

# Experiences working with longline data: catch-effort and associated datasets

IATTC 1st workshop on data improvement: industrial longline fishery  
9-11 Jan 2023

Simon D. Hoyle

Climate, Freshwater & Ocean Science



**NIWA**

Taihoro Nukurangi

# Introduction

- What longline data are needed for standardizing CPUE and for developing stock assessments?
  - Aggregated versus operational data
- Overview of experiences
  - IATTC, WCPO, IOTC, ICCAT, CCSBT
- Size data

# Issues for CPUE indices

- Usually the most influential component of a stock assessment
- Starting points
  - Basic equations
  - Fishery definitions (what are you analysing)
- Data
  - Exploring and preparing data
  - Misreporting
  - Density and catchability covariates
  - Data aggregation
  - Multispecies targeting
  - Environmental variables
  - Combining survey and CPUE data
- Analysis
  - Tools
  - Spatial considerations
  - Setting up and predicting from the model
  - Error distributions
  - Uncertainty estimation
  - Model diagnostics
  - Model selection
- Using CPUE in stock assessments

# List of Good Practices – those in red need operational data

1. Indices of abundance are the most influential part of an assessment, so invest accordingly.
- Data
2. Defining fleets is key. Definitions depend on stock and fishery structure, the type of CPUE analysis (time + space vs STM), and the assessment approach (biomass dynamic vs age/size-structured, conventional vs index fishery).
  - a. Explore and understand data (catch, effort, CPUE, sizes, ages, maturity, gear types, logbook types, vessel turnover, potential for misreporting, etc.). Plot the \*\*\*\* out of everything.
  - b. Talk to fishermen.
  - c. Use clustering techniques on species comps to explore fishing tactics
3. Structural changes based on understanding the system can be very important for the assessment, whereas many other issues just cause small changes in the index trends.
4. Revisit data exploration when updating indices. Don't just turn the handle.
5. Identify likely covariates before you start modelling. Avoid data dredging.
6. Differentiate between catchability and density variables.
7. Always include the variables that affect catchability – this is usually more important than the type of model you use. Think about potential for bias due to missing variables.
8. Consider targeting and target change through time, and how to address it. Understand the fleet well enough to know what the targeting strategies might be.
- Analysis
9. GAMs and STMs are better than GLMs. GAMs are best for exploration. STMs can be better for the final model(s). Each has unique capabilities.
10. Model the whole stock if you can do so without dropping important covariates due to data gaps or difficult spatial interactions.
11. Test your model by simulating.
12. Build multiple models using different approaches to develop your understanding of how the models are working. Start simple.
13. Use influence plots to understand how the variables and their values affect the indices.
- Assessment
14. Use the index fishery approach if you can.
15. Assume effort creep. There are catchability changes that your model hasn't captured, and  $q$  increases are almost inevitable in the long term.
16. Don't blindly split indices and assume the model will scale them correctly – this is unlikely.
17. Don't include several conflicting indices in the assessment at the same time.

