Multispecies survey design optimization for the Gulf of Alaska Groundfish Bottom Trawl Survey

Lewis Barnett*, Zack Oyafuso, Stan Kotwicki

NOAA Fisheries, AFSC, RACE Division, Groundfish Assessment Program

Plan Team Meeting 9/11/2020

Motivation and outline

- Goals
  - Increase flexibility and efficiency of stratified random survey design
  - Obtain accurate and precise estimates of abundance indices and their variance for most assessed stocks
- Outline of simulation approach
  - **operating model ->**
  - **survey optimization ->**
  - **expected performance**

Key current and proposed design elements

	Current
how many strata?	59 (55 without deep strata)
strata characteristics	depth, terrain, reporting areas
allocation criteria	Neyman: f (B, value, cost, s^2, area)
pre-specified constraints?	sample size
how many species?	52-57 species

- Potential improvements in the face of survey effort fluctuations
  - Better estimates of stratum variances with fewer strata

Key current and proposed design elements

	Current	Proposed
how many strata?	59 (55 without deep strata)	5-20 strata
strata characteristics	depth, terrain, reporting areas	depth, longitude
allocation criteria	Neyman: *f* (B, value, cost, s^2, area)	Bethel: *f* (σ^2, expected CV)
pre-specified constraints?	sample size	expected CV
how many species?	52-57 species	15 species groups

- Potential improvements in the face of survey effort fluctuations
  - Better estimates of stratum variances with fewer strata
  - Quantify expected precision and tune according to needs
  - More flexibility in species prioritization

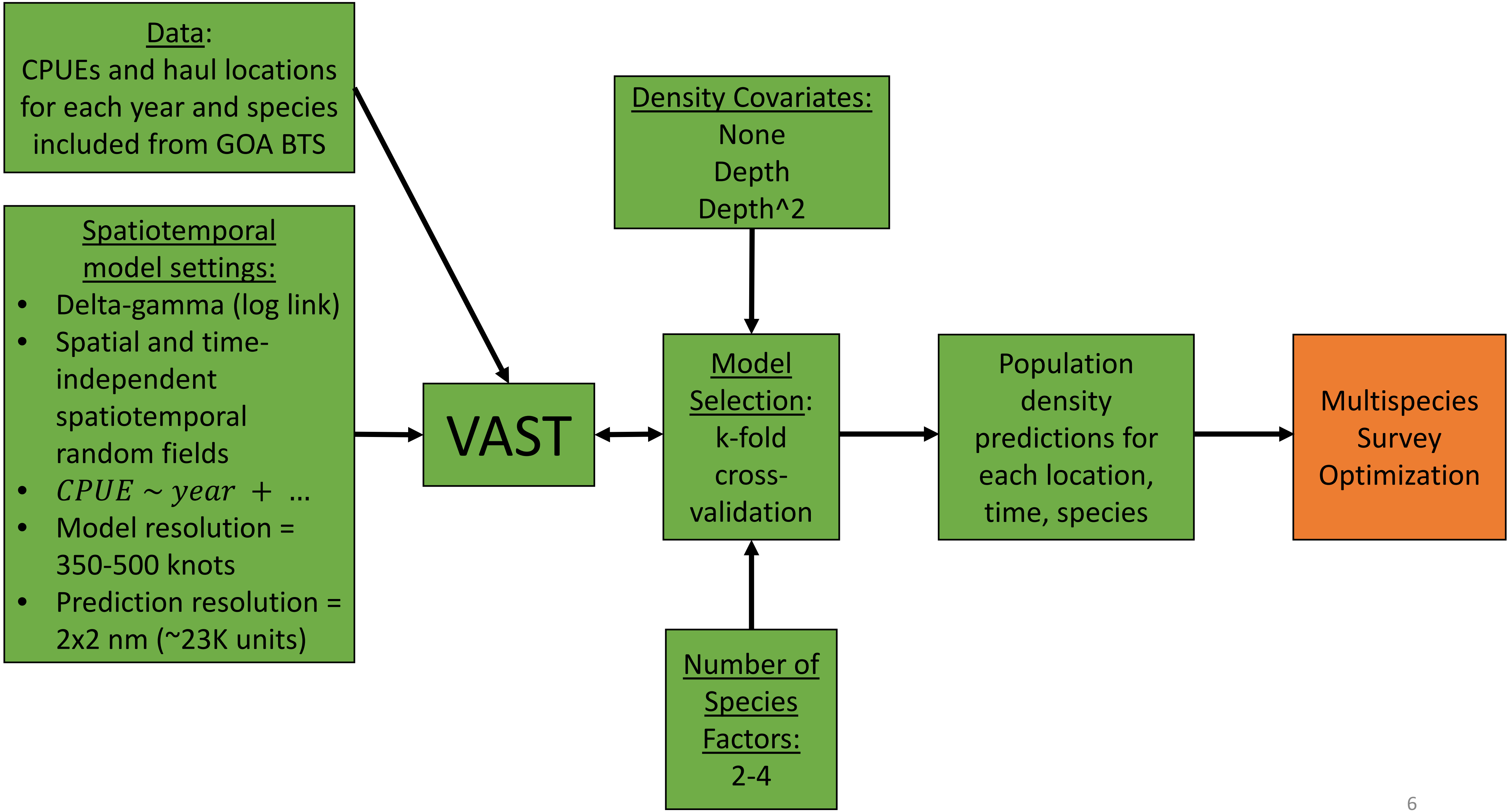
Operating model: data inputs

- Gulf of Alaska BTS
  - CPUE, haul location, bottom depth, year
  - 11 survey years 1996-2019
  - 15 species/complexes

