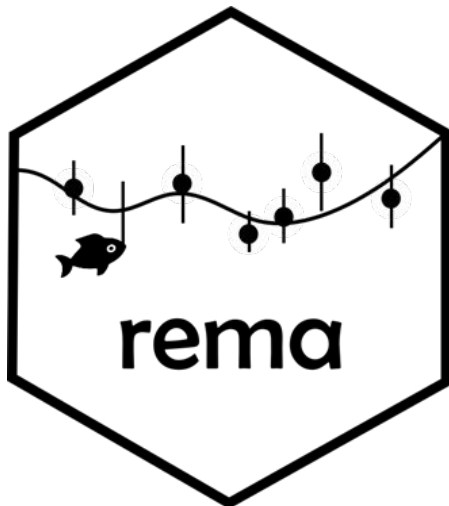




**NOAA  
FISHERIES**

# A consensus version of the random effects model for Tier 4/5 and apportionment



Jane Sullivan, Cole Monnahan,  
Pete Hulson, Jim Ianelli, James Thorson,  
and Andrea Havron

September 2022 Joint Groundfish Plan Team Meeting

# Outline for today's presentation

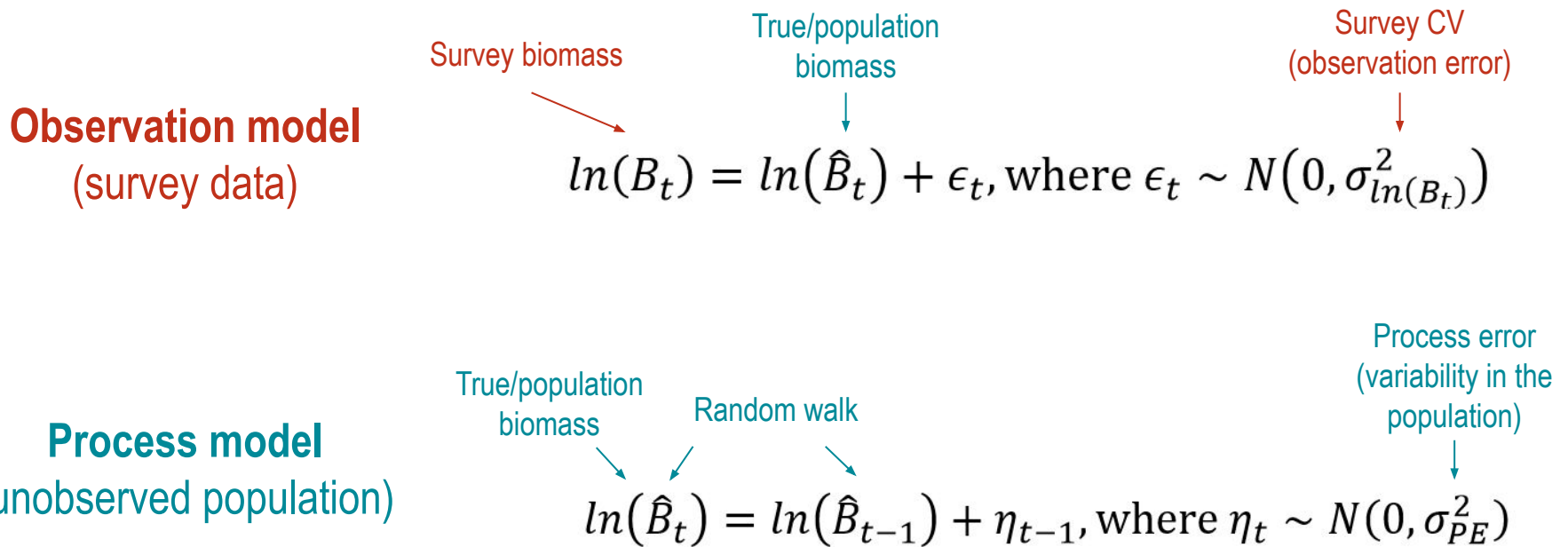
1. Background, motivation, and goals
2. Introduce *rema* R package
3. Model bridging
4. Correction to the two-survey ADMB model
5. Estimation of additional observation error
6. Zero biomass observations, experimental Tweedie distribution to model index data
7. Future work
8. Solicit your feedback and endorsement

[Link to Plan Team report](#)