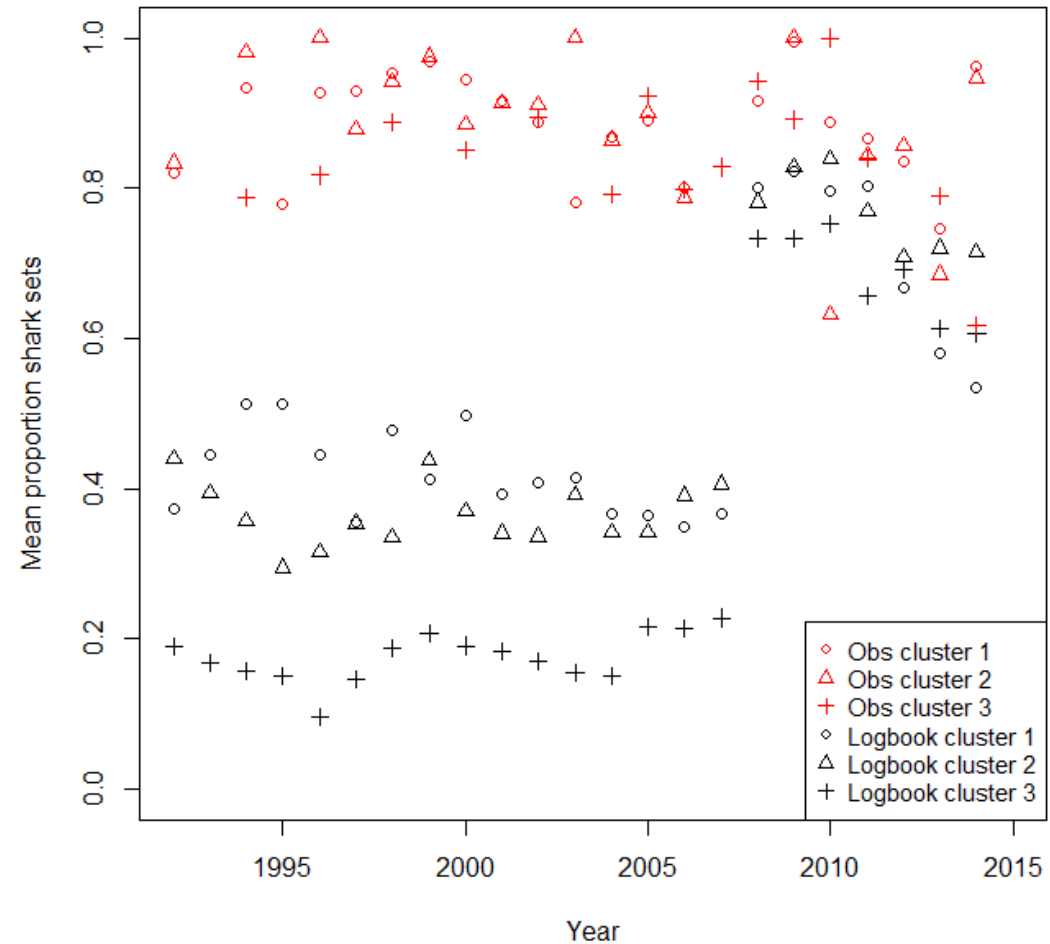
# CPUE issues that need operational data:

## 1. Starting points

- Start by exploring the data, to understand the fish population, and the fishery and its history. This is the step with the largest potential impact on the assessment.
- Identify the key sources of information.
  - Plot everything, including biological data.
    - Effort, catch, CPUE, sizes, sex ratios, maturity data, gear covariates.
    - Spatial patterns through time.
    - Are there seasonal spawning areas?
    - What changes through time in gear, logbooks, regulations, rates of misreporting?
  - Talk to fishers
- Use this information to determine the approach you use to standardize the data.
  - Define fleets / areas, identify potential covariates, change points in fishing practice.

