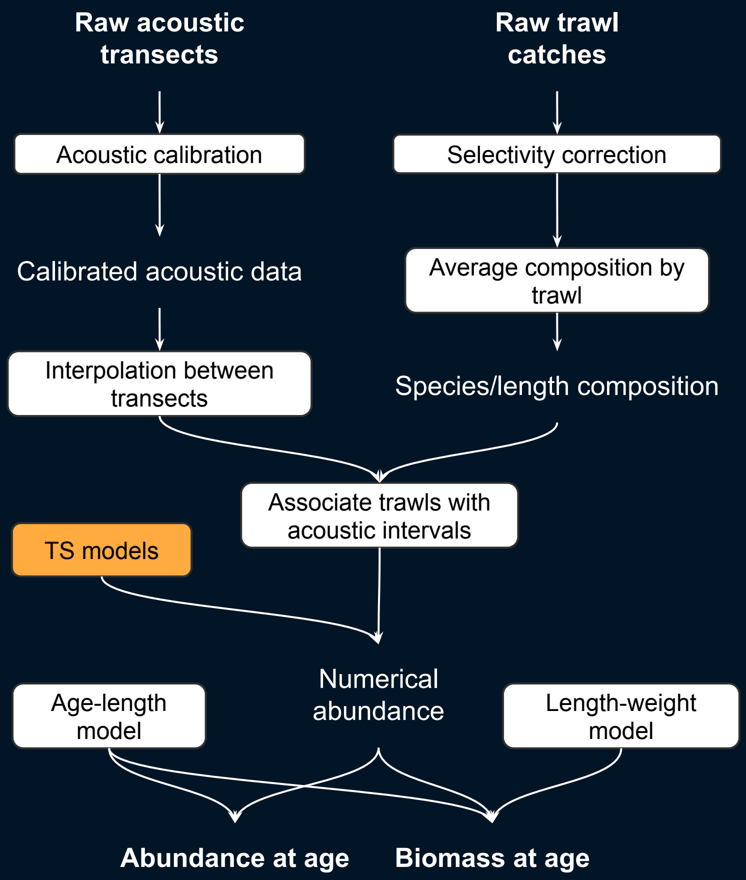


Drop one trawl at random.

Assign each grid cell to a trawl, with probability inversely related to distance.



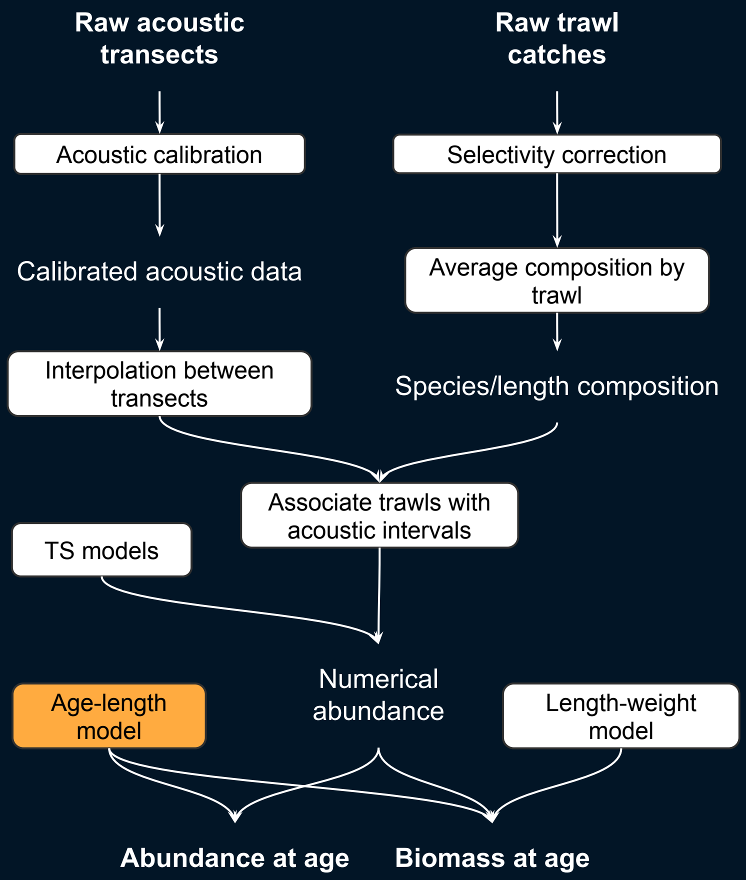
Bootstrapping procedure at each iteration:



Generate random length-TS function for pollock. TS uncertainty is 0.14 dB, about 3% in linear terms (Lauffenburger et al 2023).

For all other species, assume ± 3 dB uncertainty (100 % in linear terms–being conservative).

