One step ahead (OSA) residuals

Cole Monnahan

2023 September Plan Team

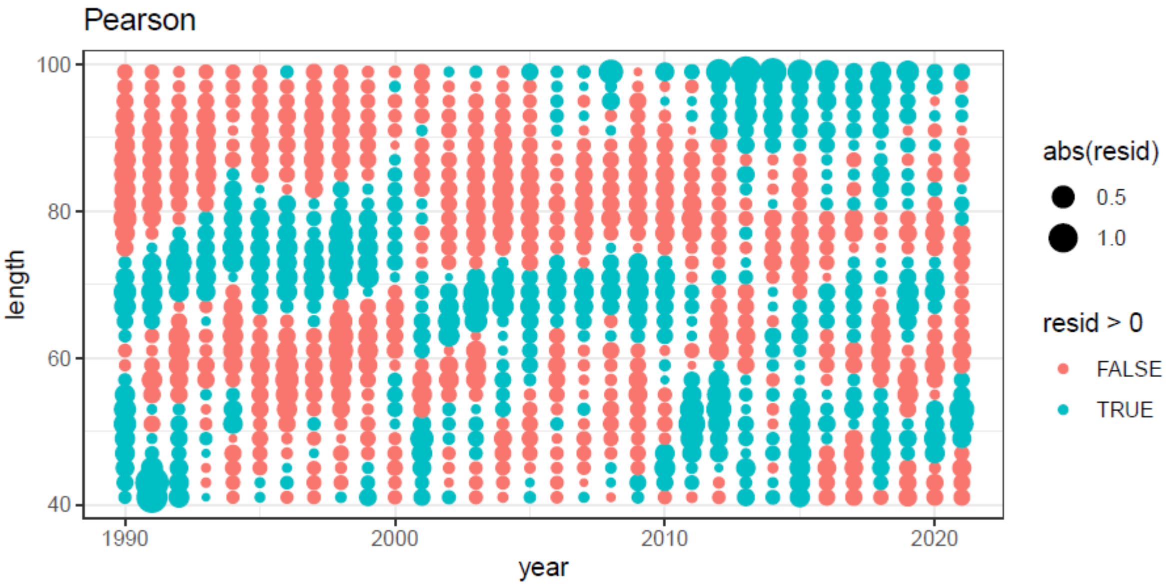
cole.monnahan@noaa.gov

Road map for today

- Review how Pearson are used
- When and why they are wrong
- Build intuition for OSA
    - Solves non-normality
    - Solves correlation
    - Properties: iid Z
- Work through set of examples
- Propose standard protocol for how to report/interpret them

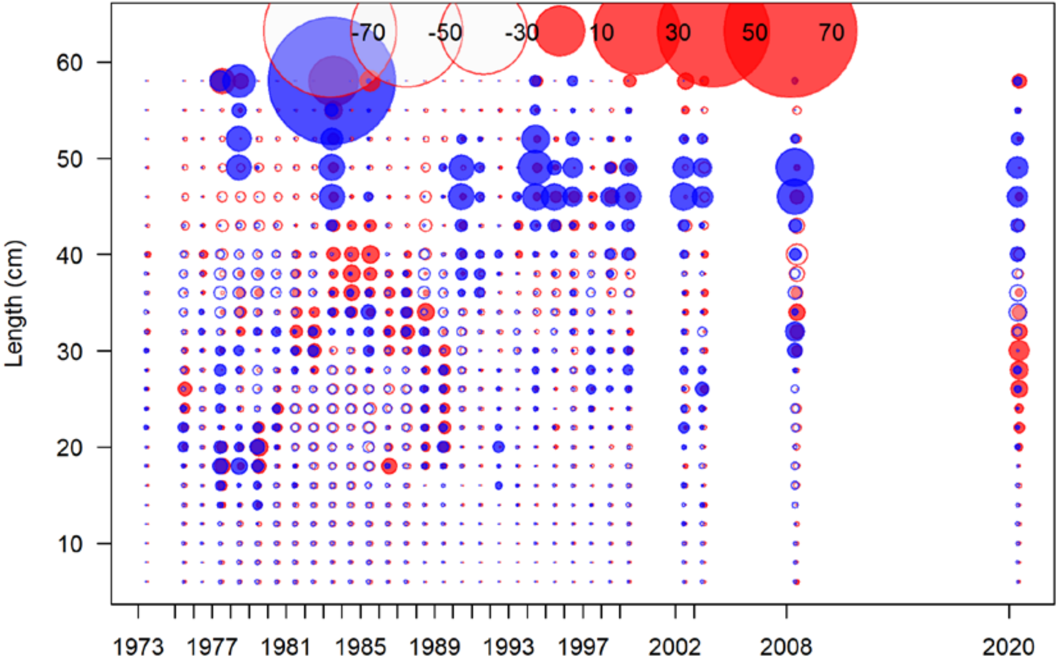
Current approach for Pearson residuals

- Pearson residuals plotted age vs year
- Implicitly assume standard normal
- Patterns and large residuals (|r|>3) suggest model misfit and potential solutions



Current approach for Pearson residuals

- Pearson residuals plotted age vs year
- Implicitly assume standard normal
- Patterns and large residuals (|r|>3) suggest model misfit and potential solutions



Pearson residual calculations

- Pearson residuals are common in GLM regressions
- "*Pearson residuals can be skewed for nonnormal responses*" (Faraway 2006)
- Age bins calculated independently
- Output implicitly assumed to be standard normal (how many SDs from the mean)
- If there is a large positive Pearson residual, the others must be negative!