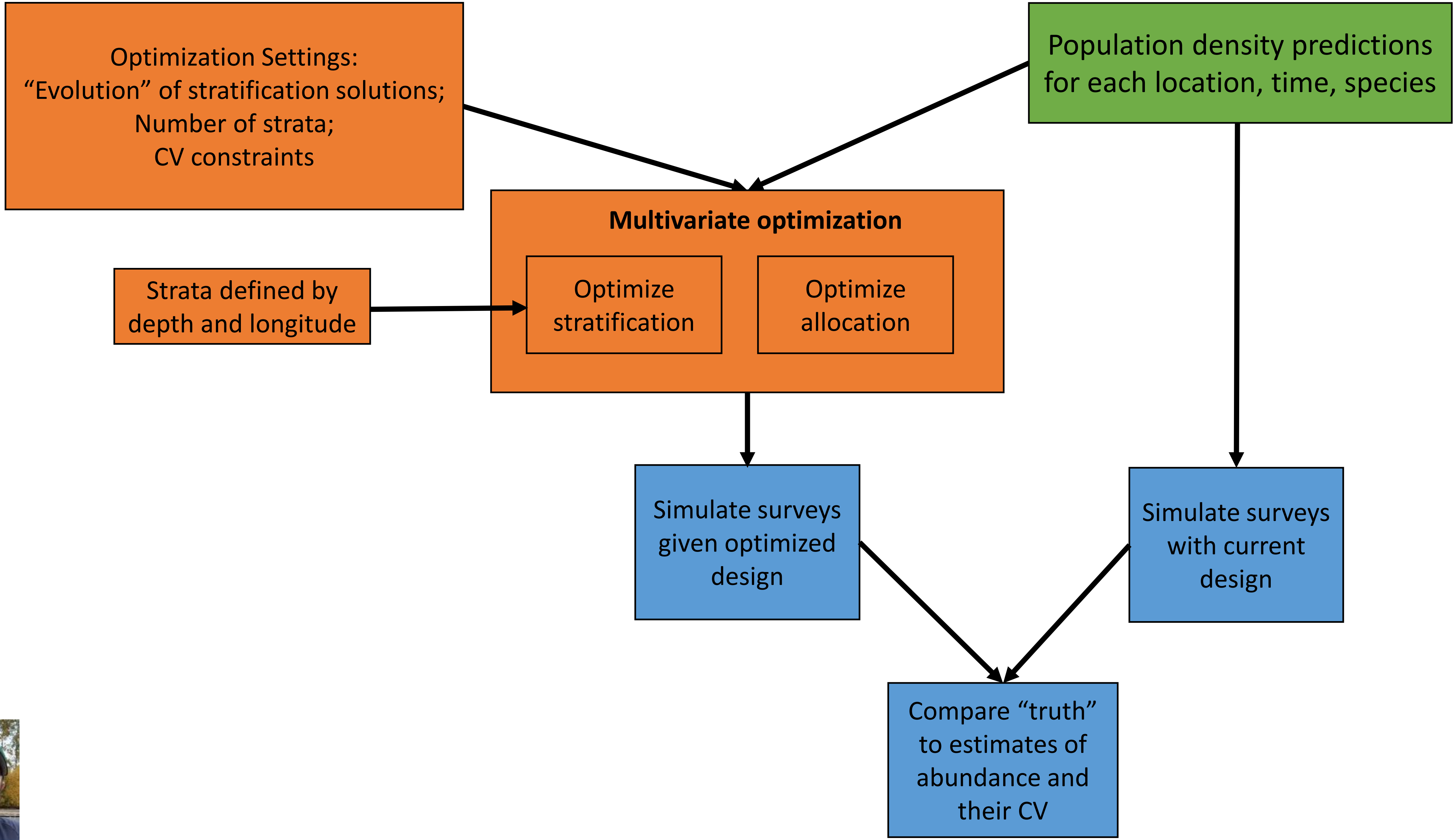
<u>Species Groups</u>

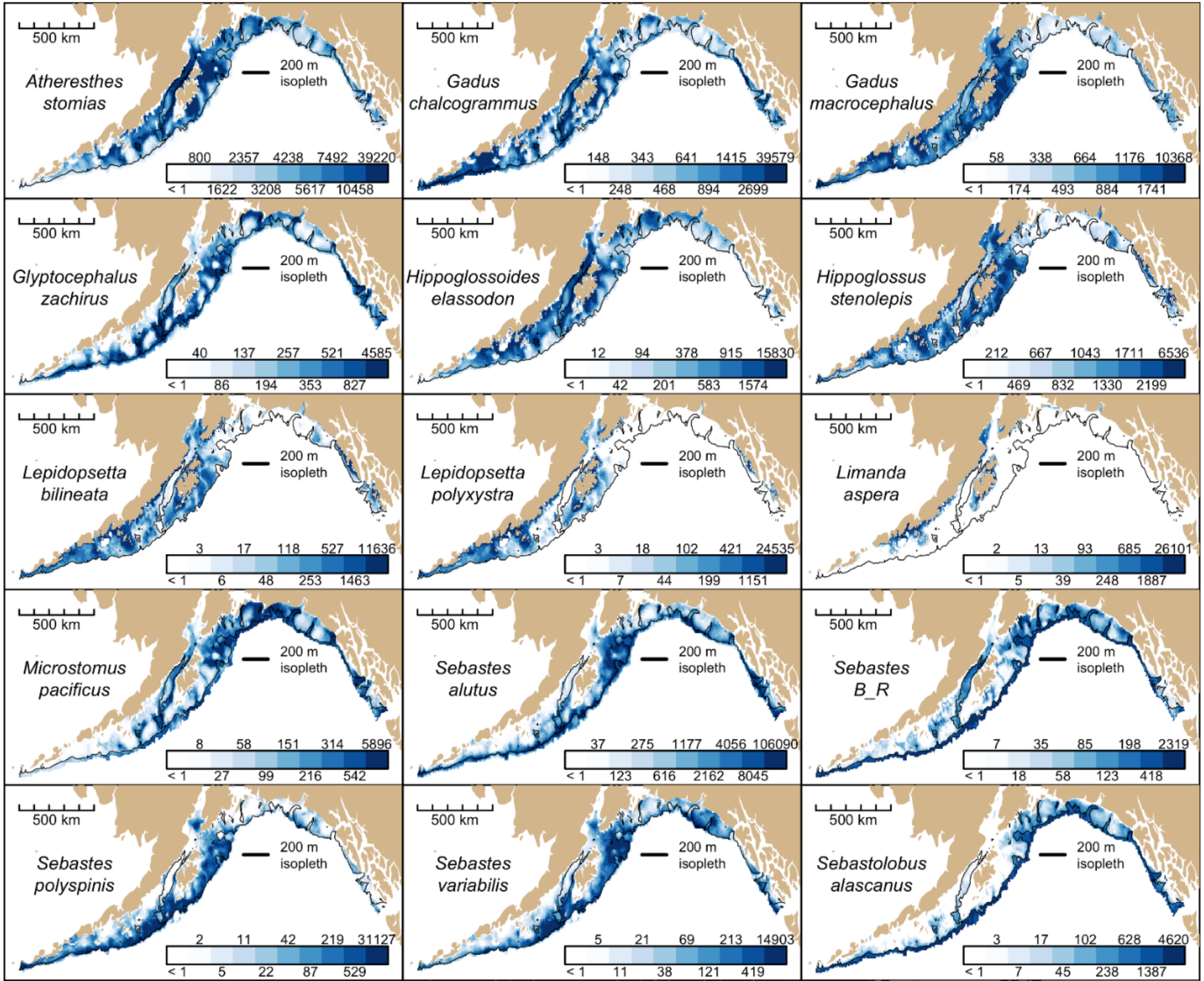
Pacific Ocean Perch
Arrowtooth Flounder
Pacific Cod
Walleye Pollock
Pacific Halibut
Rex Sole
Flathead Sole
Dover Sole
Northern Rock Sole
Southern Rock Sole
Dusky Rockfish
Northern Rockfish
Rougheye/Blackspotted Rockfish
Shortspine Thornyhead
Yellowfin Sole or Silvergray Rockfish

Project overview: operating model -> survey optimization



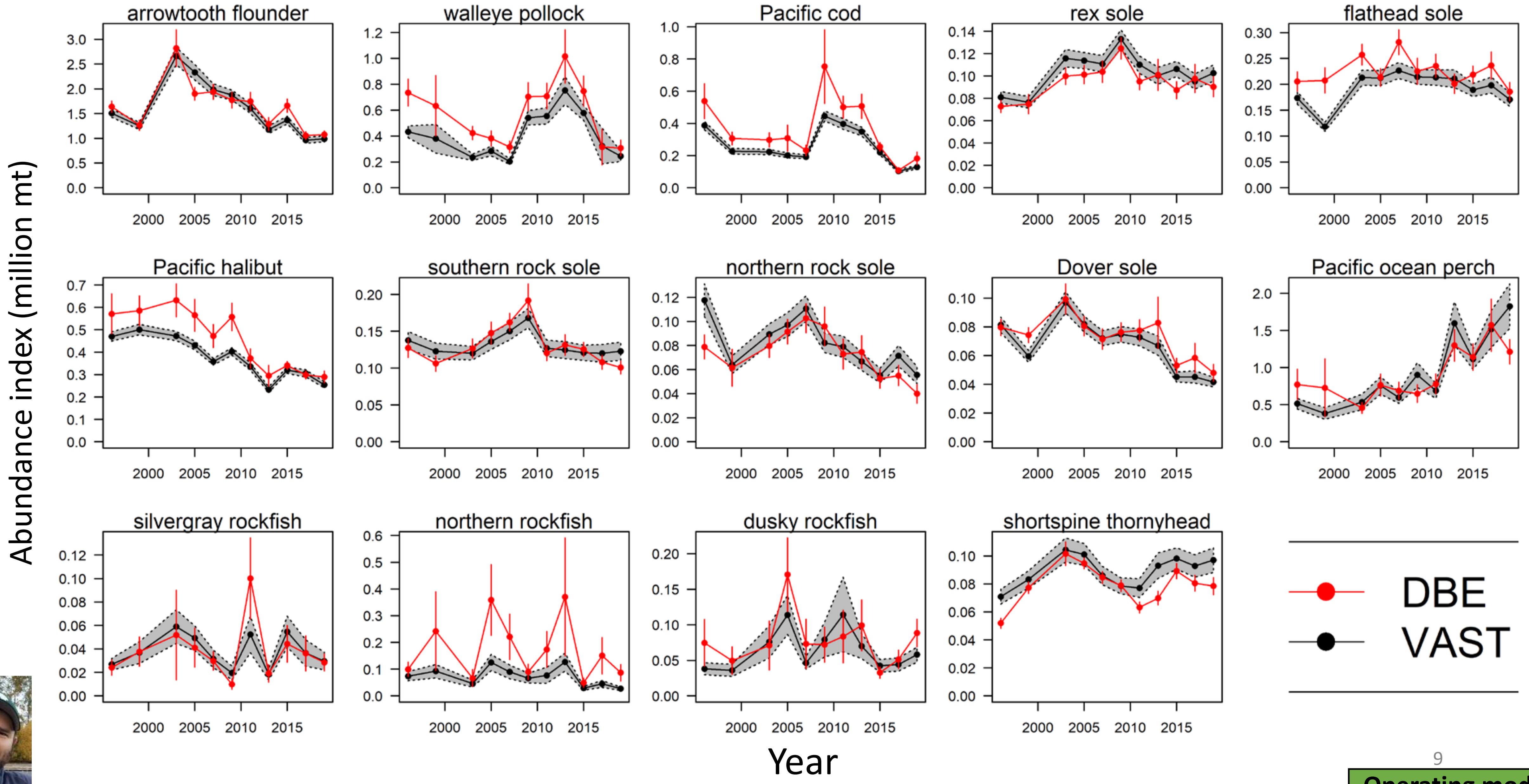
Project overview: operating model -> survey optimization -> expected performance





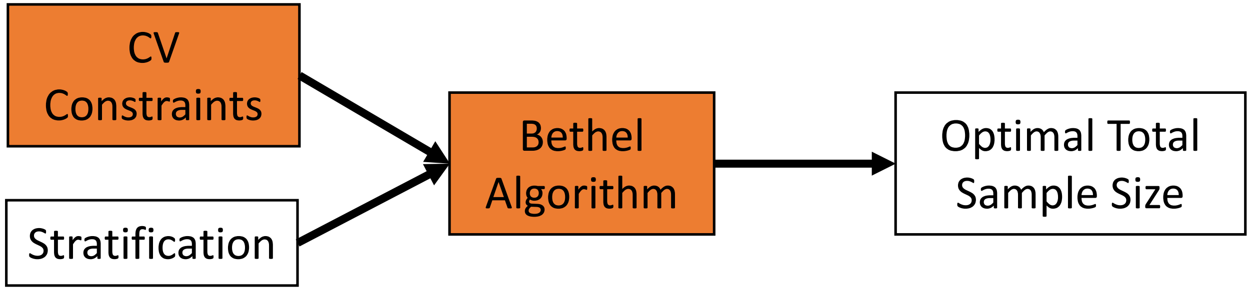
Mean population density predictions (1996-2019)