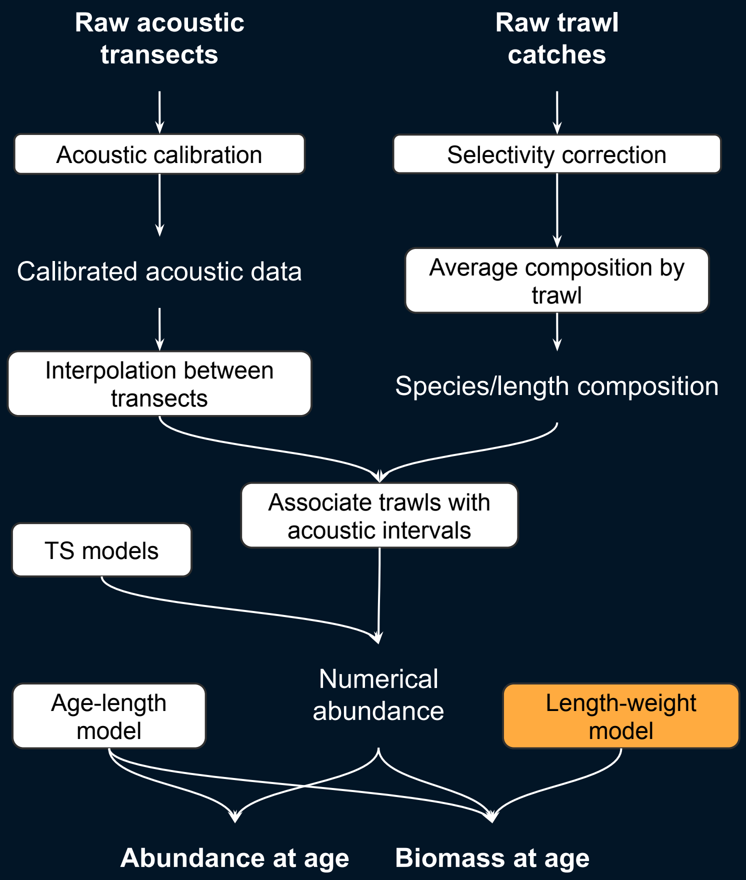
Bootstrapping procedure at each iteration:



Resample age data (otolith reads) from survey to get age composition.

Use to parameterize standard Gaussian mixture model for age-length key.

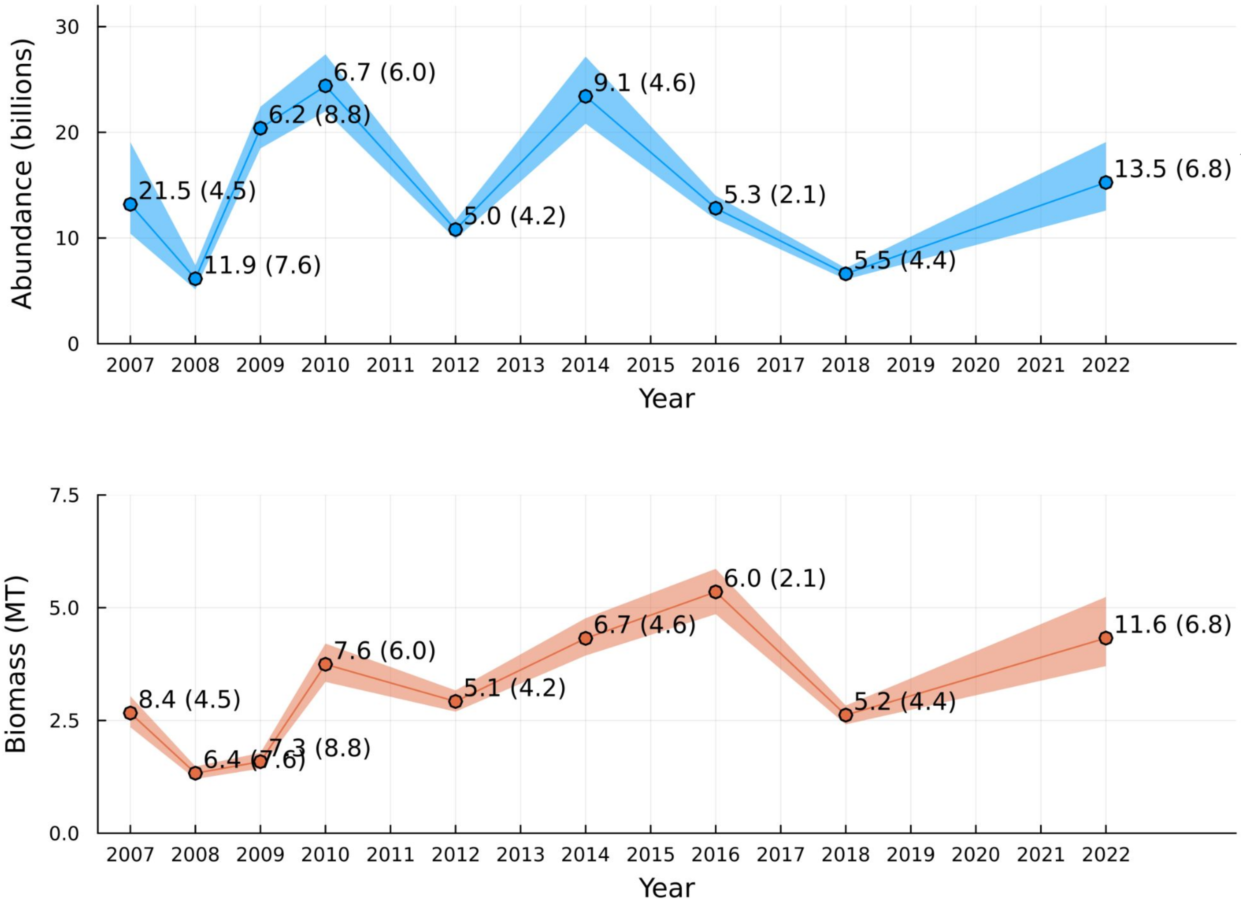
Bootstrapping procedure at each iteration:



Resample fish measured during survey with replacement.

Generate length-weight function based on resampled measurements (De Robertis and Williams 2008).

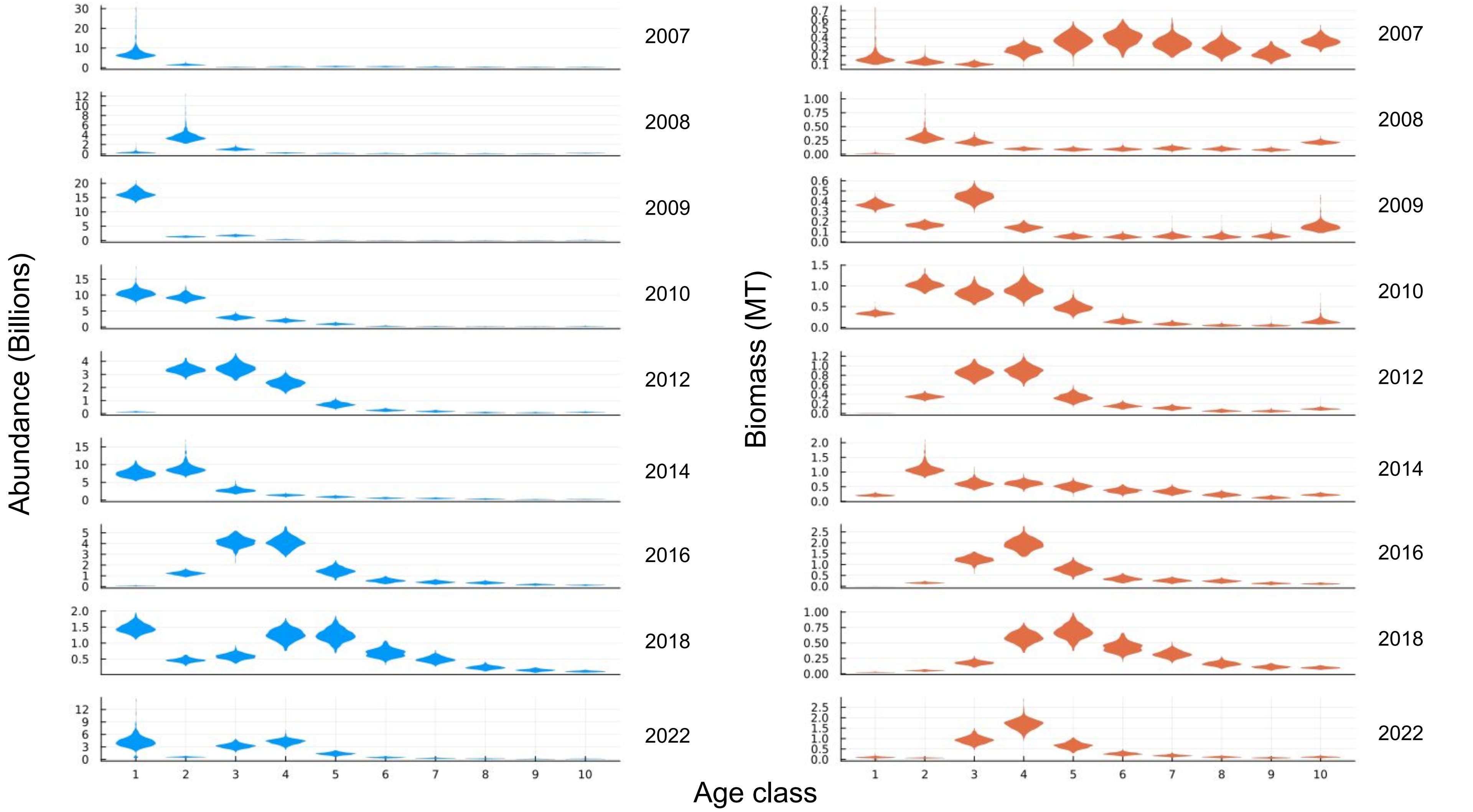
Total pollock abundance and biomass 2007-2022



CV (%) from:
Bootstrap (1D geostats)