$$r_P = \frac{y - \hat{\mu}}{sd(\hat{\mu})} = \frac{observed - expected}{sd(expected)}$$

<u>For the multinomial:</u>

$$r_P(a) = \frac{p(a) - \hat{p}(a)}{\sqrt{n\hat{p}(a)(1 - \hat{p}(a))}}$$

$p = \text{observed proportion}$

$\hat{p} = \text{expected proportion}$

$n = \text{sample size}$

$a = \text{age bin}$

Pearson residuals are wrong, OSA are right

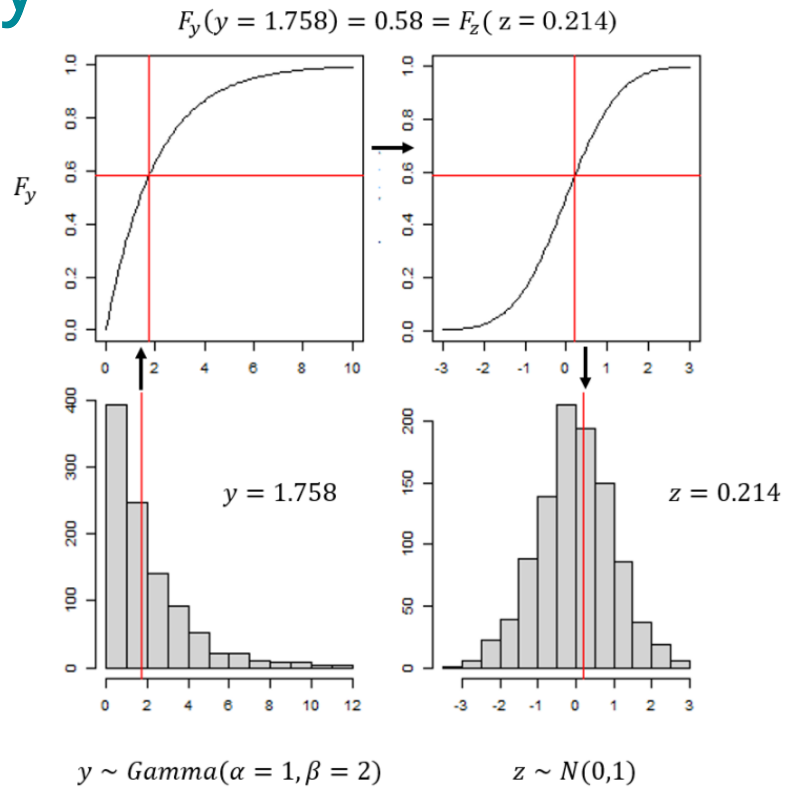
- Pearson resids don’t represent what we assume for two reasons
    - They ignore correlation between age/length bins
    - They are not iid Z when using skewed distributions
- One step ahead (OSA) residuals solve both issues
    - Expanded from timeseries to all TMB models by Thygessen et al. 2017
    - Expanded to composition data for assessments by Trijoulet et al. 2023
- Note: Pearson and OSA are identical when the data are independent and normal

Model validation basics

- Assume some statistical model fitted to some data *y*
- We want to validate the model by asking the question: *Could the observed data have come from the model?*
- The model has a multivariate distribution of predicted data
    - We compare where the observed data falls in that distribution
- This multivariate data distribution will not be normal and may have correlations for several reasons

Solving issue 1: non-normality

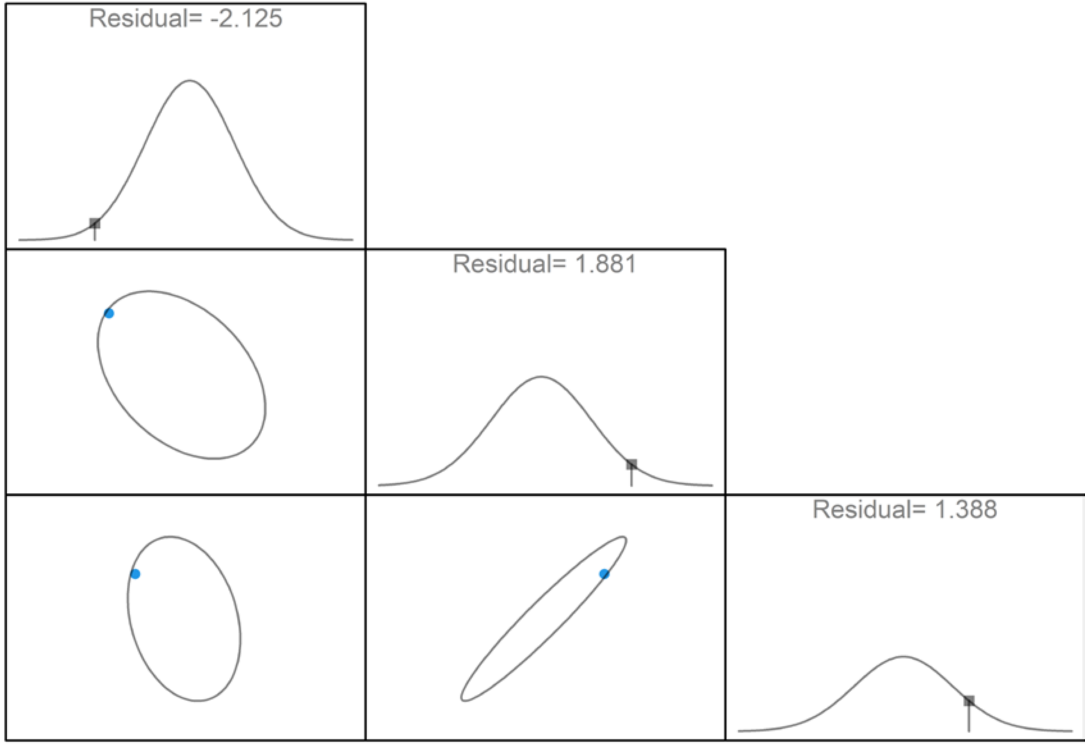
- In skewed distributions (e.g., multinomial) the residual is **not** the number of SDs from the mean
- The solution is to use quantile residuals
    - Convert quantile to u=CDF(x) in (0,1)
    - Then r=CDF^{-1}(u) of standard normal
    - u~U(0,1) so r~N(0,1)
- Can be done with simulation (DHARMa) or analytically if known CDF