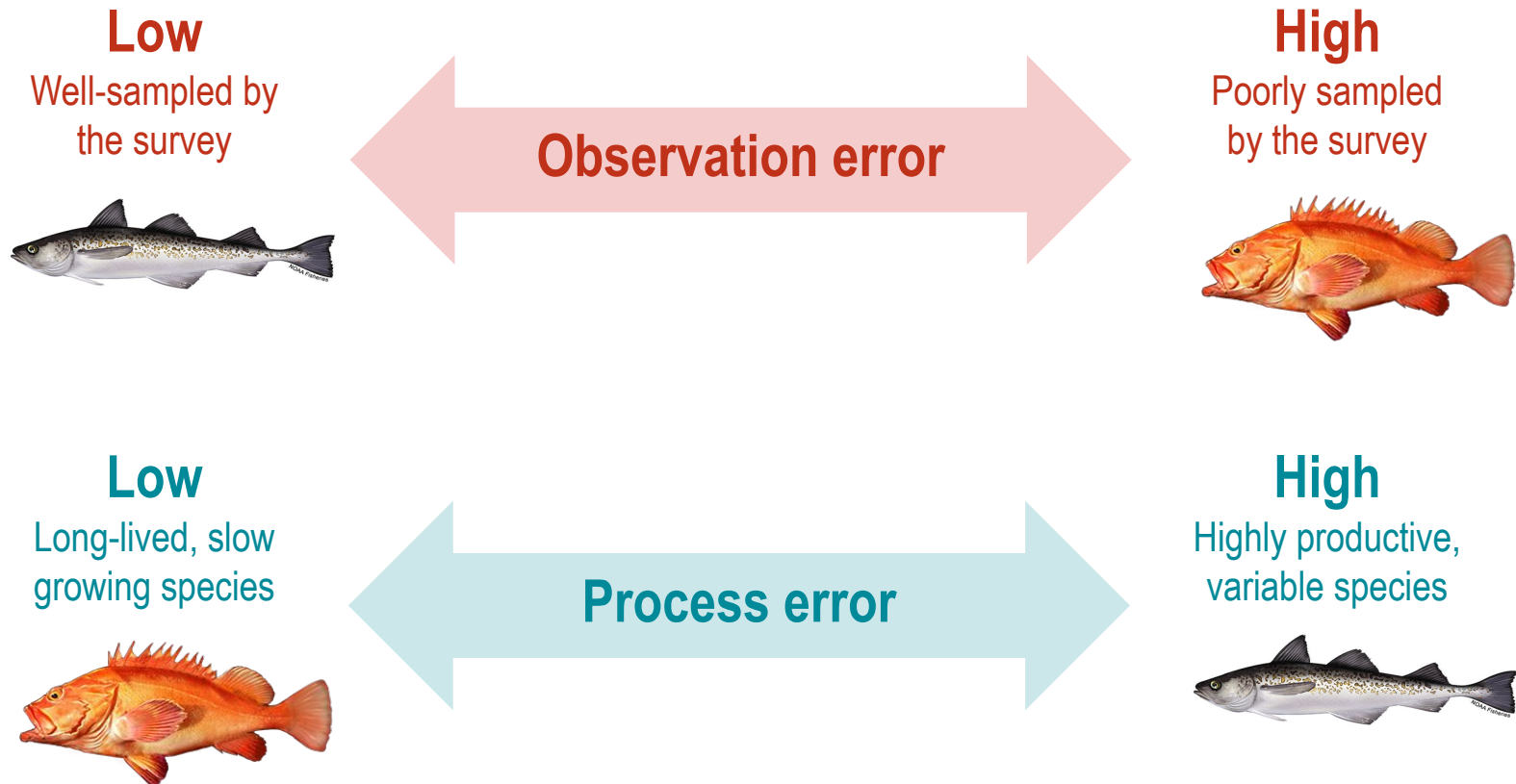
# Background

Survey averaging working group recommended the random effects (RE) model in 2013, endorsed by GPT and SSC ([Sept 2013 Joint GPT](#))

## State-space random walk model



# Now used for almost all Tier 4/5 stock assessments and most apportionment strategies



# What's the problem?

Model code and methods for implementing the RE model have diverged

Three primary variants ([Monnahan et al. 2021](#)):

1. **RE** model developed/tested (2012-2013)
  - Accepts a single biomass time series
2. **REM** extension (2015)
  - >1 index, user discretion pooled process errors
3. **REMA** development began in 2017
  - Additional survey index data (Hulson et al. 2021)

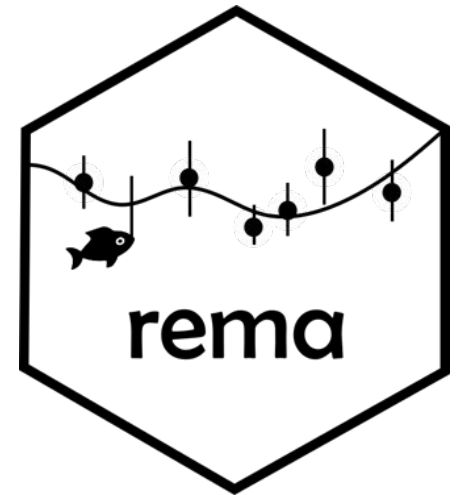
# Goals

Recommendation to “**create a consensus version of the RE model**” that is “**documented and version-controlled online**” ([Sept 2021 Joint GPT](#))

- Unify RE, REM, and REMA models
- Flexible enough to accommodate all Tier 4/5 models and apportionment strategies
- Improve transparency and reproducibility across assessments

# Intro to *rema* R package

- Coded in TMB ([Kristensen et al. 2016](#))
- Documentation and tutorials:
  - <https://afsc-assessments.github.io/rema/>
- Informative error messages
- Testing framework
- Extensible
- Modeled after *wham* R package ([Miller and Stock 2020](#))
- Applied to four stocks in 2022: GOA DSR, GOA thornyheads (Appendix A), BSAI shortraker (Appendix B), and BSAI other rockfish (Appendix C)