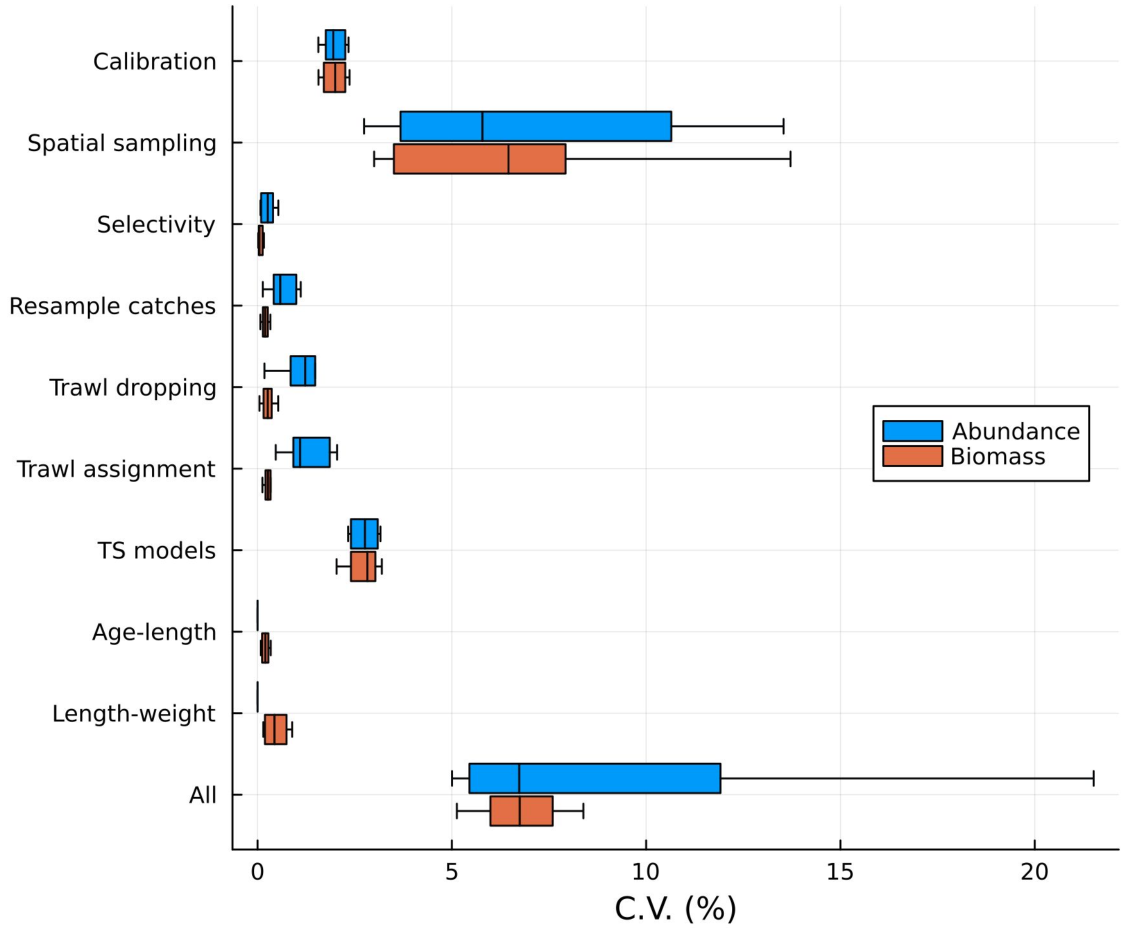
Median new CV / old CV:

- 1.57 for numbers
- 1.27 for biomass



Abundance (Billions)
Biomass (MT)
Age class
2007
2008
2009
2010
2012
2014
2016
2018
2022
1 2 3 4 5 6 7 8 9 10

Contributions of individual uncertainty sources



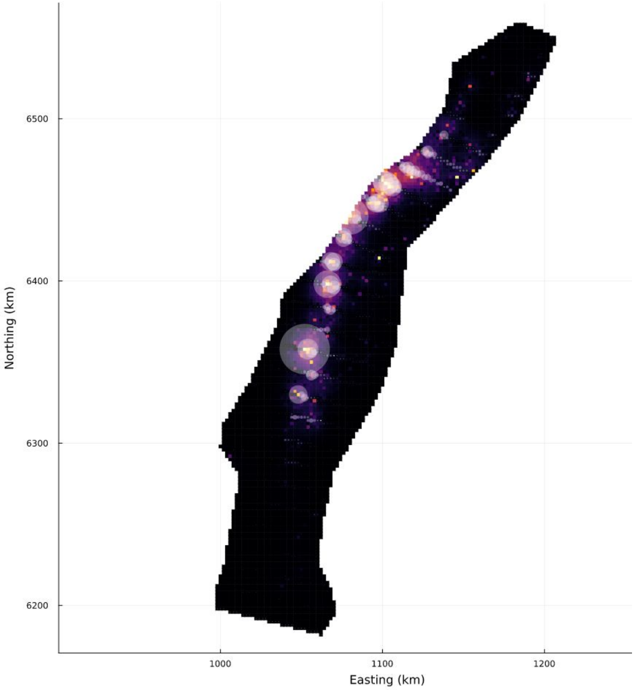
- Re-analyzed all surveys, turning on one component at a time
- Largest individual sources, on average:
  - Spatial sampling error
  - TS models
  - Echosounder calibration
  - Trawling-related sources (mostly for abundance)