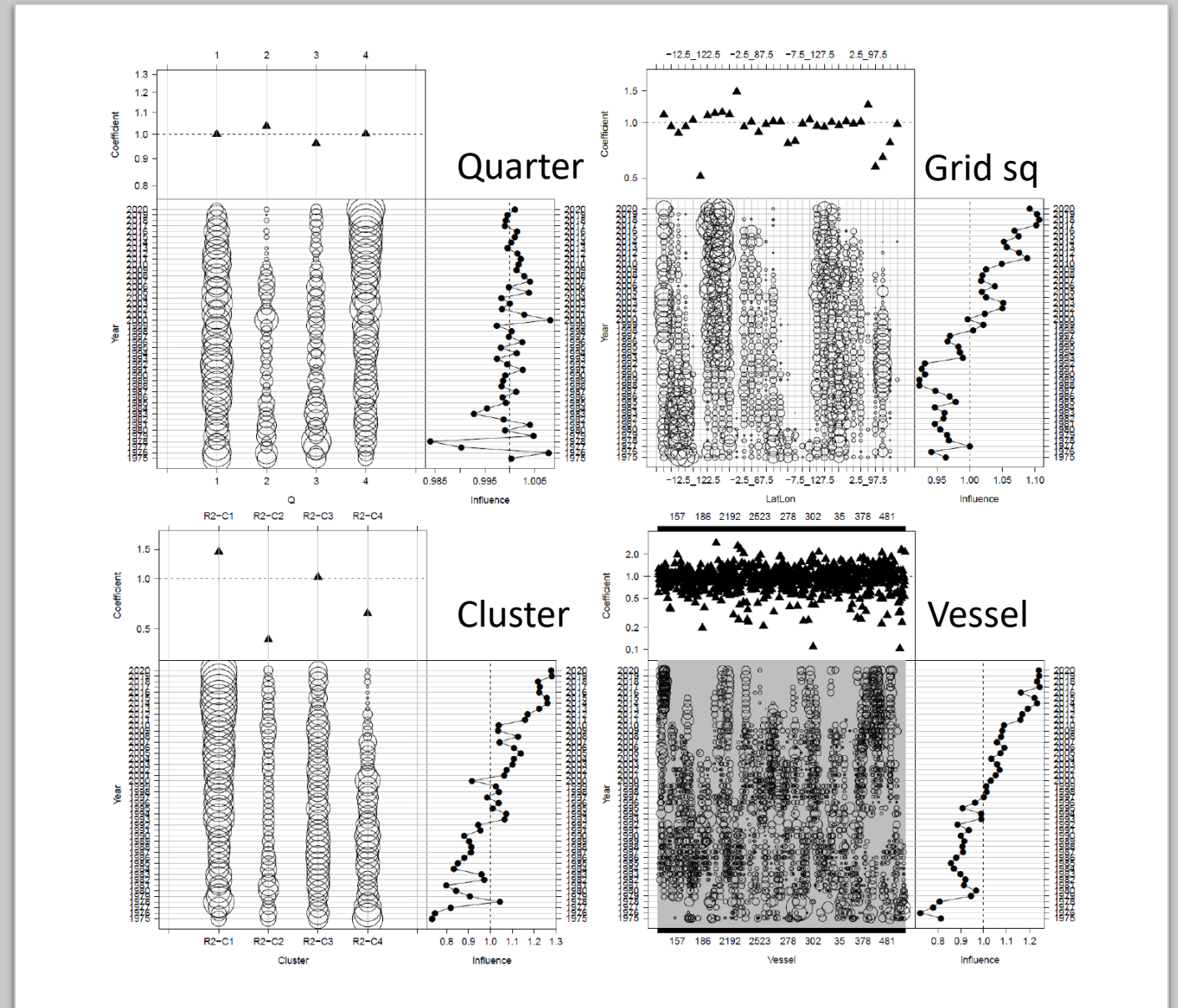
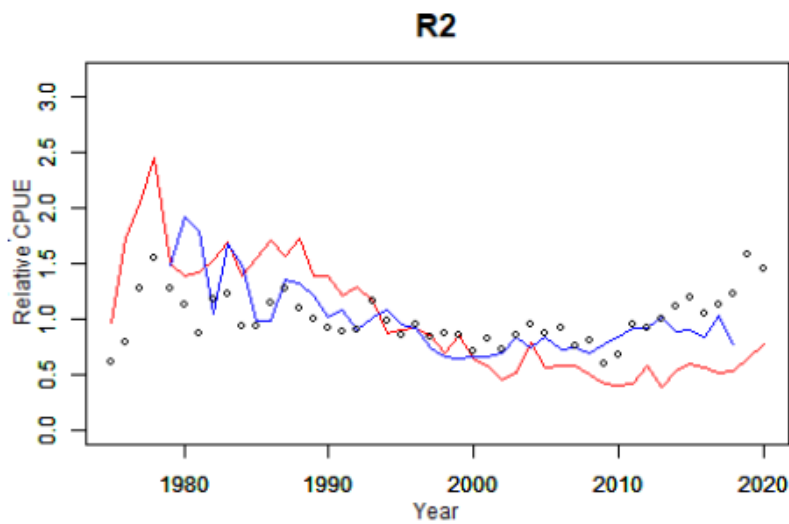
## 2. Temporal change in rate of misreporting