Biomass trends are similar between design- and model- based estimates



Operating model

Optimization methods



$$\min \sum_{h=1}^{H} n_h$$

$$s.t.$$

$$CV(Y_1) < U_1$$

$$\dots$$

$$CV(Y_G) < U_G$$

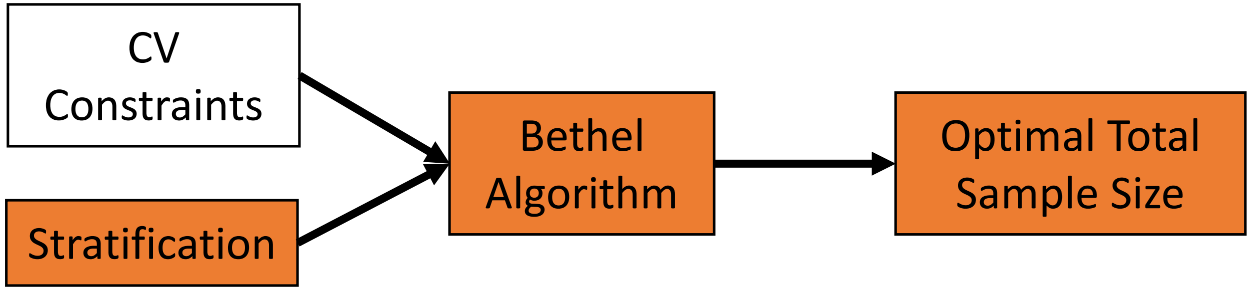
- n_h: samples allocated to stratum h
- U_g: upper CV constraint of species g
- Y_g: mean density of species g, of G total species

$$Y_g = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right)(Y_{g,h})$$

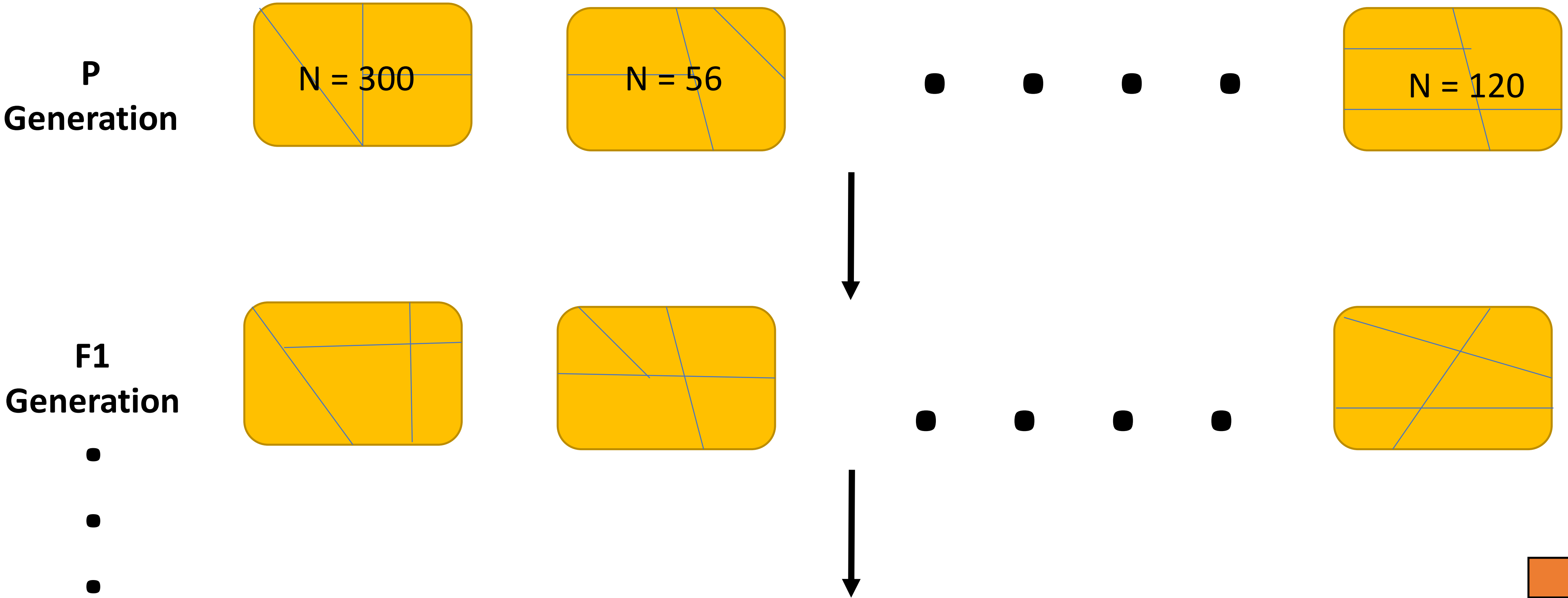
$$Var(Y_g) = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right)^2 \frac{S_{h,g}^2}{n_h}(1 - \frac{n_h}{N_h})$$

- $S_{h,g}^2$: variance of of species g in stratum h
  - Can incorporate either spatial variability only, or both spatial and temporal variability
- N_h: total number of units in stratum h

Optimization methods



- Complete search not feasible, so we perform a partial search using a **genetic algorithm**



Estimation steps

- Objective: calculate expected uncertainty and bias in abundance index for optimized and current design
  - How does this change as a function of
    - sampling effort?
    - number of strata?
    - location of strata boundaries?

Expected performance

Estimation steps

- Objective: calculate expected uncertainty and bias in abundance index for optimized and current design
  - How does this change as a function of
    - sampling effort?
    - number of strata?
    - location of strata boundaries?

- Computation: simulate D=1000 surveys, compare abundance index precision and accuracy relative to their true value, for each species and year (using mean density as a proxy for total abundance)
  - Precision of estimated mean:

$$CV_{true} = \frac{\sqrt{D^{-1} \sum_{d=1}^{D} (y_d - \bar{y})^2}}{Y}$$

  - Relative bias of estimated mean:

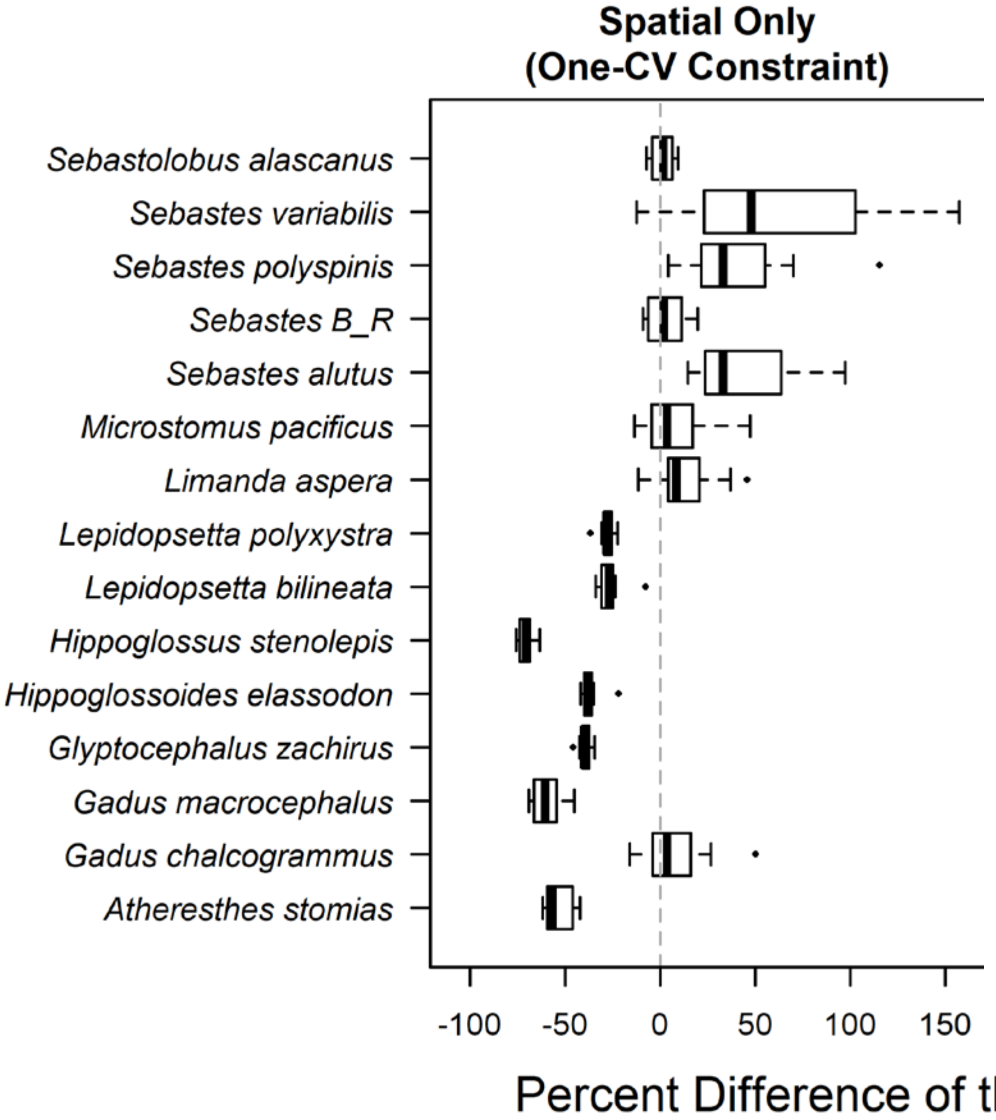
$$RB(y) = 100\% \frac{\sum_{d=1}^{D} (y_d - Y)}{D\,Y}$$

  - Accuracy of uncertainty estimate:

$$RRMSE(CV) = \frac{\sqrt{D^{-1} \sum_{d=1}^{D} (CV_d - CV_{true})^2}}{\overline{CV}}$$