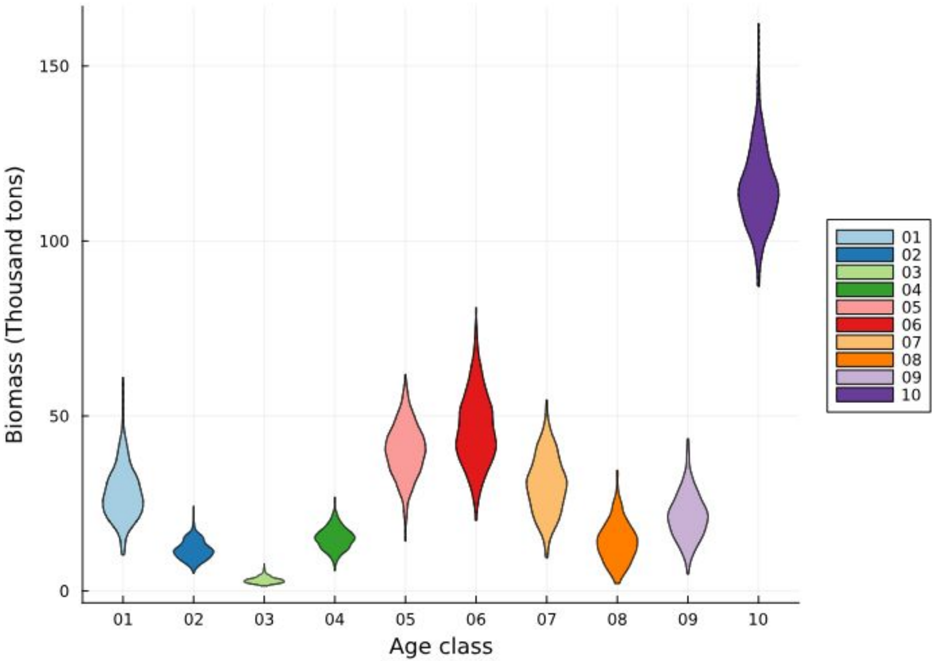
- Some differences year-to-year
- EBS is homogenous–trawling likely more important in GOA

Caveats

- A few sources of uncertainty remain unaccounted for:
  - Acoustic classification errors
  - Near-bottom acoustic dead zone
  - Survey domain/geographic availability
  - Fish movement
  - Some remaining questions about calibration and TS
- Currently, ignoring bottom 3 meters of water column (~25% of biomass)
- Uncertainty of absolute biomass vs. relative index
- Results may vary in other ecosystems, but…