Havron et al. in prep., http://florianhartig.github.io/DHARMa/

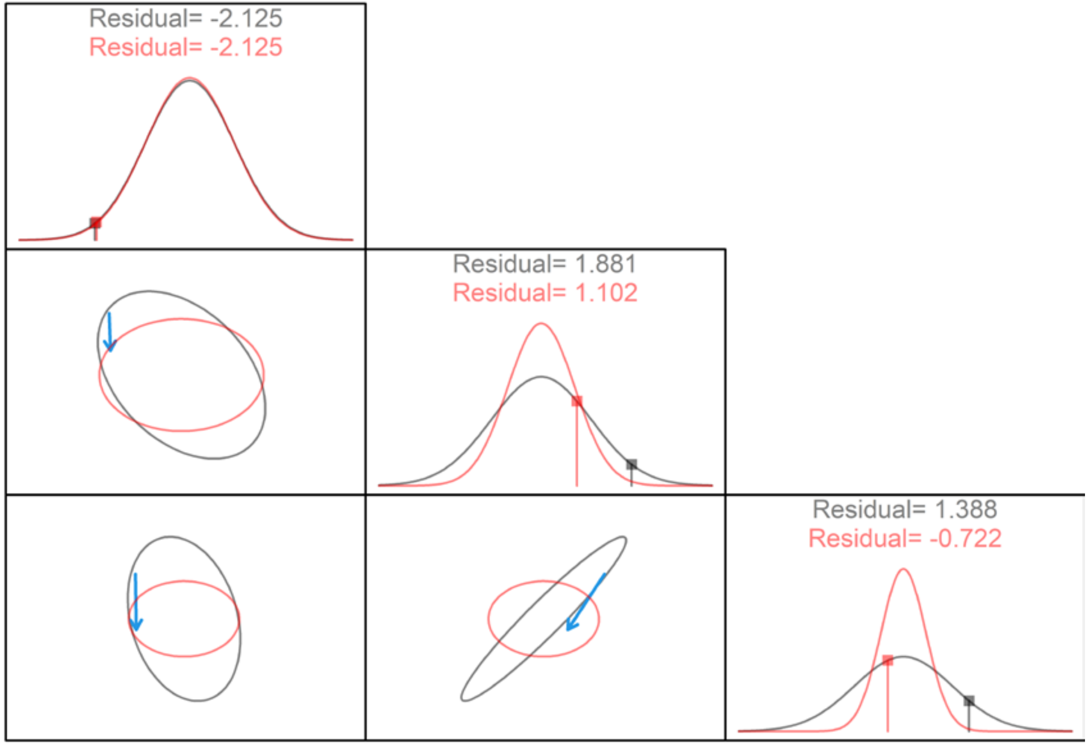
Solving issue 2: correlation

- Assume we have 3d multivariate normal with a single vector of observations (blue point)
- First, ignore correlations and calculate marginal residuals



Solving issue 2: correlation

- Assume we have 3d multivariate normal with a single vector of observations (blue points)
- Now, decorrelate (red lines) via a Cholesky decomposition rotation
- Blue arrow shows how observations moves
- **3^{rd} residual goes from 1.4 to -0.7**
- The residuals are wrong w/o rotation. (Even the sign!)



Solving issue 2: correlation

- Rotation **changes order and thus quantiles** so residuals are invalid when correlations exist
- What causes correlations?
  - Random effects. This is why OSA is used for mixed effects models more broadly.
  - Multivariate distributions which have inherent constraints such as sum to N (multinomial) or sum to 1 (Dirichlet). E.g., the covariance for a multinomial is $-np_ip_j$ i.e., if one bin goes up, others must go down
- OSA solves nonnormality and correlations simultaneously.
  - I think of OSA **as an analytical method for quantiles for non-normal correlated distributions.**
  - The math is intense and uses TMB's Laplace approximation machinery. There are several "methods" (ignored here)

$$U_i = \frac{\mathbf{P}^M(Y_i \le y_i, Y_1^{i-1} = y_1^{i-1})}{\mathbf{P}^M(Y_1^{i-1} = y_1^{i-1})} = \frac{\int_{(-\infty,y_i]} f_i(y)\,d\mu(y)}{\int_{(-\infty,y_i]} f_i(y)\,d\mu(y) + \int_{(y_i,+\infty)} f_i(y)\,d\mu(y)}$$

Thygesen et al. 2017

OSA overview

OSA residuals are statistically better

- They are available now, calculated externally via R package (https://github.com/fishfollower/compResidual)
- They are distributed iid N(0,1) <u>under a correct model</u>
- Large residuals (>3 ?) or non-normality suggest model misfit