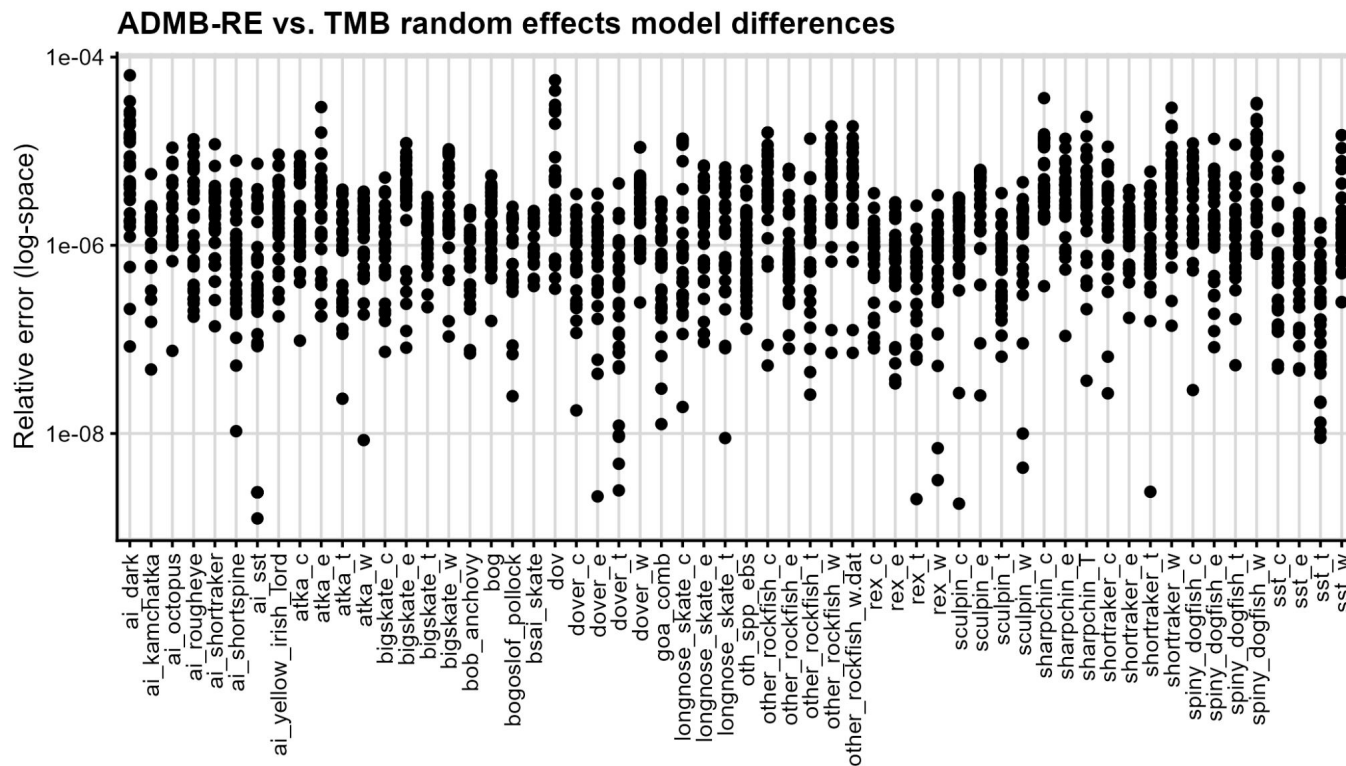
# The rema workflow

1. Load `rema` and data. The user can read biomass or other abundance index data from file (e.g. csv files), or they can use the `rwout.rep` report file from the ADMB version of the RE model using `read_admb_re()`.
2. Define the model structure and assumptions using `prepare_rema_input()`. This function allows users to quickly transition from a single to two survey model, specify alternative process error structures, add likelihood penalties or priors on parameters, and evaluate alternative assumptions about zero biomass observations.
3. Fit the specified REMA model using `fit_rema()` and determine whether the model has met basic convergence criteria.
4. Extract `rema` model output into clean, consistently formatted data frames using `tidy_rema()`. The user can visualize this model output using `plot_rema()`, or quickly format it into tables for a report.
5. Compare alternative REMA models and conduct model selection using `compare_rema_models()`. Output from this function includes a table of Akaike Information Criteria (AIC) when appropriate, figures, and tidied data frames. This function also accepts model output from the ADMB version of the RE model for easy comparison to past models.



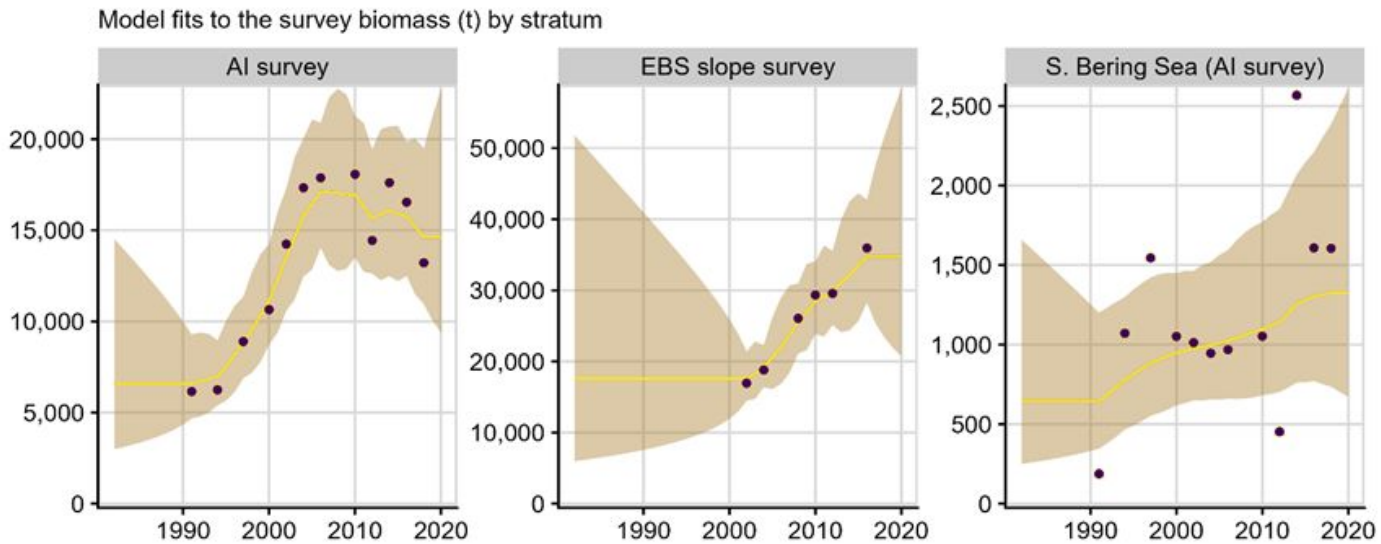
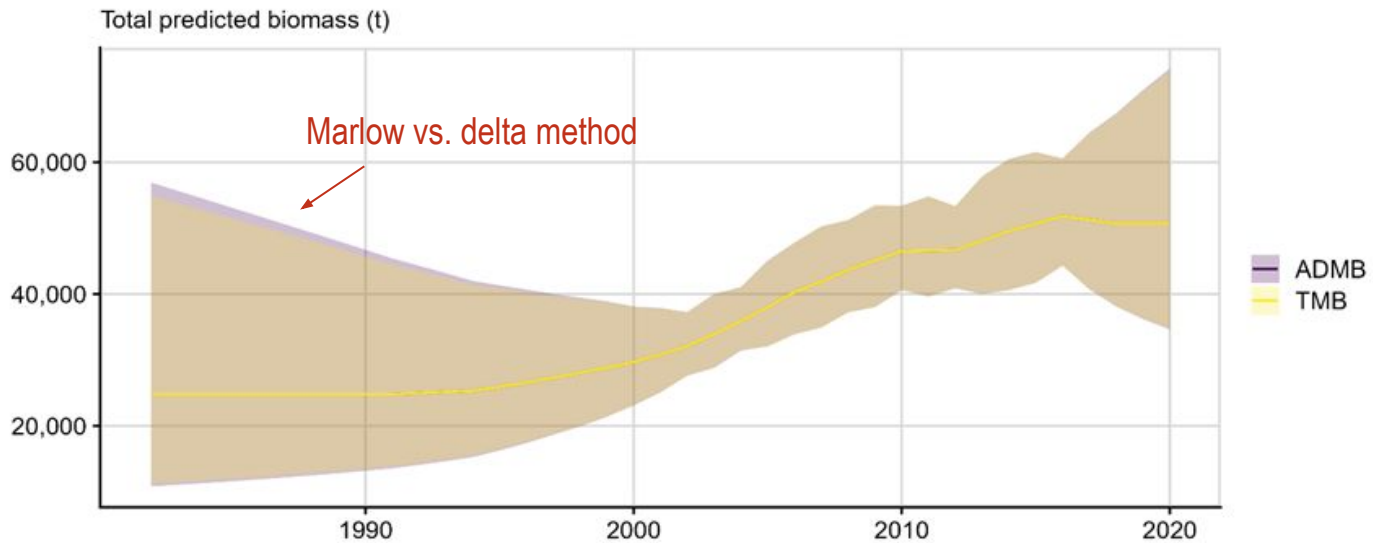
# Bridge from ADMB to TMB