Works in GOA too: Shelikof Strait, Winter 2023



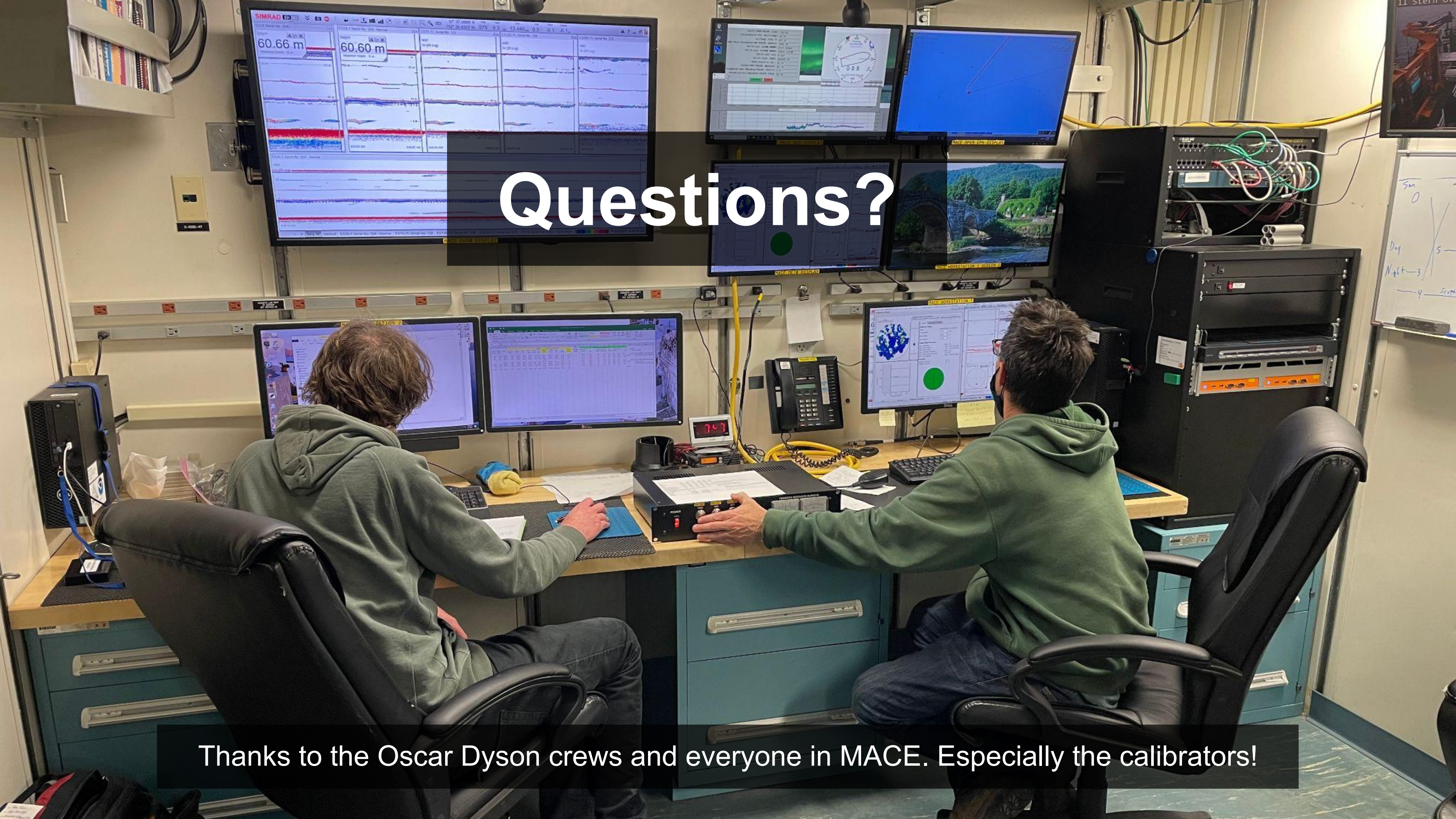


Age	1	2	3	4	5	6	7	8	9	10+
CV (%)	27.1	24.2	30.2	21.2	18.8	12.9	27.3	41.3	31.1	9.6

Conclusions

- Total biomass uncertainty for MACE EBS pollock surveys is typically 5-11%
  - For individual age classes, 10-30%
- Spatial sampling error, TS, and calibration are main sources
  - For less abundant age classes/species, uncertainty may be higher/have different drivers
  - More transects = less uncertainty
- On average 1.9 and 1.5 x 1D geostatistical estimates for numbers and biomass
- 2-4 times smaller than assumed in stock assessment
- Framework can be used to think about effort allocation/reduction

Questions?



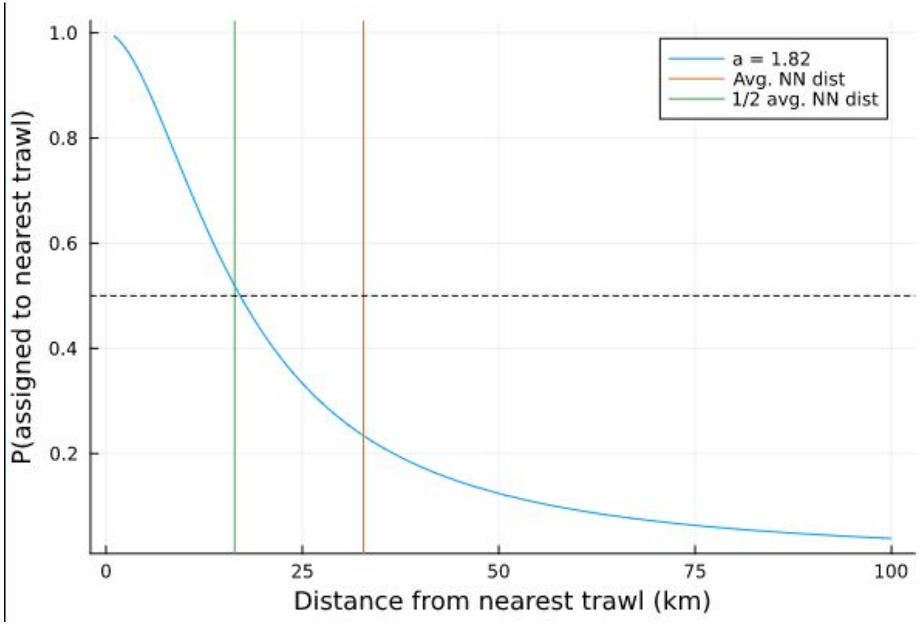
Thanks to the Oscar Dyson crews and everyone in MACE. Especially the calibrators!