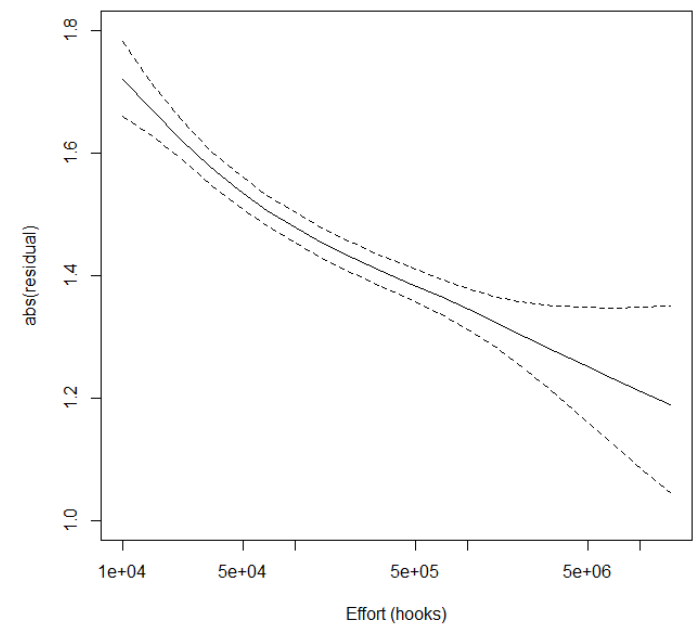
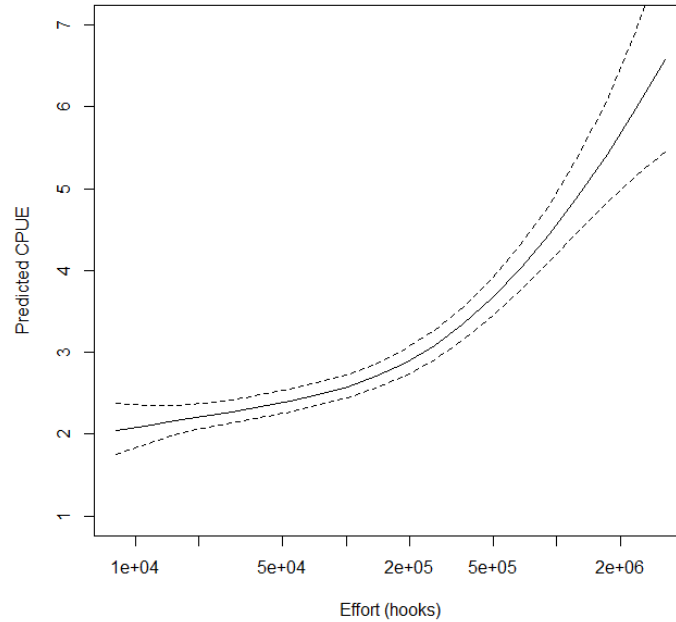
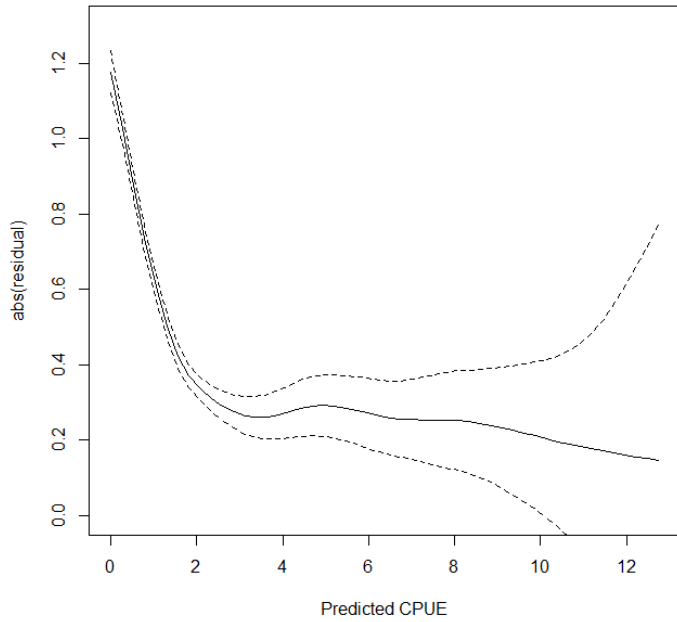
- Index of abundance for porbeagle sharks
- Proportions of sets with nonzero shark catch
- In 2008 there was a transition
- Logbooks suddenly reported shark catches at almost the same rate as observers.