58 case studies, <0.01% error. Small differences attributed to (1) different implementations of the Laplace approximation, and (2) different optimizers (p. 4, Fig. 1)






# Bridge from RE to REM

- Example: 3 survey indices
  - RE workflow: Fit 3 RE models, sum results
  - REM workflow: Fit 1 REM model
- Get the same answer
- ADMB version of REM uses Marlow method to get variance of total log biomass (Monnahan et al. 2021)
- In *rema* we recommend using delta method



REM model methods (p. 4) and variance method comparison (Fig. 2)

# Questions thus far?

1. **Background, motivation, and goals** 
2. **Introduce *rema* R package** 
3. **Model bridging** 
4. Correction to the two-survey ADMB model
5. Estimation of additional observation error
6. Zero biomass observations, experimental Tweedie distribution to model index data
7. Future work
8. Solicit your feedback and endorsement

# REMA extension: additional abundance index

Example: NMFS longline survey (LLS) relative population weights (RPWs)

- Species that are well-sampled by hook gear
- Missing surveys or survey reductions (e.g., EBS slope survey)

**Add additional observation equation:**

$$\overset{\text{LLS RPW}}{\ln(I_t)} = \ln(\hat{I}_t) + \epsilon_I, \text{ where } \epsilon_I \sim N(0, \overset{\text{LLS RPW CV}}{\sigma_{\ln(I_t)}^2})$$

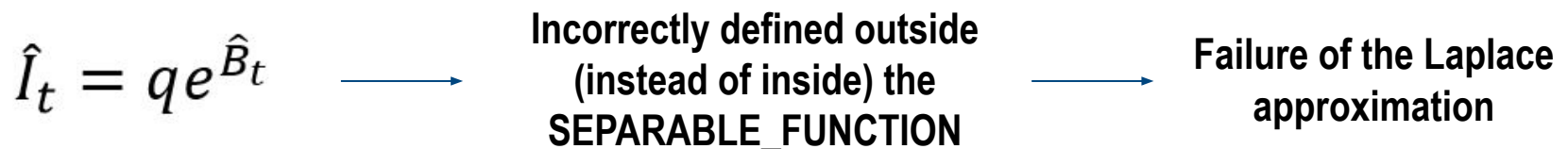
**New scaling parameter ( $q$ ) estimated:**

$$\hat{I}_t = qe^{\hat{B}_t}$$

p. 5-6

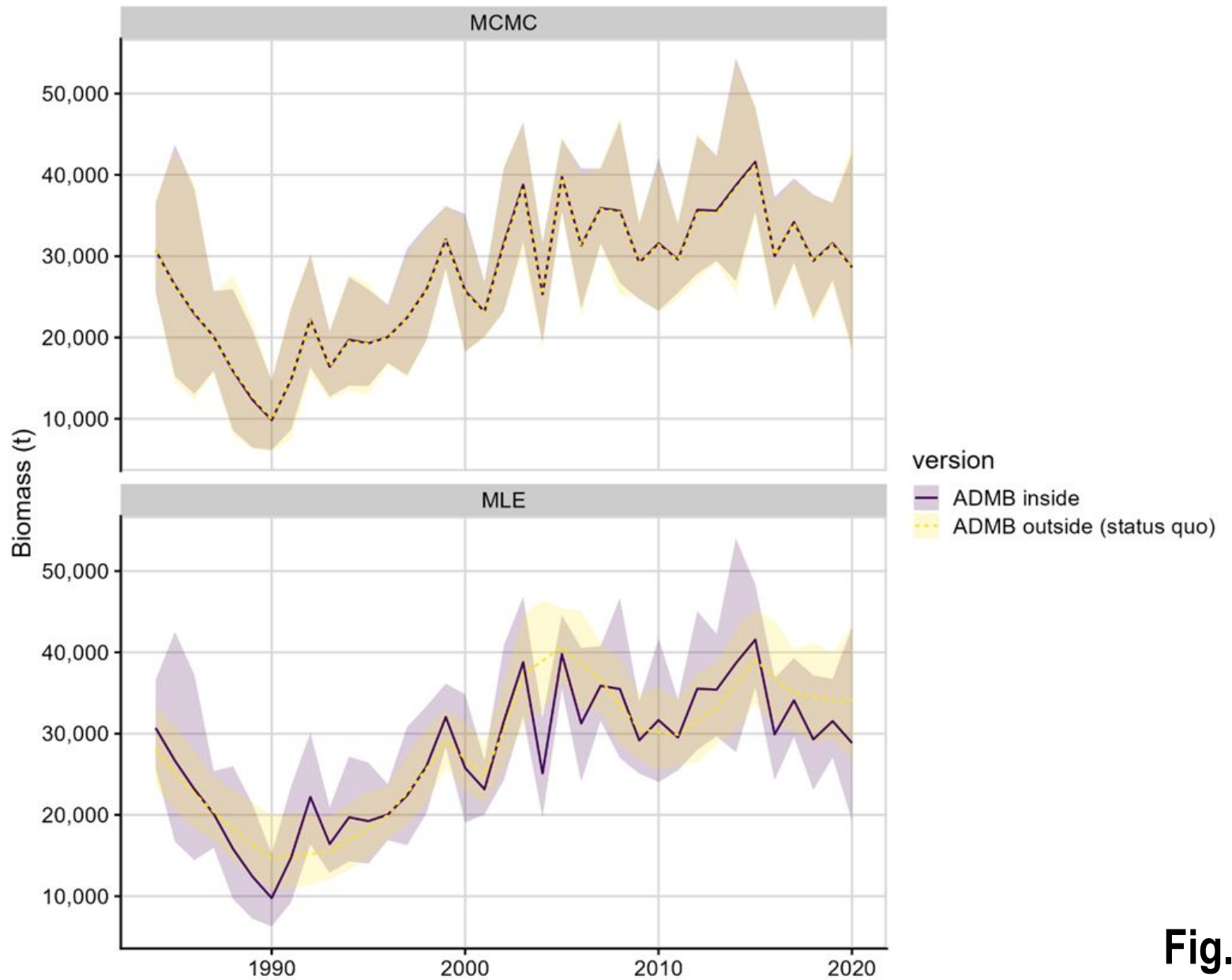
# Correction to the ADMB version of REMA