LUNGS details and the “Cholesky trick”

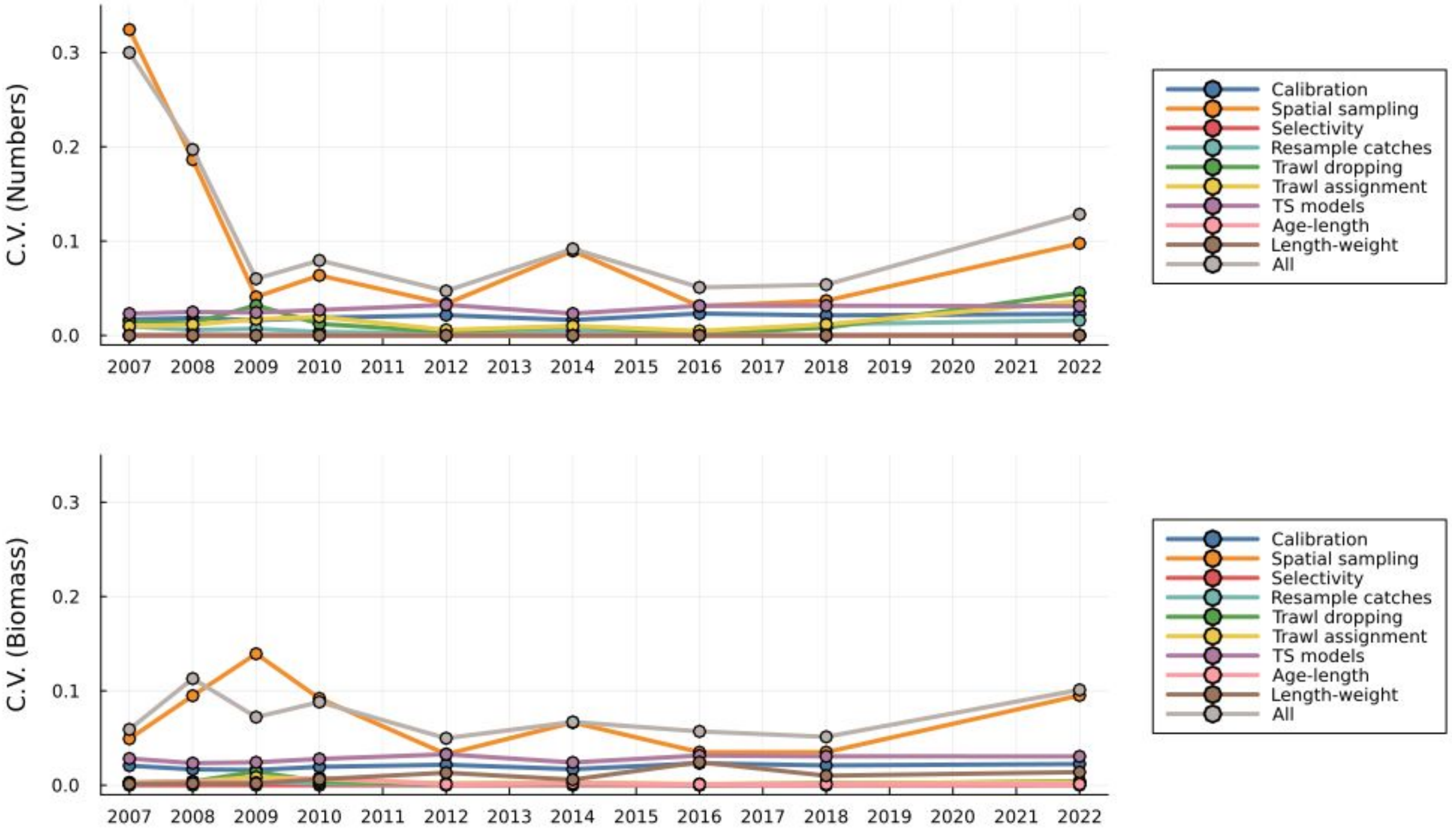
- Normal “lower-upper Gaussian simulation” (LUGS):
  - Use data + variogram model to define mean (μ_x) and covariance (Q) for simulation locations
  - Cholesky (LU) decomposition of covariance matrix $Q = L\ L^{-1}$
  - If the vector $\boldsymbol{z}$ is i.i.d standard normal, product $\boldsymbol{x} = L\ \boldsymbol{z}$ will have covariance Q
  - Q = cov(x) = E[$x\ x'$]
  - cov(z) = $E[z\ z'] = I$
  - $Q = L\ L' = L\ I\ L' = L\ E[z\ z']\ L' = E[(L\ z)\ (L\ z)']$
  - From definition of covariance, cov($L\ z$) = Q
- But, $\boldsymbol{z}$ does not have to be normal–as long as var($\boldsymbol{z}$) = 1.0, cov($L\boldsymbol{z}$) = Q
  - Get means to match: $\mu_z = L^{-1}\ \mu_x$
  - Know required mean and variance (1.0) for each element of $\boldsymbol{z}$, can then translate into parameters for whatever non-negative distribution you want