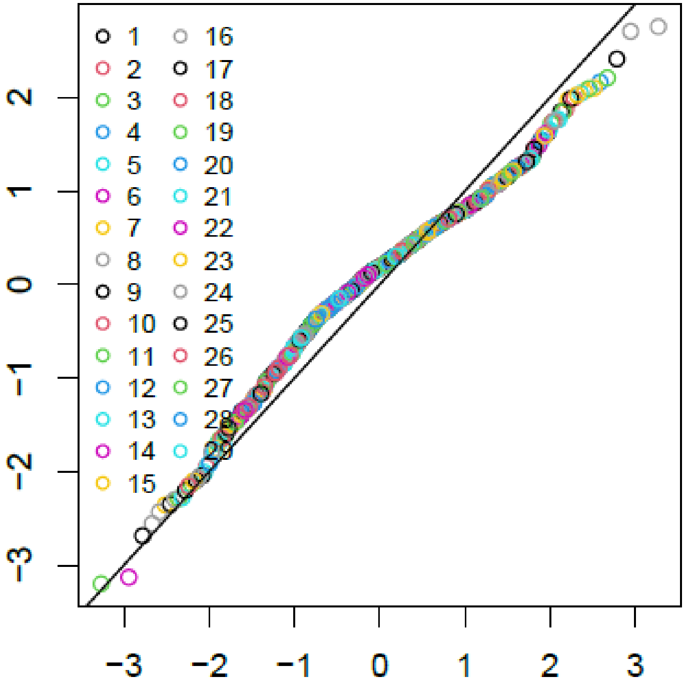
OSA residuals for composition data introduce a new issue

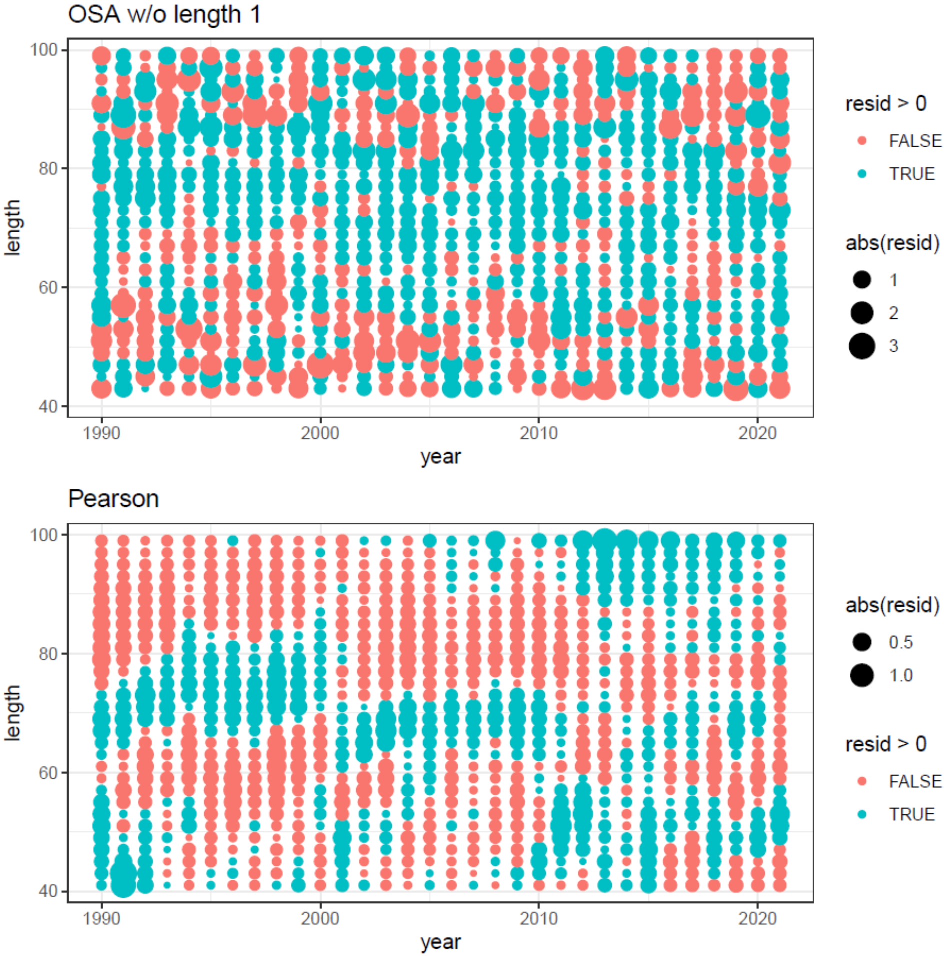
- Due to the constraint (sum to N, etc.) it only makes sense to produce K-1 residuals, in other words **we must drop an age/length bin**
- The order of bins also matters, although we likely won't alter that
- Consequently there is no single residual for a certain data point (proportion in a bin).
- Instead the residuals are realizations with the 'correct properties'

OSA overview

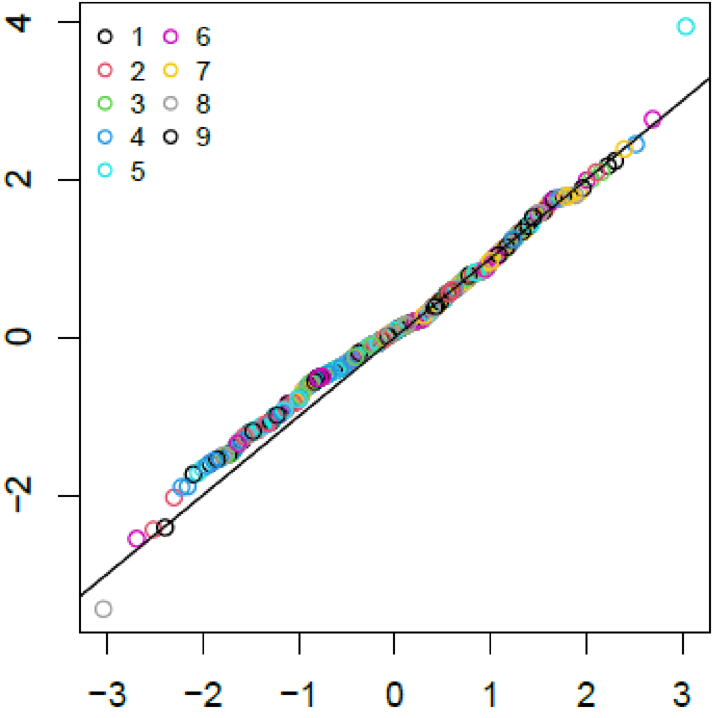
Questions before we go into examples?