- Separability: break complex, multivariate integrals into a product of simpler, univariate integrals to increase computational efficiency of the Laplace approximation
- Parameters defined in ADMB's PARAMETER\_SECTION of the template file cannot be used within the SEPARABLE\_FUNCTION unless they are passed as arguments to the function ([Skaug and Fournier 2013](#))

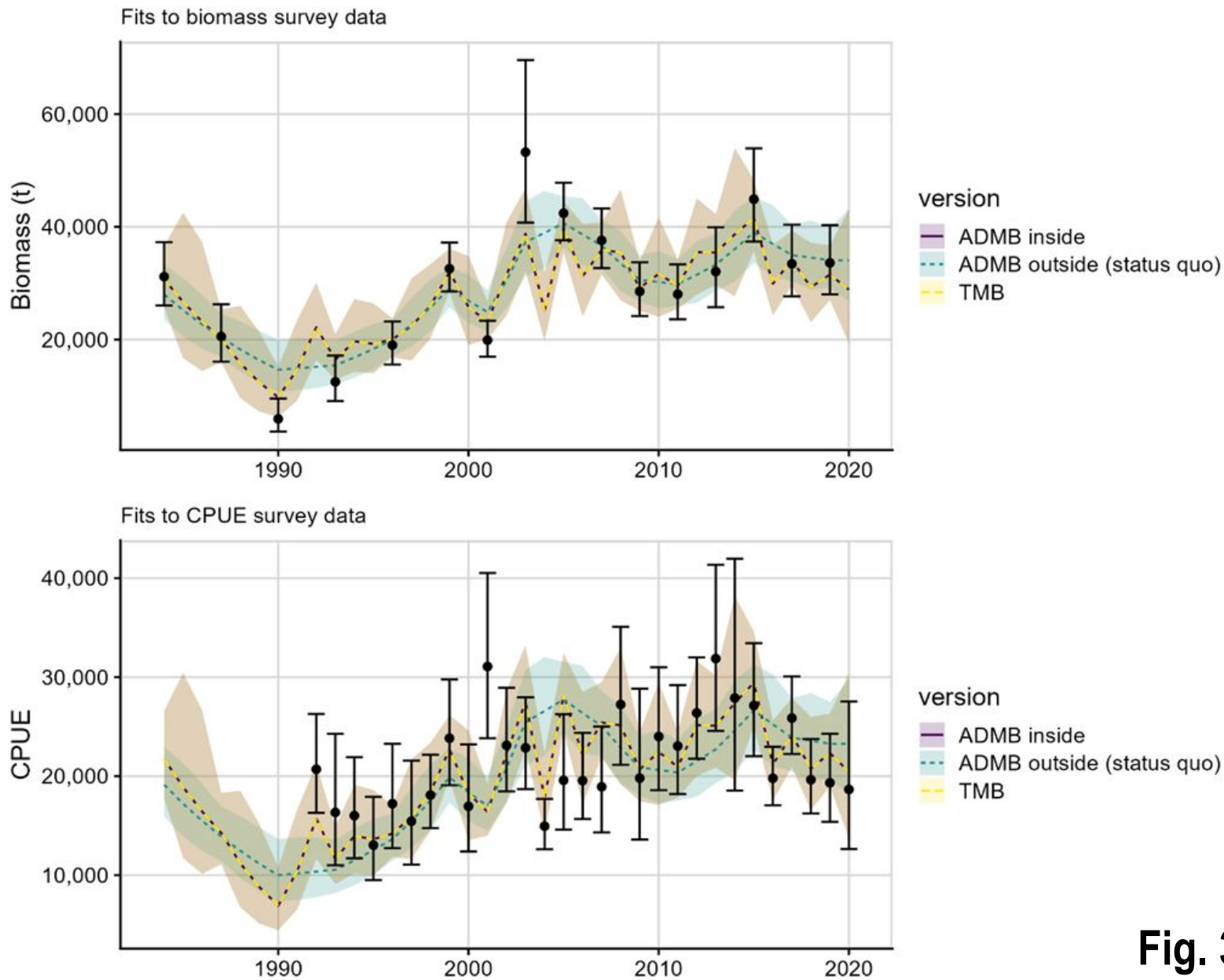


- Analyzed this problem using MCMC and by examining the marginal negative log-likelihood ([reproducible example](#))