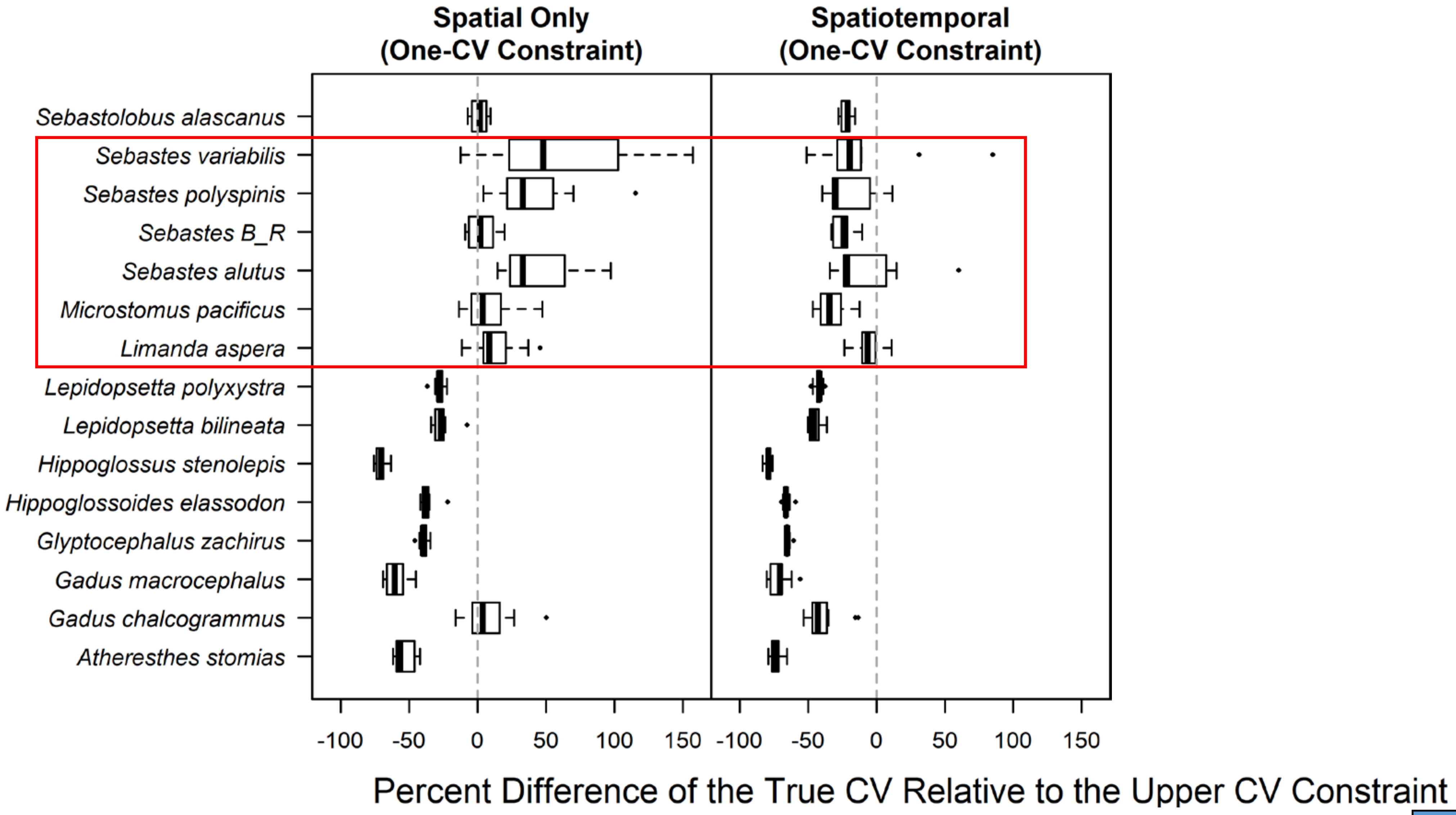
Expected performance

Spatial CV constraint does not cap uncertainty



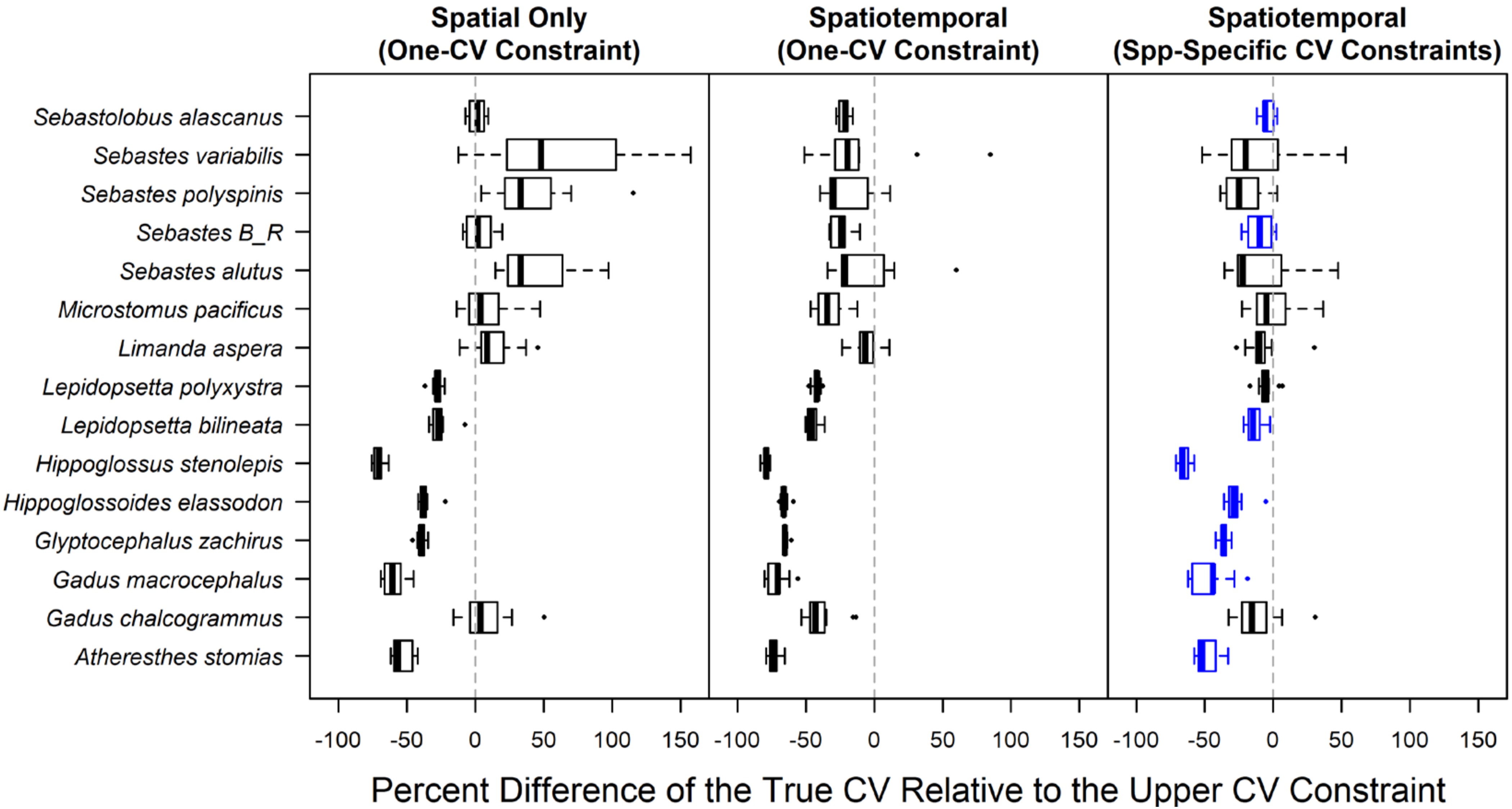
Expected performance

Spatiotemporal CV constraint caps uncertainty



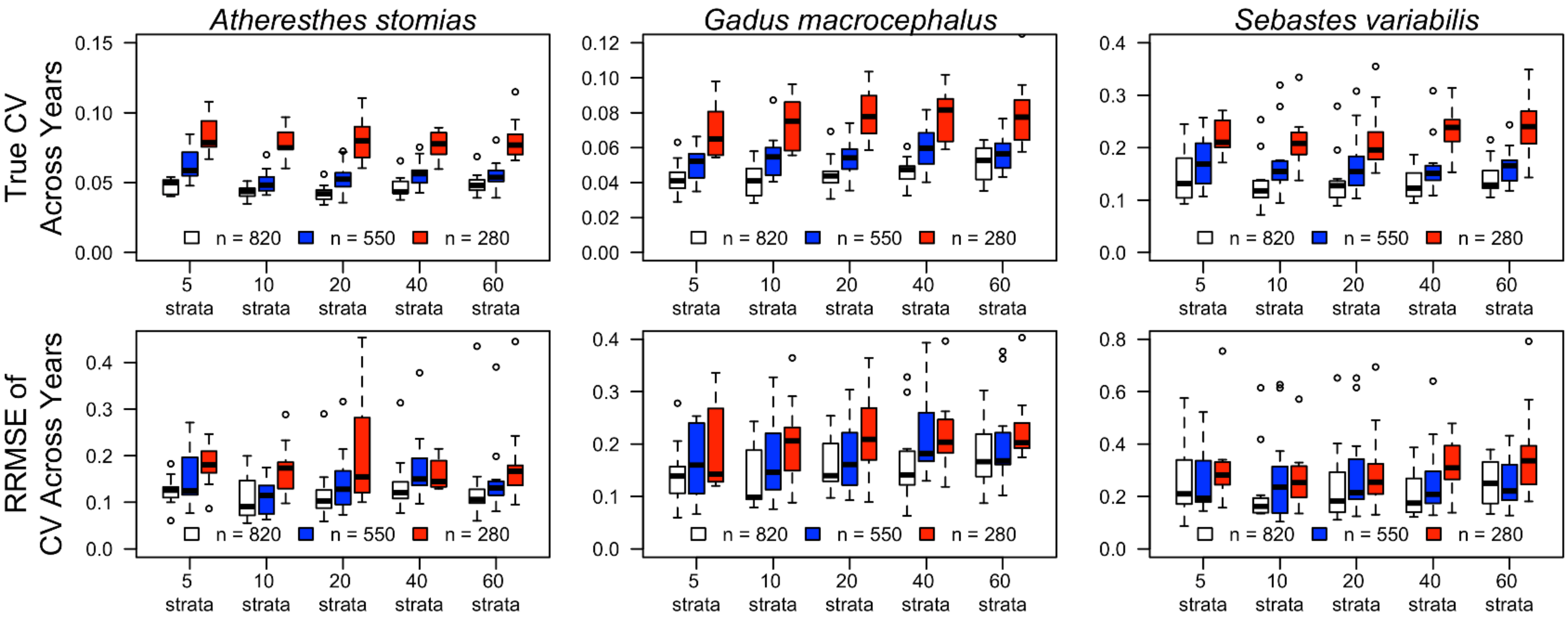
Expected performance

Species-specific CV constraints improve efficiency



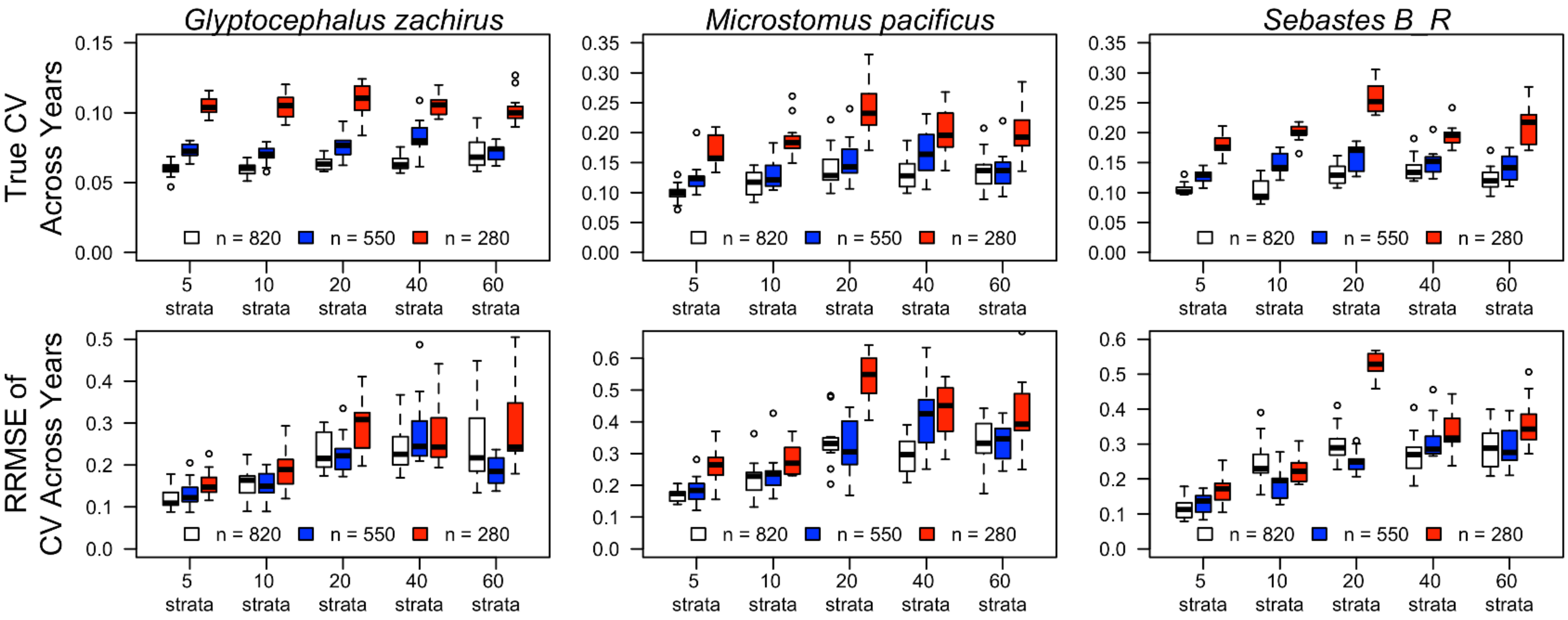
Expected performance

Precision and accuracy of index uncertainty estimates differs little with N strata for some species



Expected performance

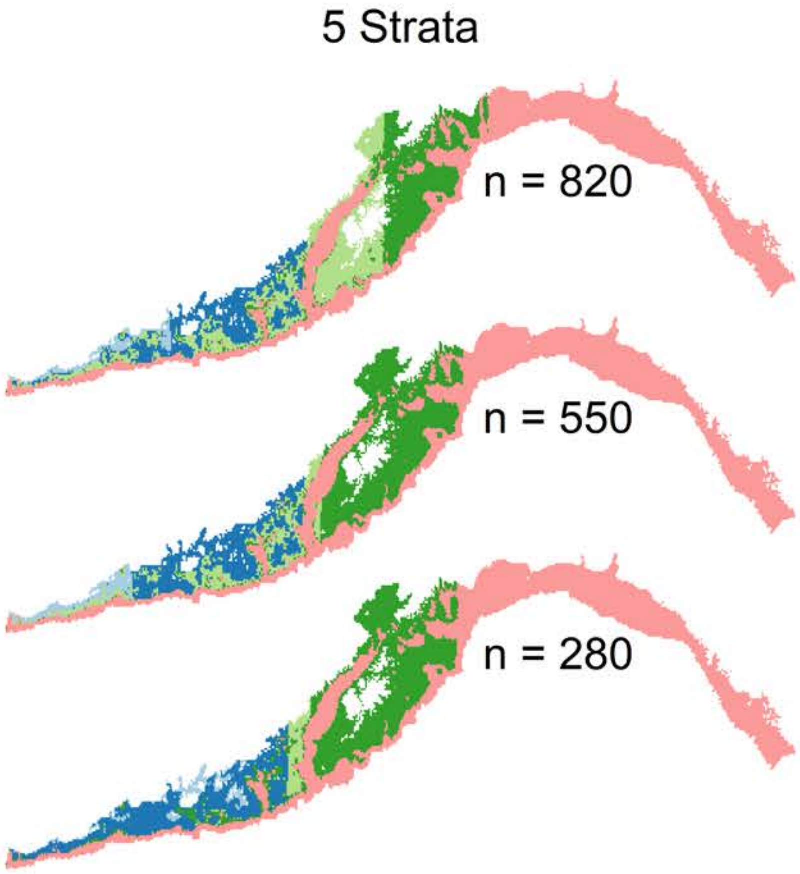
Accuracy of index uncertainty estimates *decreases* with N strata for some species