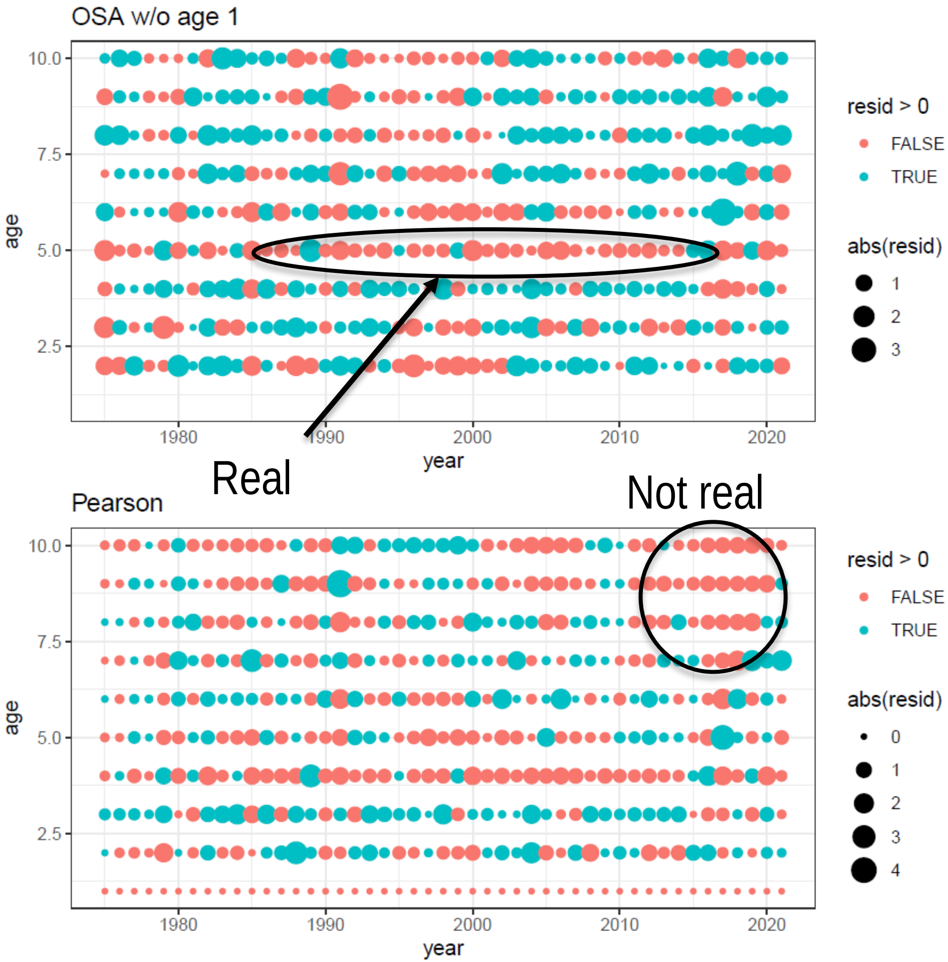
Sablefish fishery LL fem





GOA pollock fishery ages



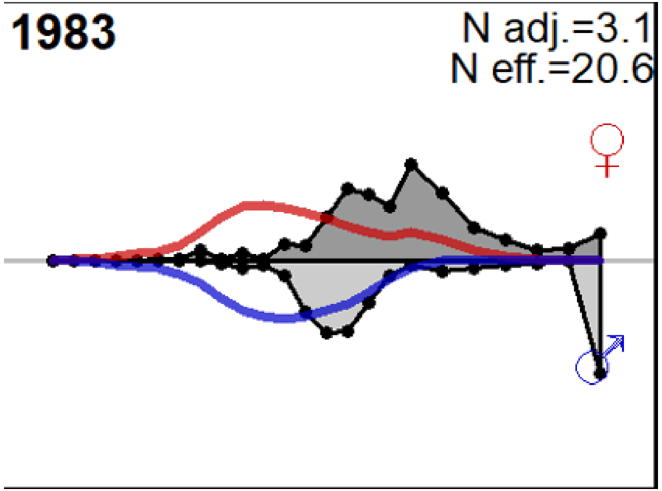


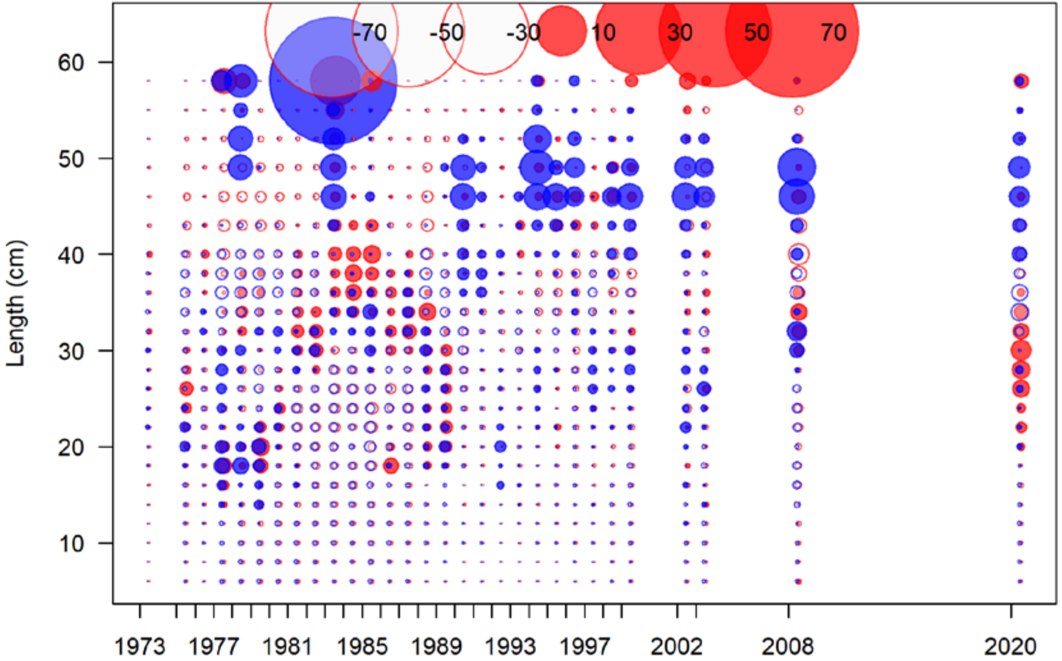