p. 5-6, Fig. 3 and 4



**Fig. 4**



**Fig. 3**



# Three stock assessments impacted

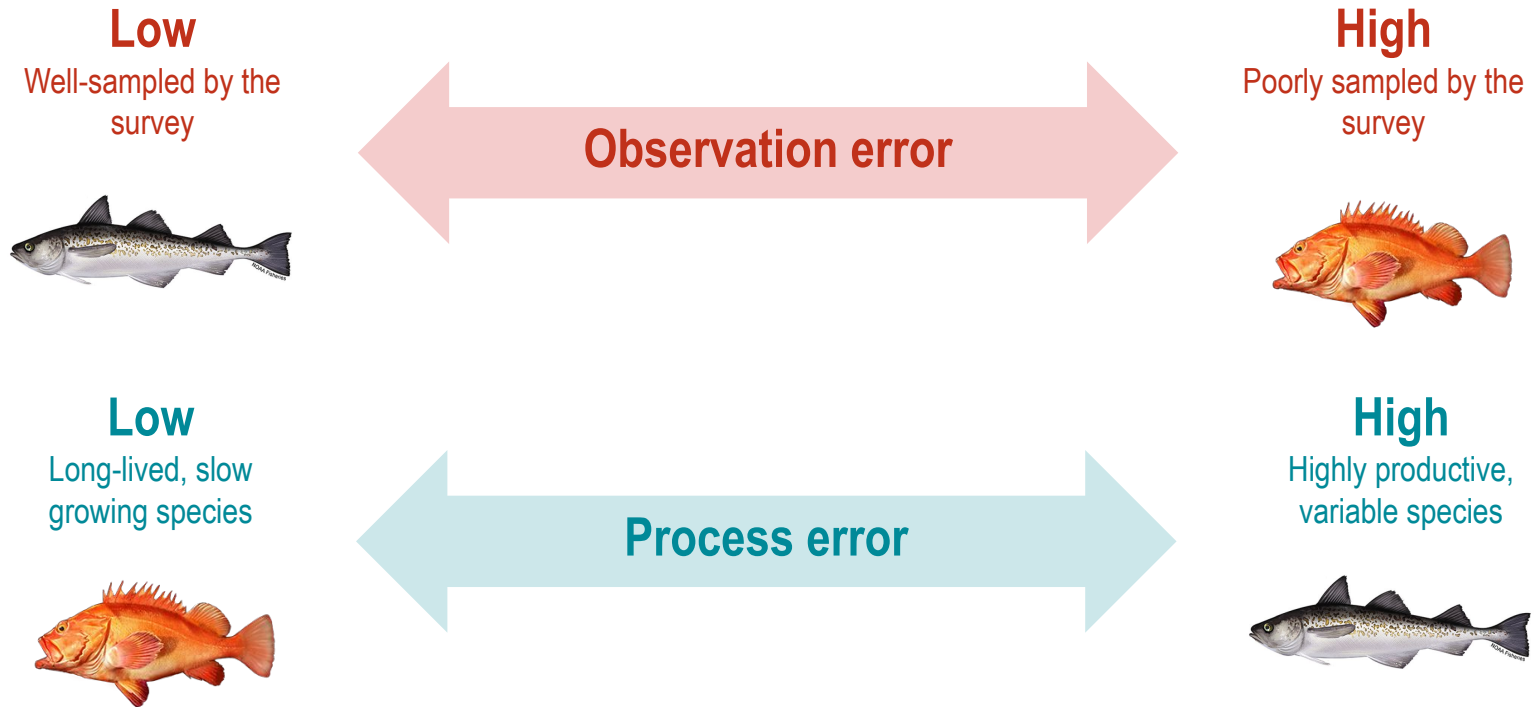
1. GOA thornyheads (corrected in Appendix A)
2. GOA shortraker (next full assessment in 2023)
3. GOA rougheyeye apportionment (next full assessment in 2023)

Corrected estimates of process error variance were often much higher than previously thought, leading to noisier biomass trajectories.



# Trade-off between observation and process error

If survey data are noisy but estimates of observation error are low, process error estimates may be too high for some species' life histories



# One potential solution: Estimate additional observation error

Common in crab stocks and has been explored in some groundfish stocks

Observation error

Survey estimates of biomass CV

New extra CV (estimated)

$$\sigma_{\ln(B_t)} = \sqrt{\ln\left(\frac{\sigma_{B_t}^2}{B_t} + \sigma_\tau^2 + 1\right)}$$

# Estimate additional observation error

```
> compare <- compare_rema_models(rema_models = list(m0, m1, m2, m3))
> compare$aic
```

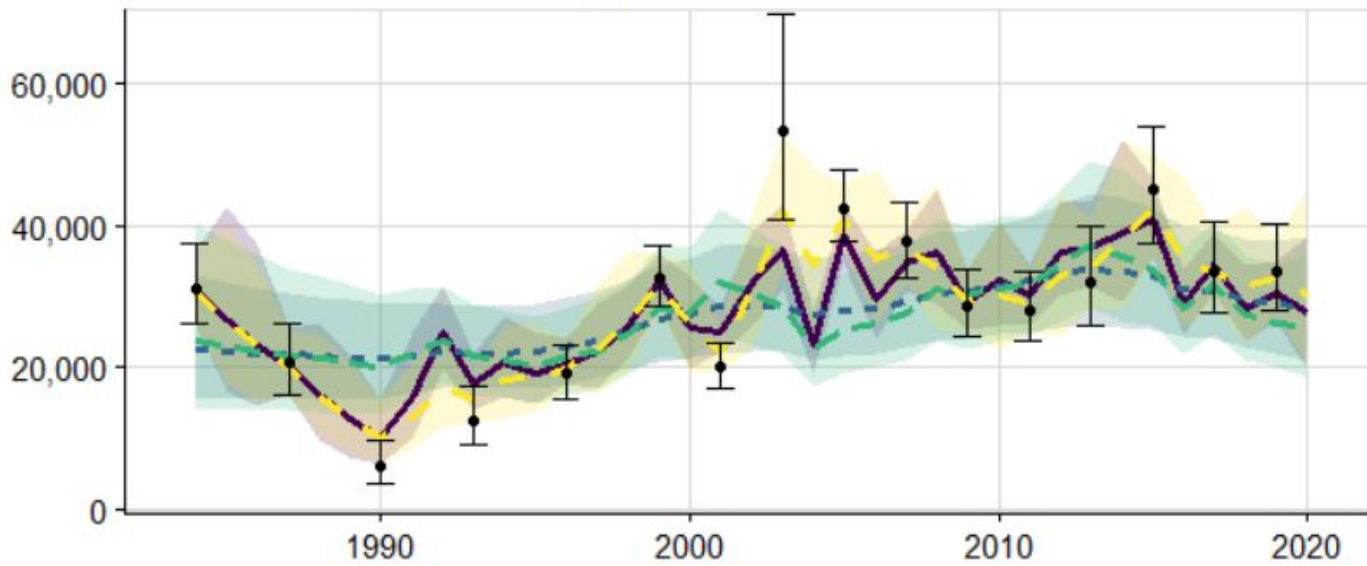
	model_name	objective_function	n_parameters	aic	delta_aic
1	Xtra Biom + CPUE Obs Err	6.6	4	21.2	0.0
2	Xtra Biom Obs Err	8.6	3	23.2	2.0
3	Xtra CPUE Obs Err	20.5	3	47.0	25.8
4	Base	31.1	2	66.2	45.0

```
> compare$output$parameter_estimates
```

	model_name	parameter	estimate	std_err	lci	uci
1	Base	process_error	0.28	0.05	0.19	0.40
2	Base	scaling_parameter_q	0.71	0.03	0.66	0.77
3	Xtra Biom Obs Err	process_error	0.14	0.04	0.08	0.24
4	Xtra Biom Obs Err	scaling_parameter_q	0.76	0.09	0.61	0.95
5	Xtra Biom Obs Err	extra_biomass_cv	0.42	0.11	0.25	0.65
6	Xtra CPUE Obs Err	process_error	0.24	0.05	0.15	0.37
7	Xtra CPUE Obs Err	scaling_parameter_q	0.71	0.05	0.63	0.81
8	Xtra CPUE Obs Err	extra_cpue_cv	0.26	0.06	0.17	0.40
9	Xtra Biom + CPUE Obs Err	process_error	0.07	0.03	0.03	0.18
10	Xtra Biom + CPUE Obs Err	scaling_parameter_q	0.75	0.08	0.60	0.92
11	Xtra Biom + CPUE Obs Err	extra_biomass_cv	0.40	0.10	0.24	0.62
12	Xtra Biom + CPUE Obs Err	extra_cpue_cv	0.13	0.04	0.07	0.24