Thus we favor solutions for fewer than ~ 20 strata

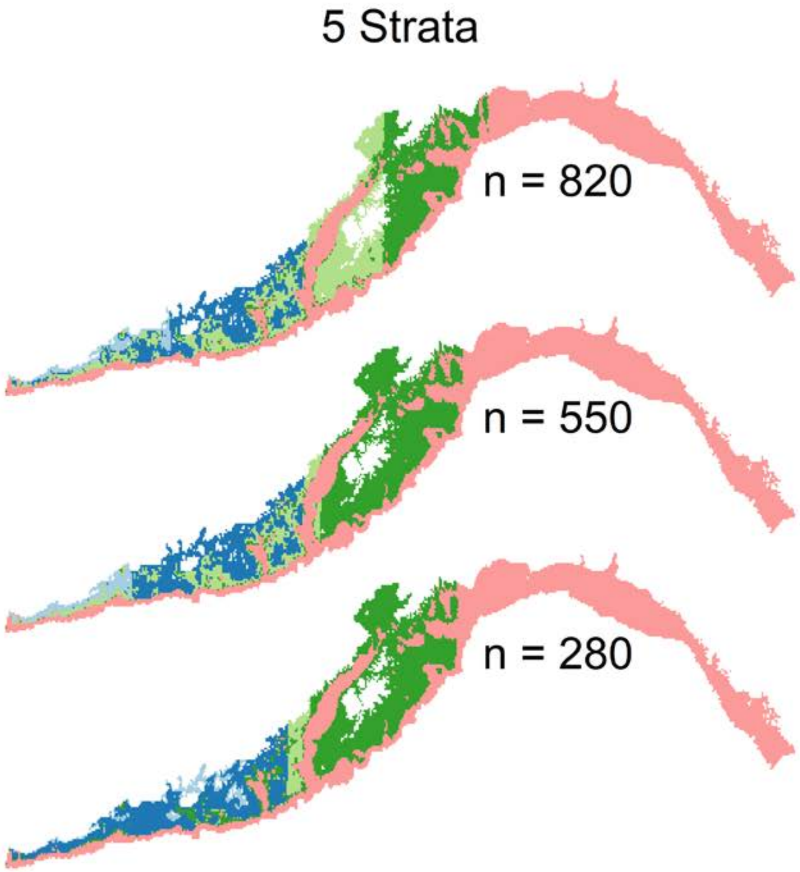
Expected performance

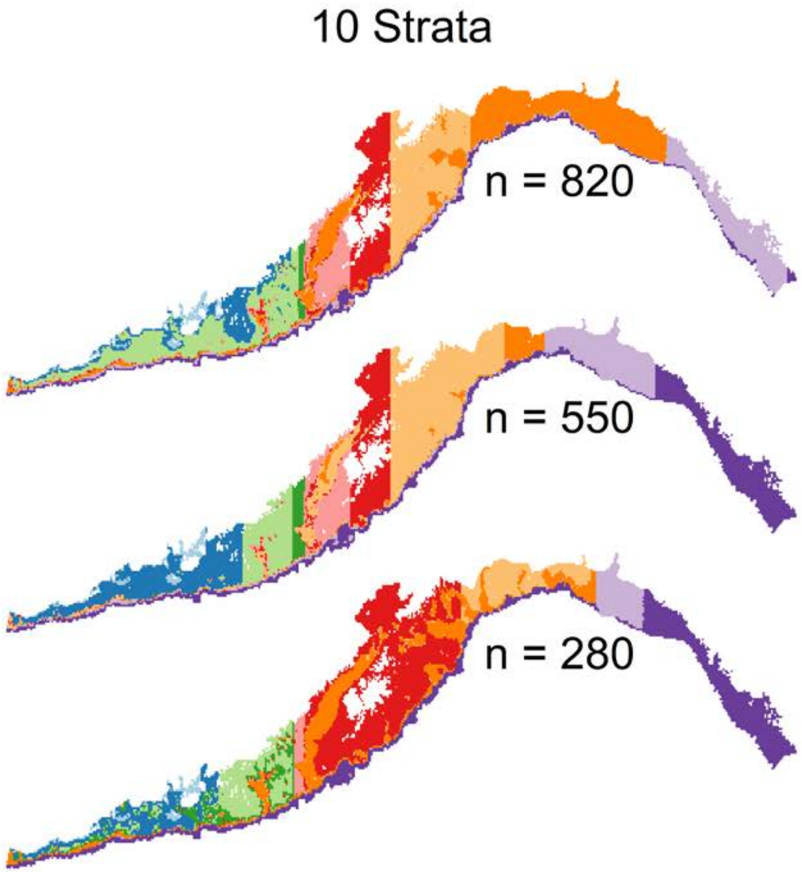
Optimal strata boundaries given sample size and N strata