Trawl shuffle details

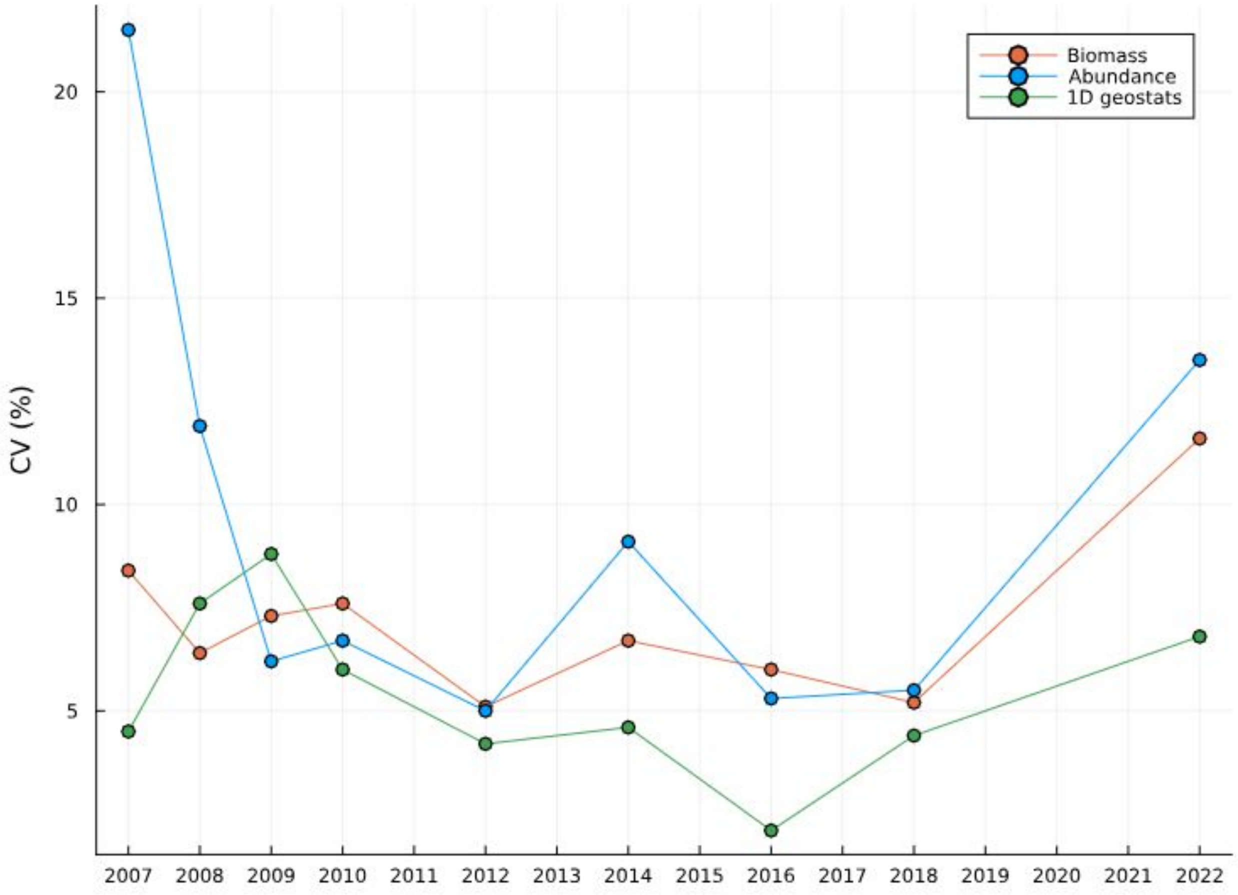
- Randomly assign each acoustic cell to a trawl
- Inverse distance weighted: $1 / d^a$
- Exponent *a* set so average pixel has 50/50 chance of getting nearest trawl when ½ distance to nearest neighbor



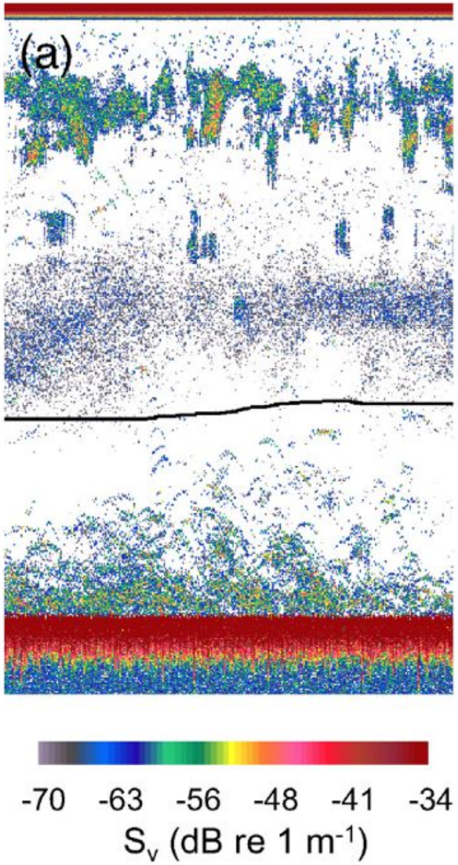
Individual error contributions through time



Coefficients of variation over time



Separating backscatter into "scaling strata"



Scaling Stratum 2
(small, dense schools)

Scaling Stratum 1
(continuous
near-bottom
backscatter)