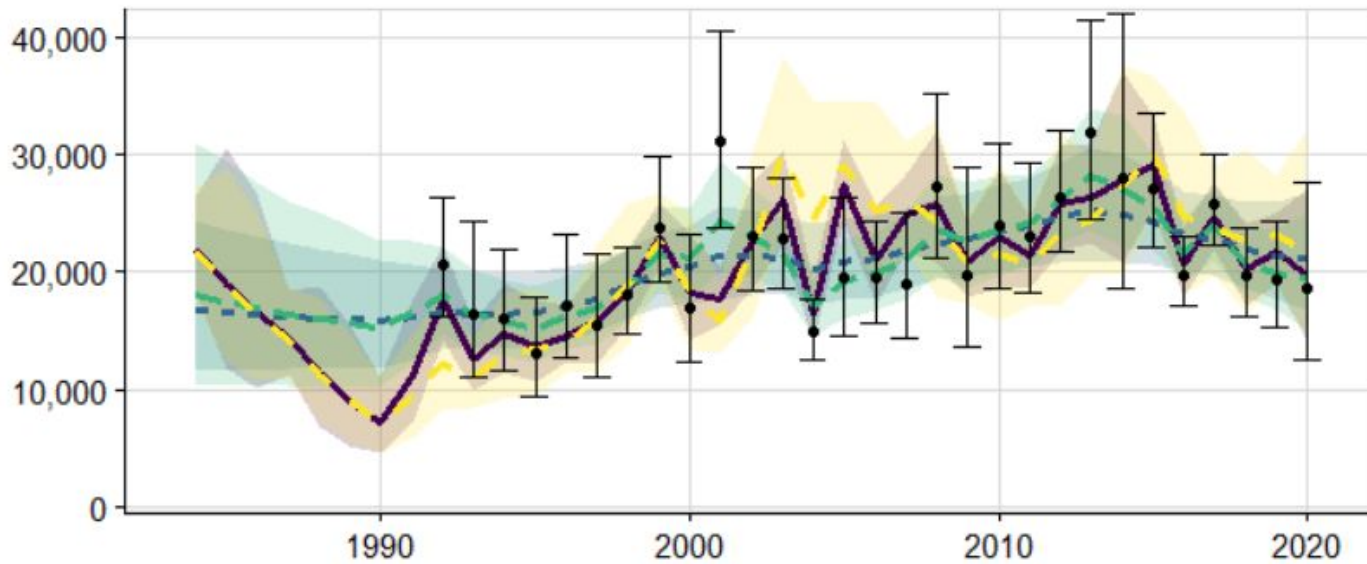


# Trawl survey biomass (t)








- Base
- Xtra Biom + CPUE Obs Err
- Xtra Biom Obs Err
- Xtra CPUE Obs Err

# Longline survey RPW

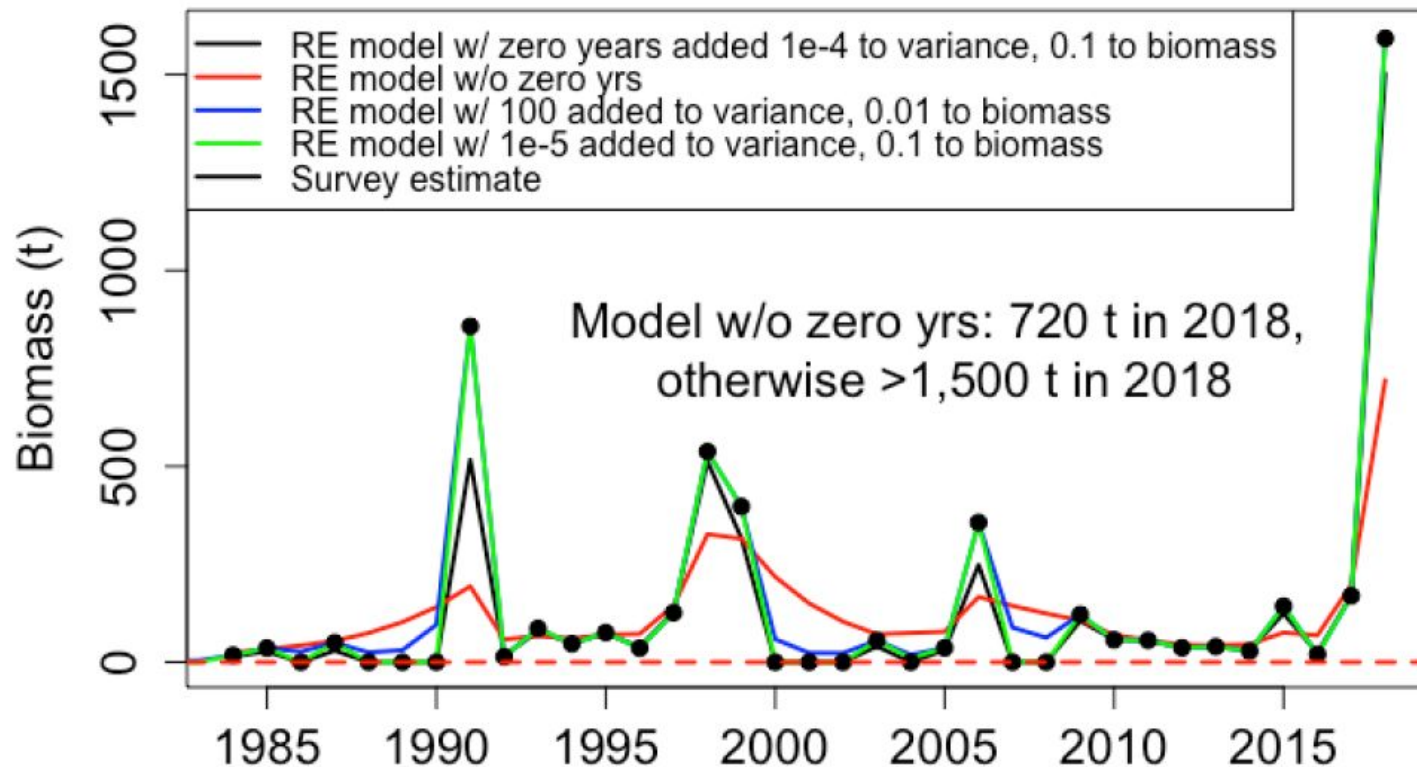


# Checking in

1. **Background, motivation, and goals** 
2. **Introduce *rema* R package** 
3. **Model bridging** 
4. **Correction to the two-survey ADMB model** 
5. **Estimation of additional observation error** 
6. Zero biomass observations, experimental Tweedie distribution to model index data
7. Future work
8. Solicit your feedback and endorsement