# 3. Covariate effects

- Influence plots show the effects of each covariate on the trend
- Japanese bigeye CPUE in eastern tropical Indian Ocean (Matsumoto 2022).
- Covariates have large impact on index trend





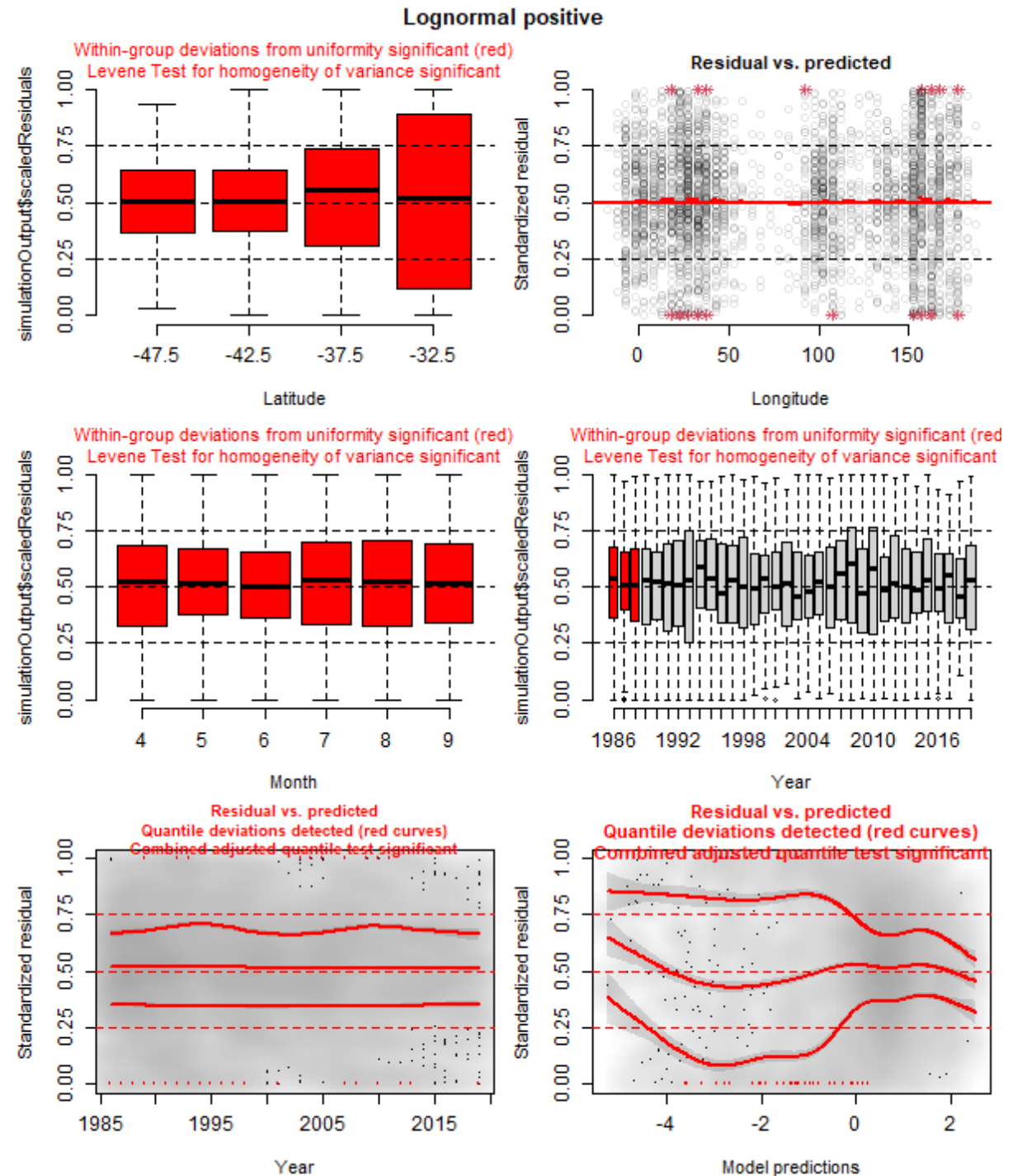
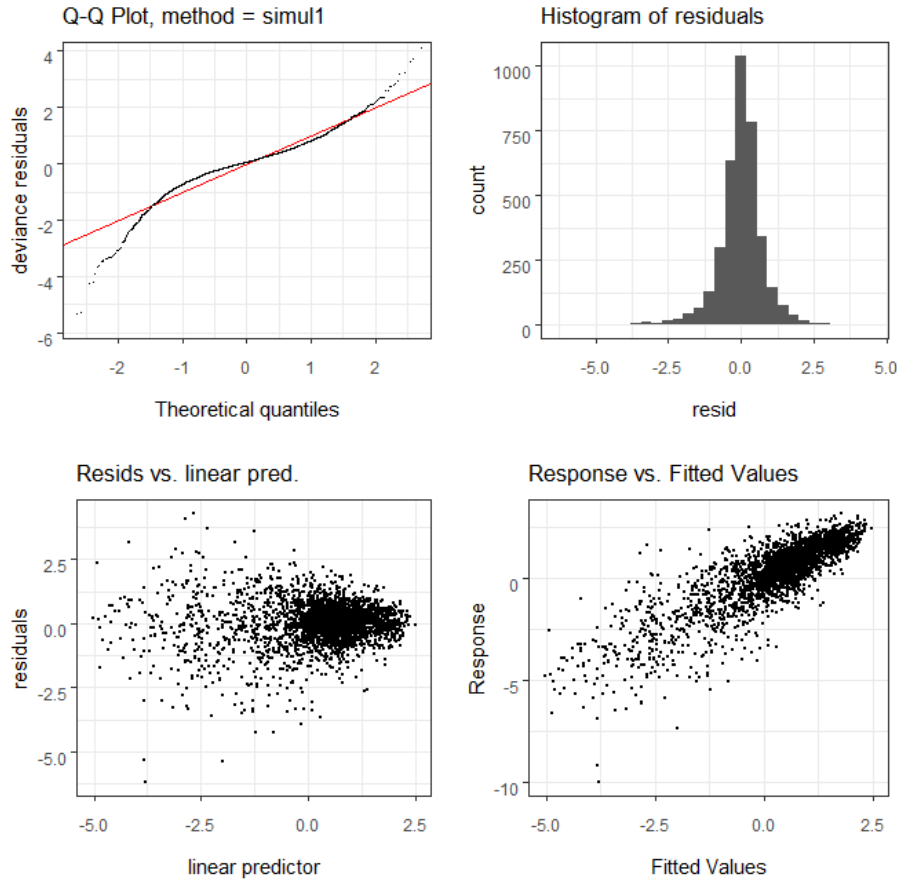
## 4. Data aggregation and error distributions

(Analyses of aggregated Japanese SBT CPUE data for CCSBT)