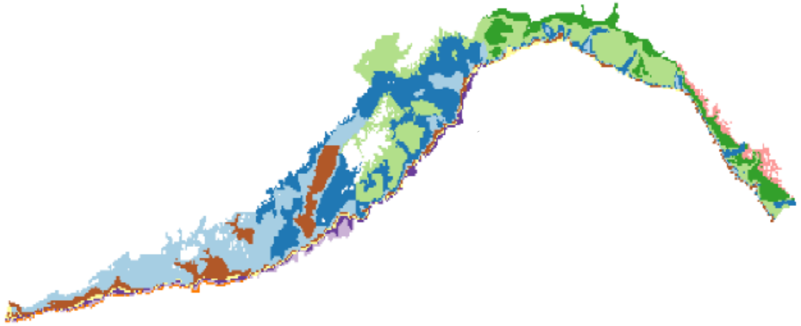
Expected performance

Optimal strata boundaries given sample size and N strata



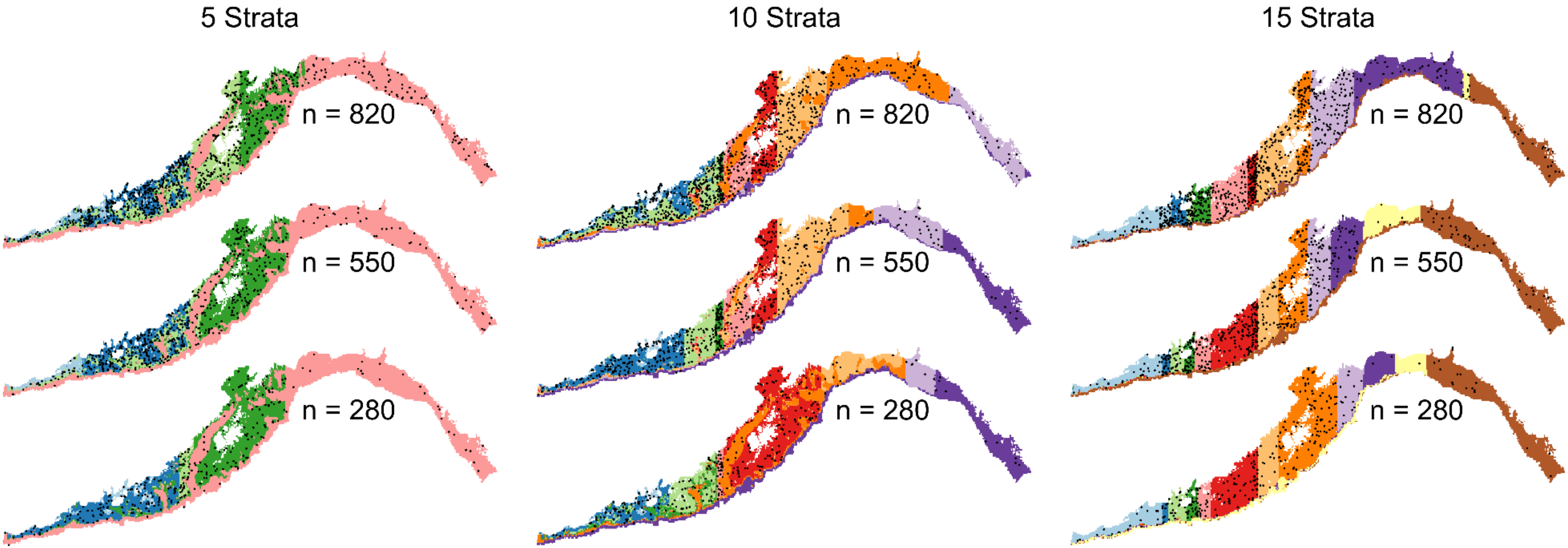


Current Strata



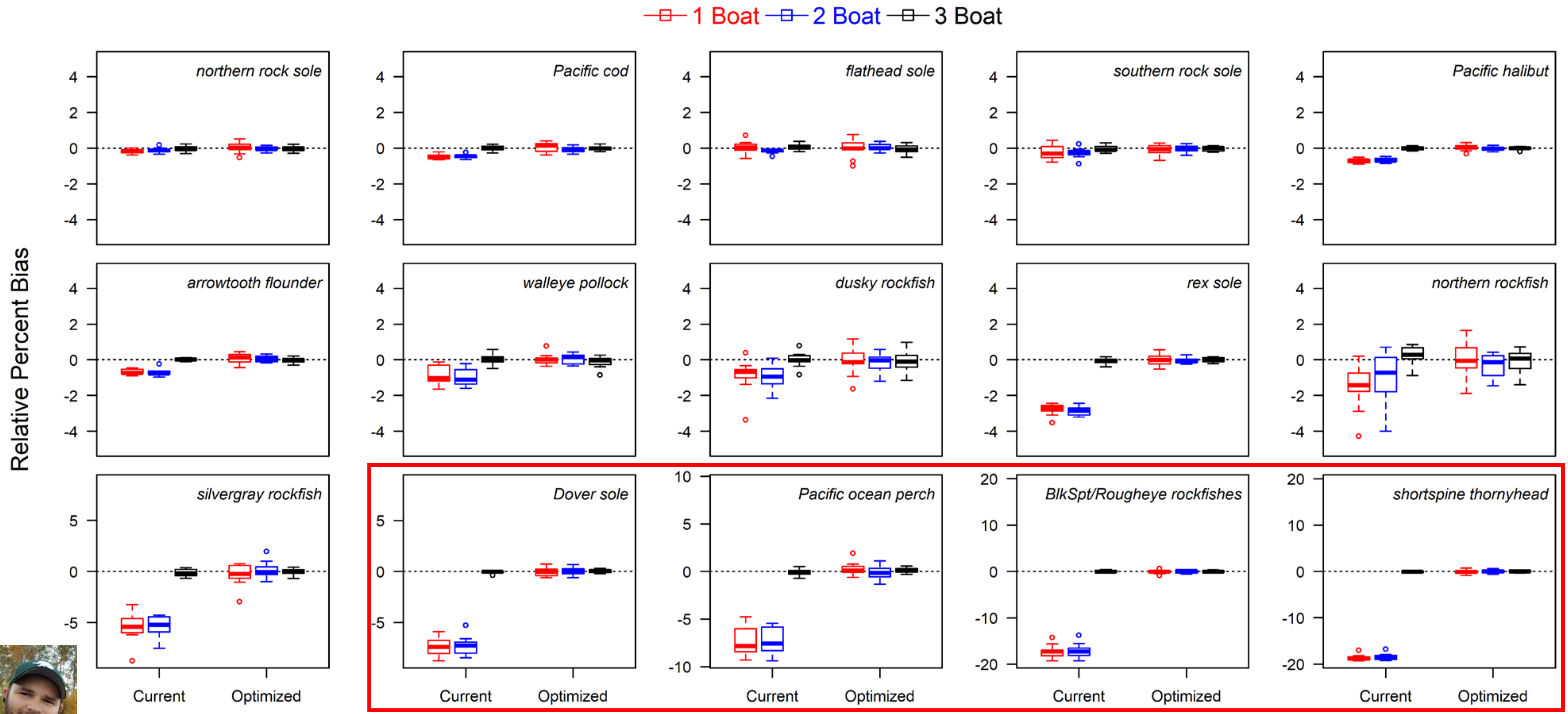
Expected performance

Optimal strata with examples of optimal sampling density



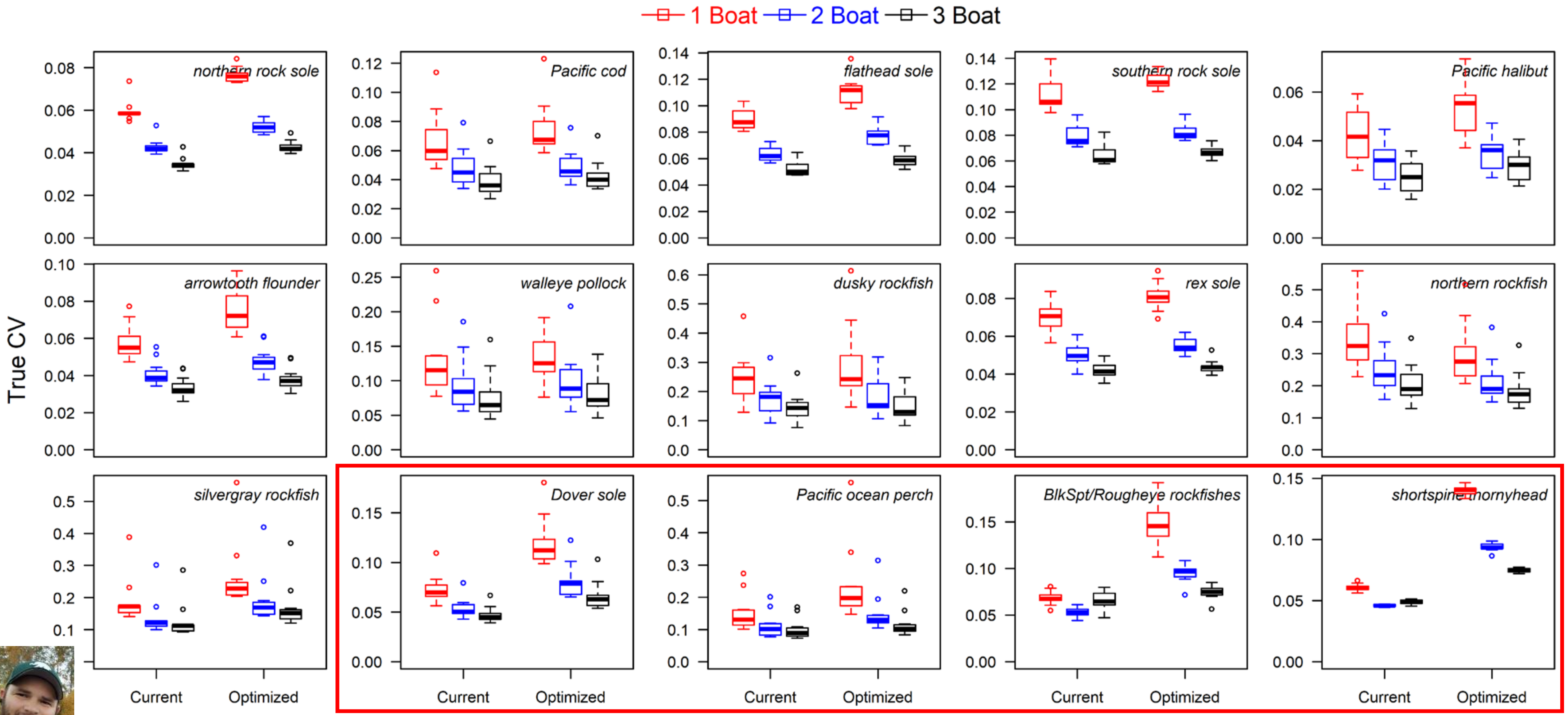
Optimal allocation indicates highest sampling density in western GOA where biomasses of many species are highest

Current vs. optimal design: optimization mitigates bias



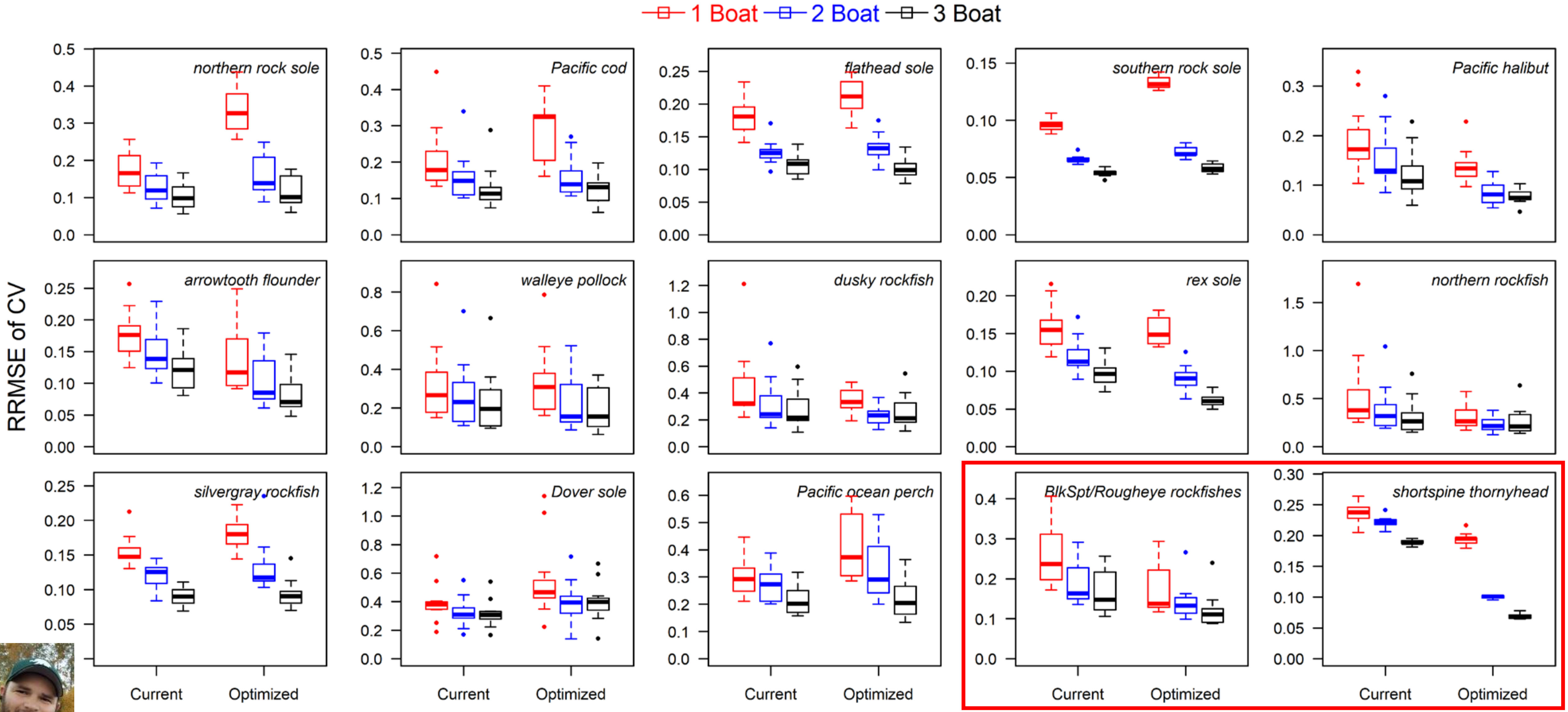
Expected performance

Precision is similar but sometimes lower than current



Expected performance

Accuracy of uncertainty estimate is similar or improved



Expected performance

Summary: advantages of optimized design

- Can design a survey to meet user-specified precision constraints
- Improve abundance index estimation