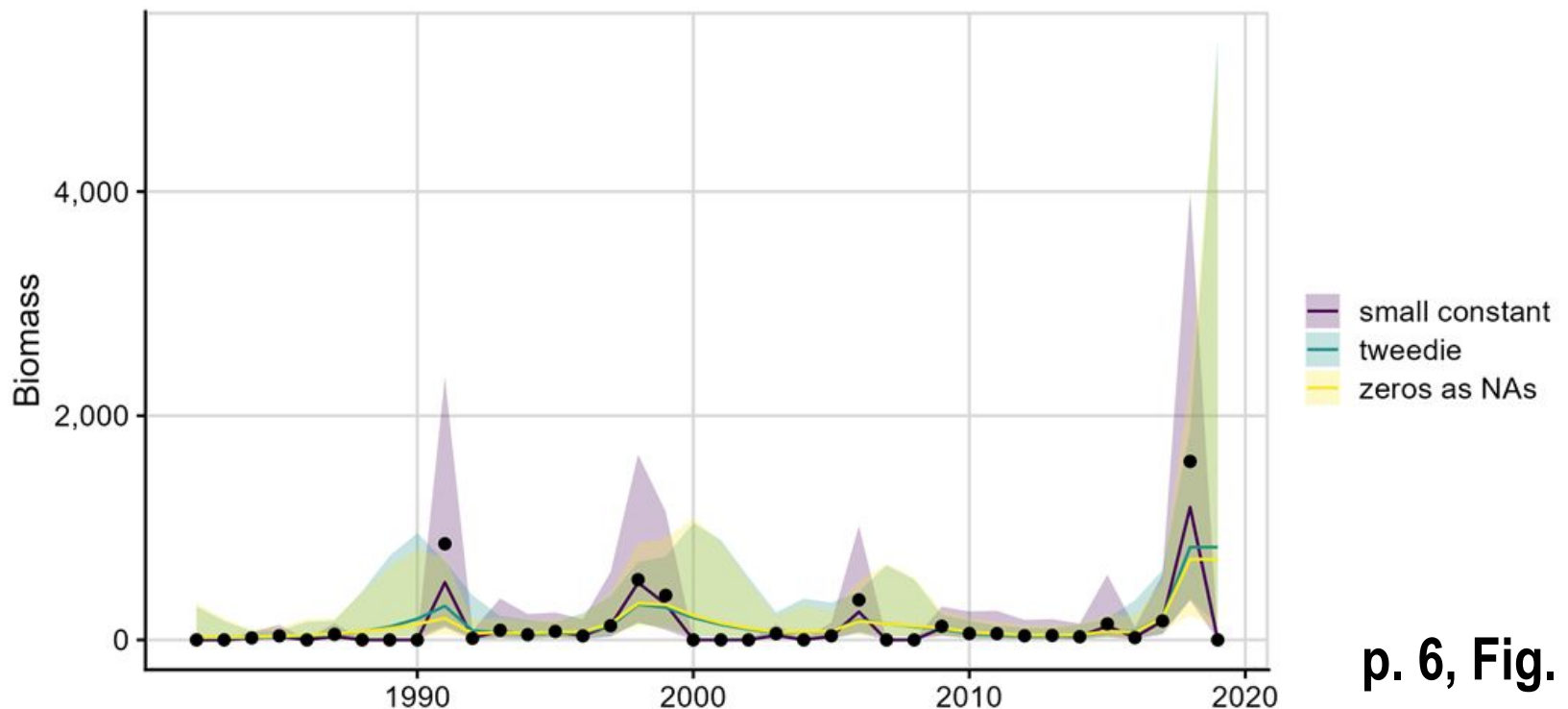
# Importance of dealing with zeros

Spies et al. (2018) found ~**double the terminal biomass** depending on if the zeros were ignored or a small constant was added



# Experimental Tweedie for observation model

- Three parameters: mean ( $\mu$ ), power parameter ( $\rho$ ), and dispersion ( $\phi$ )
- When  $\rho > 1$  and  $\rho < 2$ , Tweedie is positive, continuous, and can equal zero
- Tweedie model can be slow to run and often did not converge



p. 6, Fig. 5



# Recommendations on zeros

ADMB models are not consistent in their approach to zeros

- Check your data and explicitly state assumptions
- By default *rema* treats zeros as NAs and will return warning message if user inputs a zero without defining an assumption

# Next steps

- Model validation: Pearson residuals are not independent  $N(0,1)$ . Use one-step ahead (OSA) residuals instead:

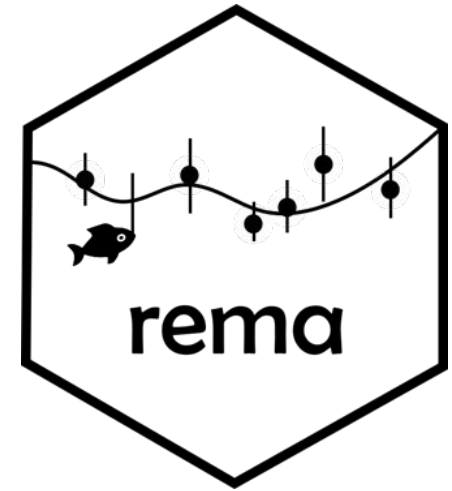
$$(y_i - \hat{y}_{i|i-1}) / \hat{\sigma}_{i|i-1}$$

- Continued experimentation with Tweedie
- Priors on process error variation based on life history ([Spencer et al. 2015](#))
- Off-season: code review, feedback from authors, implementation if endorsed

# Questions?

Tutorials and documentation:

<https://afsc-assessments.github.io/rema/>



Thanks again to my coauthors and to Katy Echave, Phil Joy, and Kalei Shotwell for letting me implement *rema* for their stocks.