Real

Not real

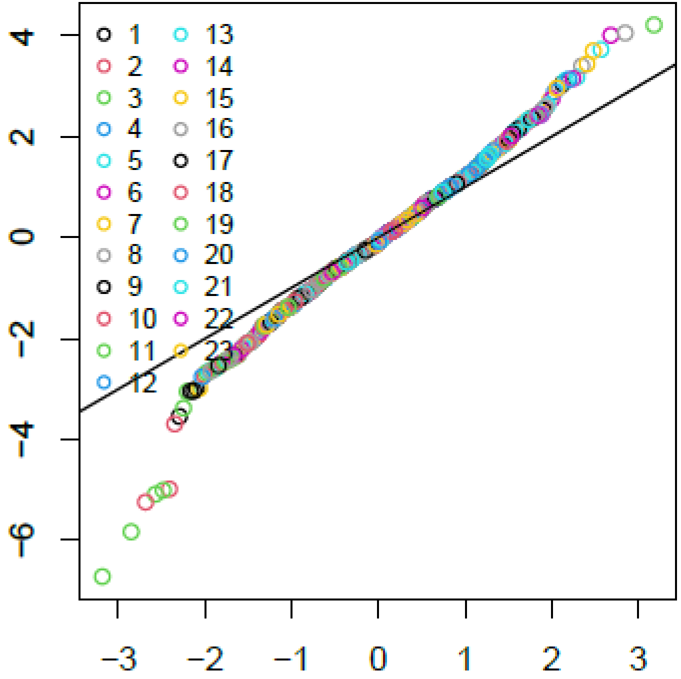
BSAI Flathead fish lengths

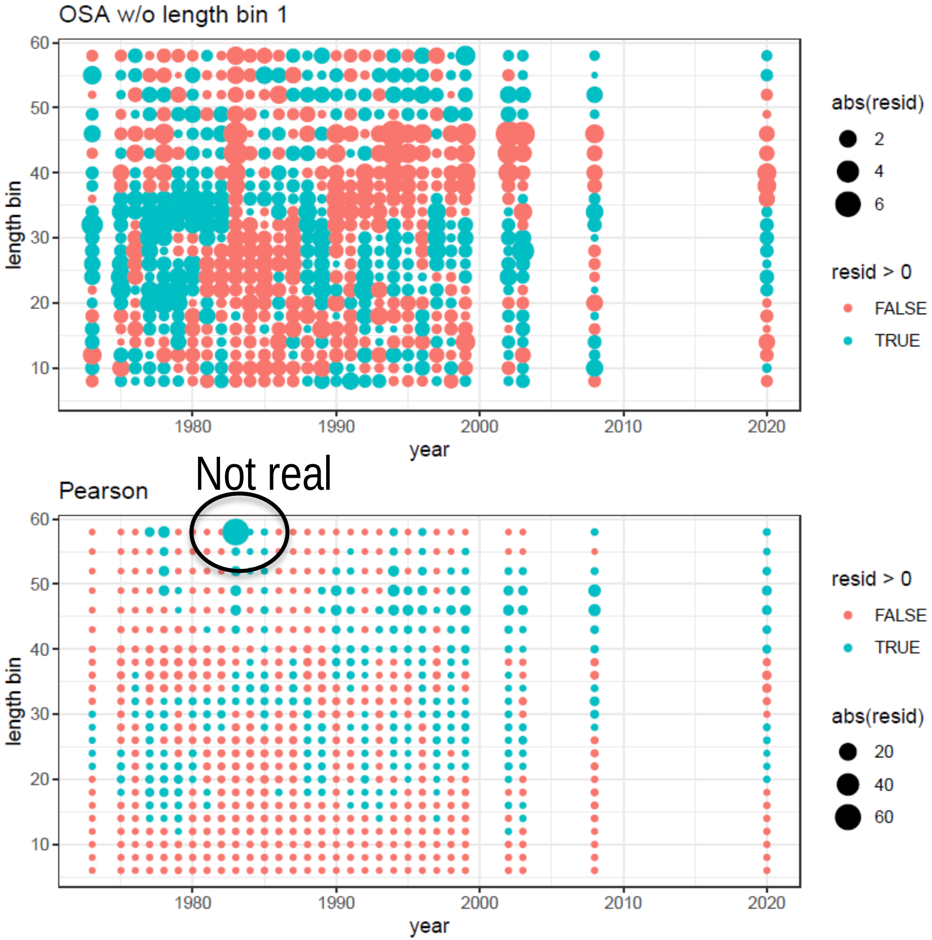
Is the Pearson residual of 60 real?