  - Reduce bias (important for tier 5 stocks)
  - May increase accuracy of uncertainty estimate (important for data-weighting)
- Improved flexibility of surveys given modular approach
  - Can generate population density predictions with multiple methods
  - Enabling quick, data-driven decisions on where to cut samples when necessary
    - Can use to optimize allocation among existing strata (fast!)

Future and ongoing work

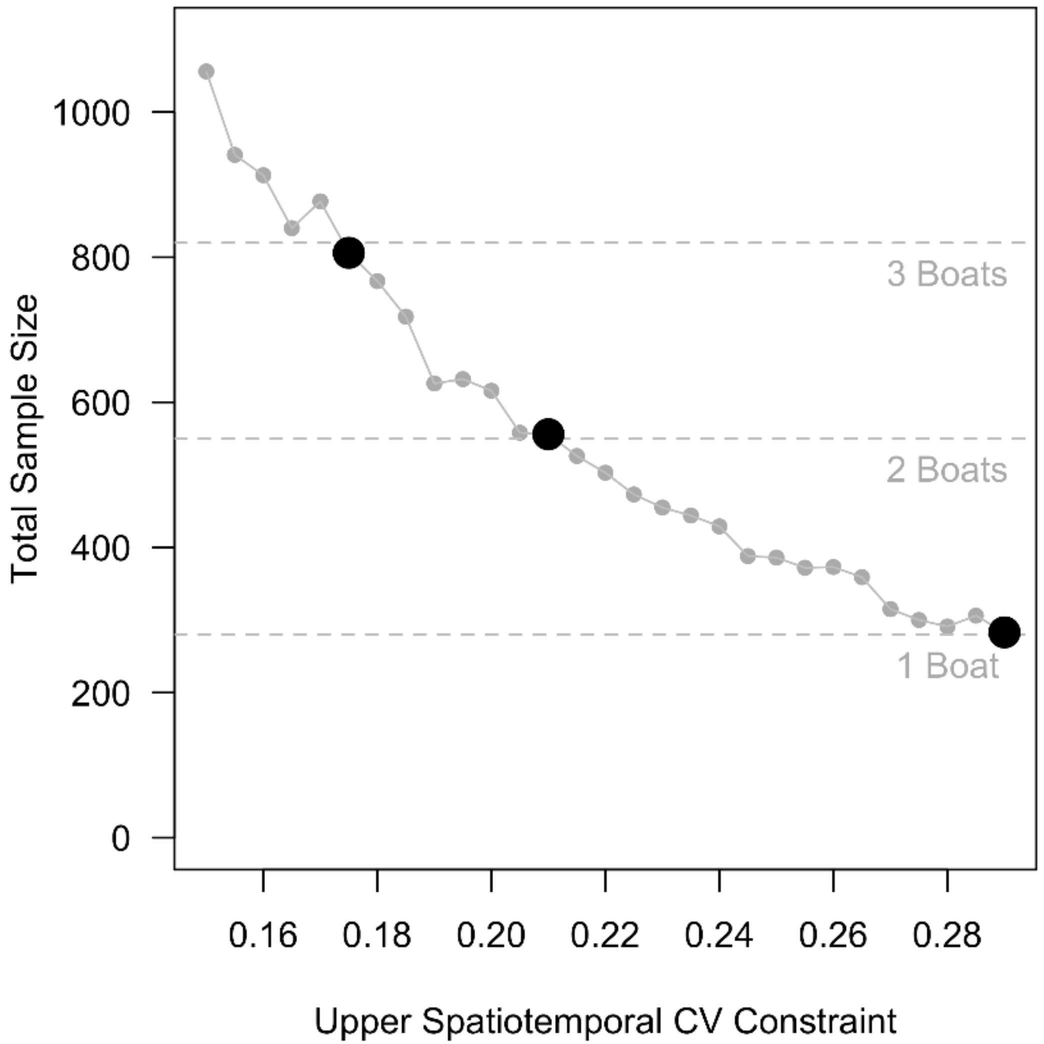
- Operating model
  - Species included
  - Improving predictive skill
    - Multivariate vs. univariate models
    - Covariate nuances
  - Exclude untrawlable habitat?
  - Simulate replicate predictions?
- Optimization
  - Species-specific CV constraints
  - How often to update?
- Tactical adjustments post hoc
  - Adjust strata based on expert knowledge or other analyses
  - Practicalities (e.g., travel and sampling duration)