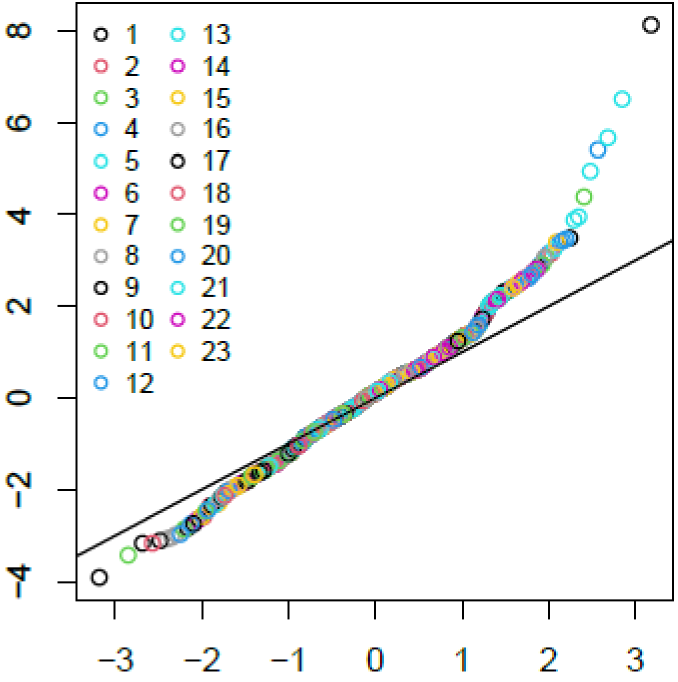


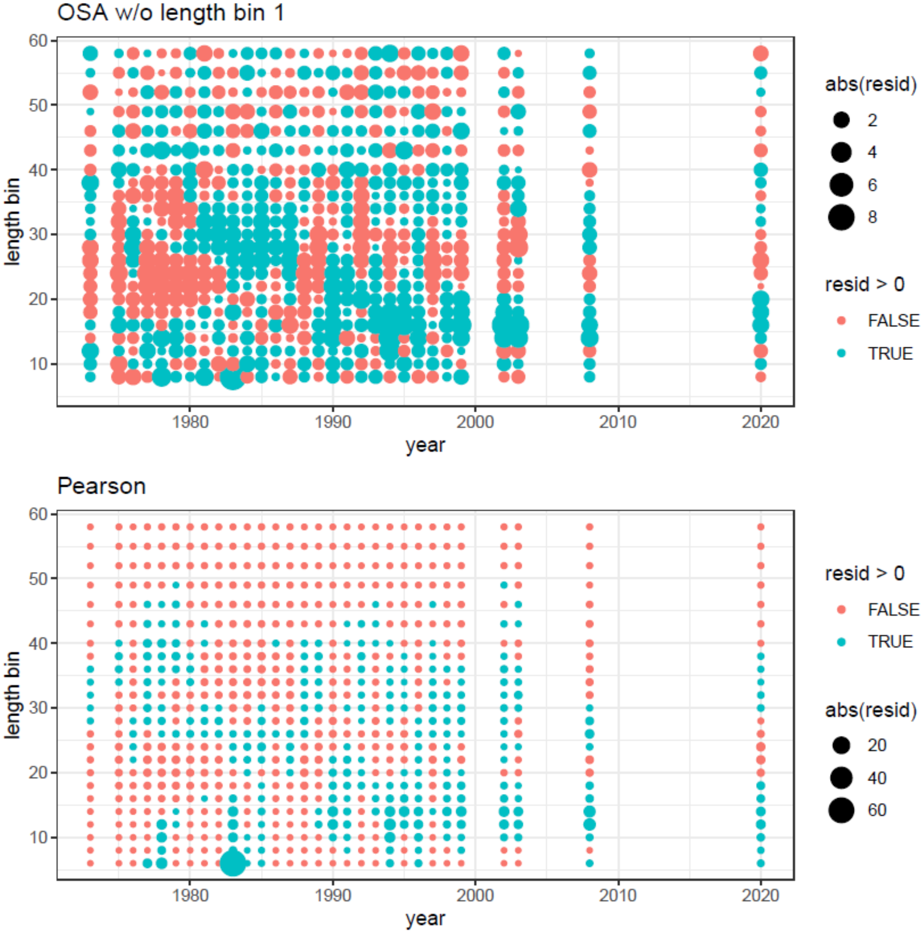
BSAI Flathead fish lengths



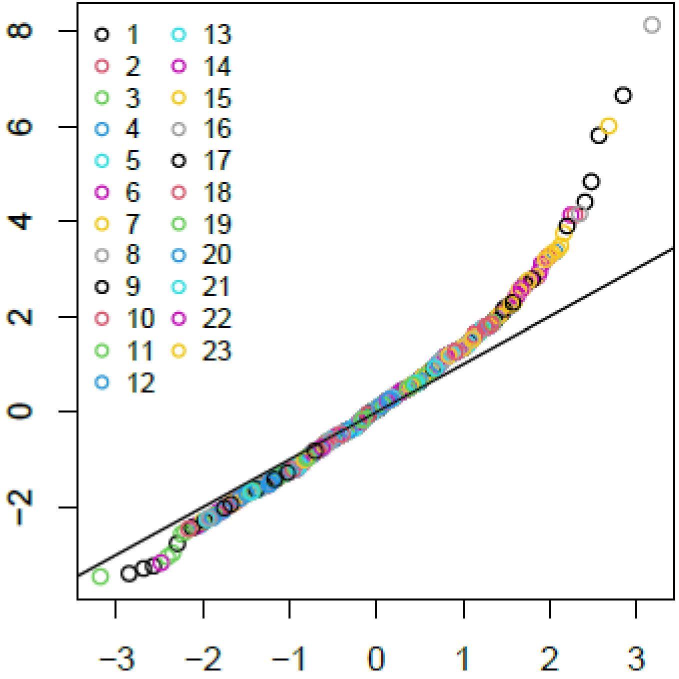


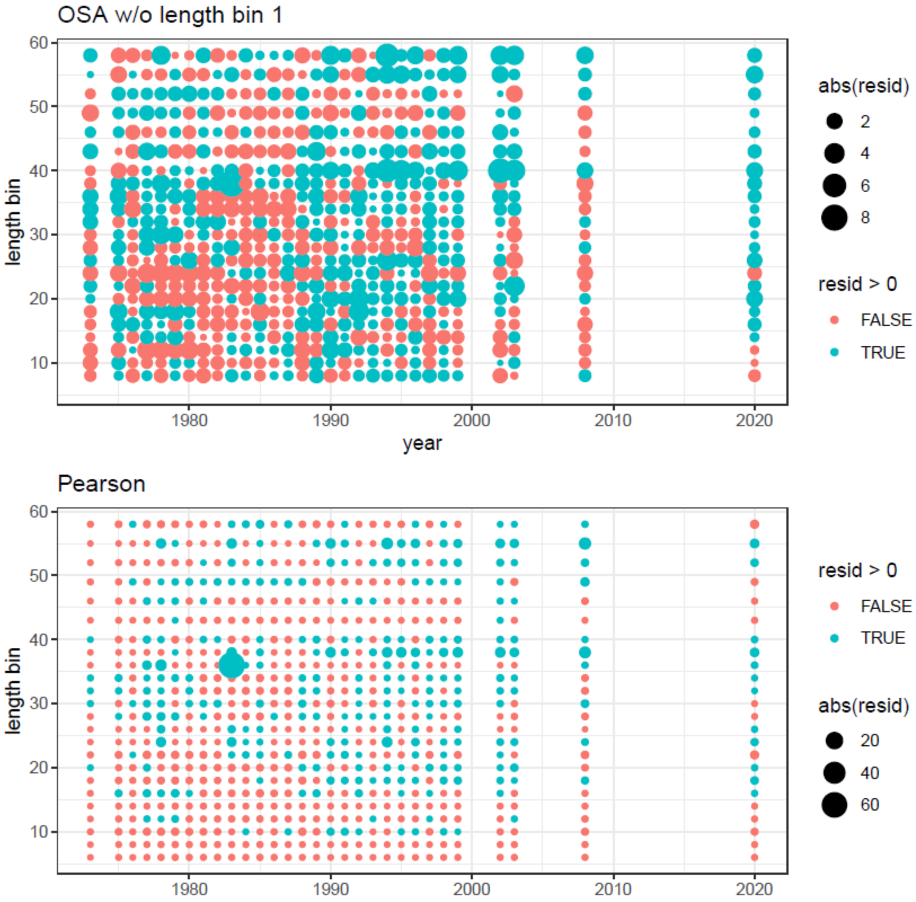
BSAI Flathead fish lengths length bins reversed





BSAI Flathead fish lengths length bins randomized





Summary

- OSA residuals have better statistical basis and interpretation [N(0,1)] for assessments. I **don't recommend normality tests**.
- For a year, they are the residual <u>conditioned on all previous bins</u>
- There is no single residual for an observation, it depends on order and which bin is dropped.
    - But each realization will have the correct statistical properties
- Can be calculated externally for ADMB models, or internally for TMB models with hierarchical structure.

Recommendations for 2024 and beyond

“Pearson residuals could therefore be wrongly interpreted and result in rejecting a suitable model. This could have direct consequences when these residuals are used for validating stock assessment models. Poor composition residuals are typically the basis for changing estimates of selectivity, which could lead to changes in the resulting management advice.” Trijoulet et al. 2023

It is time to retire Pearson residuals at the AFSC?

I propose AFSC authors

1. Report OSA residuals (bubble + QQ plot) and compare to Pearson (1 year)
2. Use visual inspection to detect model misfit and justify model changes
3. Do not formally use statistical tests (no p-values)
4. Continue research on AFSC applications

References

Havron, A, C.C. Monnahan, J.T. Thorson, K. Kristensen (in prep). A practical guide to residual analysis for validation of mixed effects models in Template Model Builder.

Thygesen, U. H., C. M. Albertsen, C. W. Berg, K. Kristensen, and A. Nielsen. 2017. Validation of ecological state space models using the Laplace approximation. Environmental and Ecological Statistics 24:317-339.

Trijoulet, V., C. M. Albertsen, K. Kristensen, C. M. Legault, T. J. Miller, and A. Nielsen. 2023. Model validation for compositional data in stock assessment models: Calculating residuals with correct properties. Fisheries Research 257:106487.

https://github.com/fishfollower/compResidual