Questions?
Comments?
Requests?

Supplementary methods: Optimization steps

- Find optimal solution for each increment of CV constraint

- Find solution with closest sample size to the 1, 2, or 3 “boat” effort scenarios (black points)

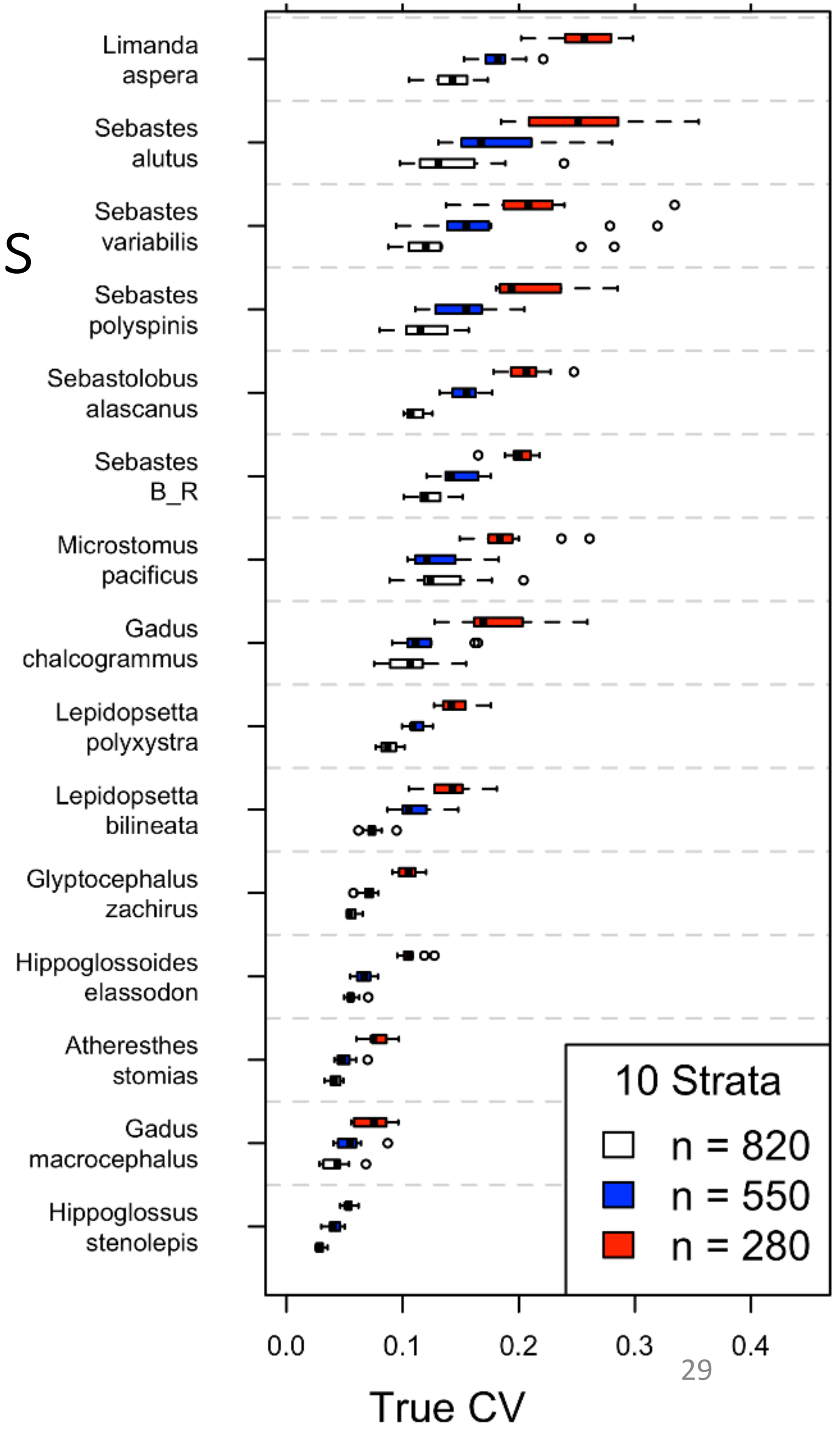


Optimization

Supplementary results:
True CV varies among species and sample sizes

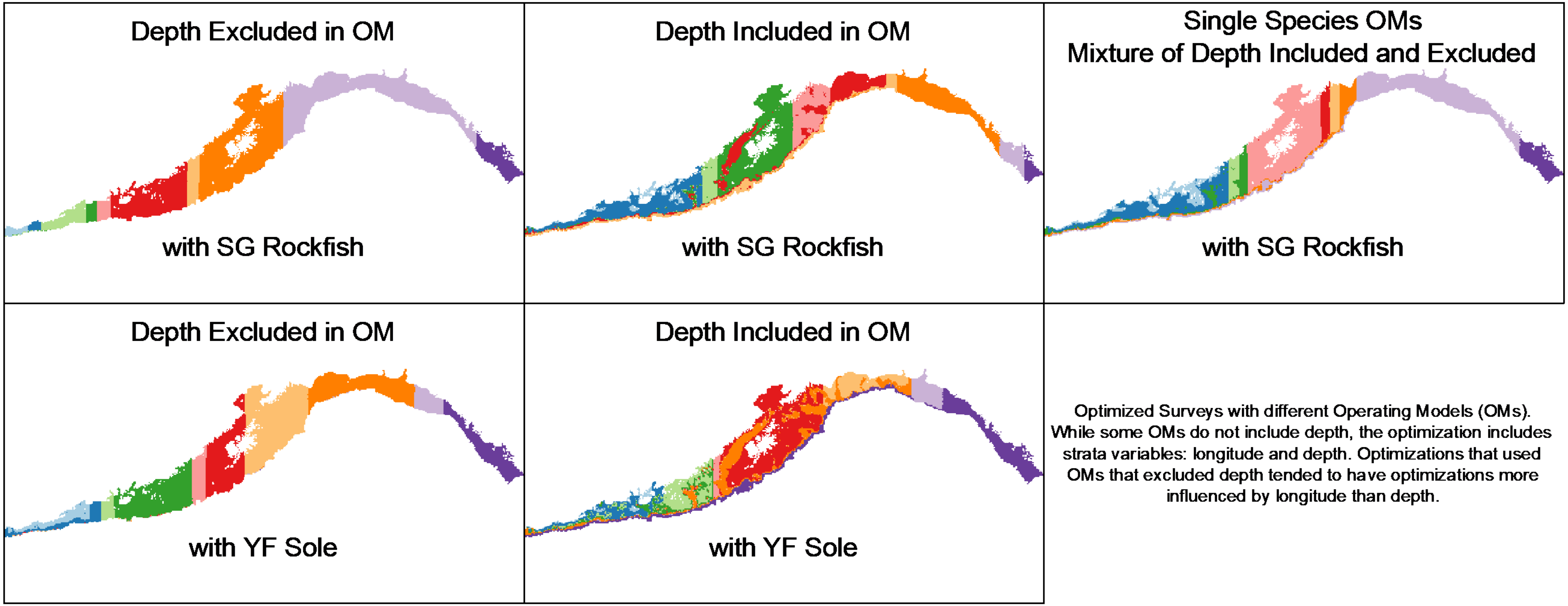
- Species with higher uncertainty
  - Rockfishes
  - Flatfishes with constrained or patchy distributions

- Some species are well-sampled even at low sampling intensity
  - Abundant and diffusely-distributed flatfishes



Expected performance

Supplementary results: Sensitivity to operating model structure



Optimized Surveys with different Operating Models (OMs). While some OMs do not include depth, the optimization includes strata variables: longitude and depth. Optimizations that used OMs that excluded depth tended to have optimizations more influenced by longitude than depth.

Few strata are defined by depth when not included as a covariate in OM