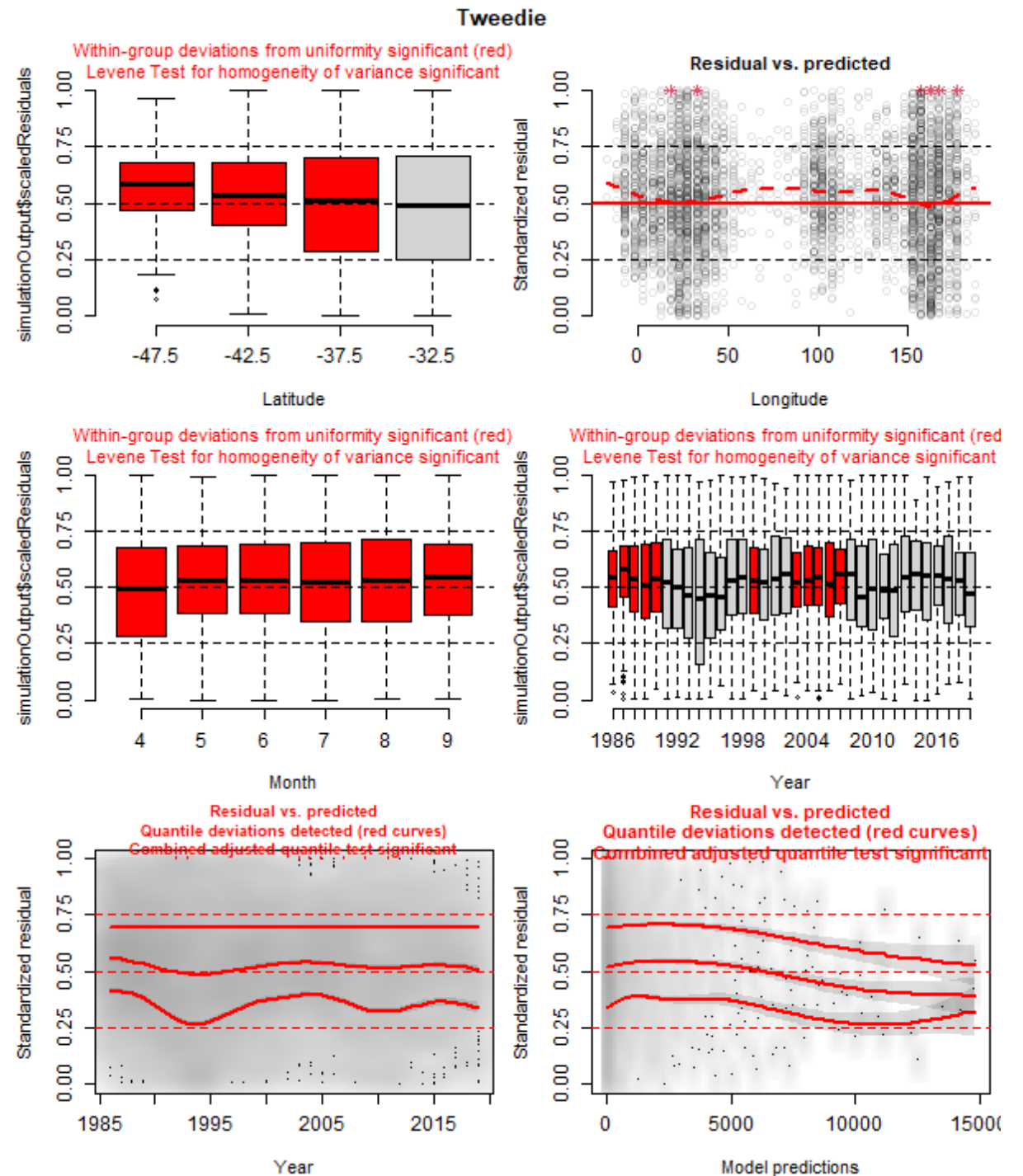
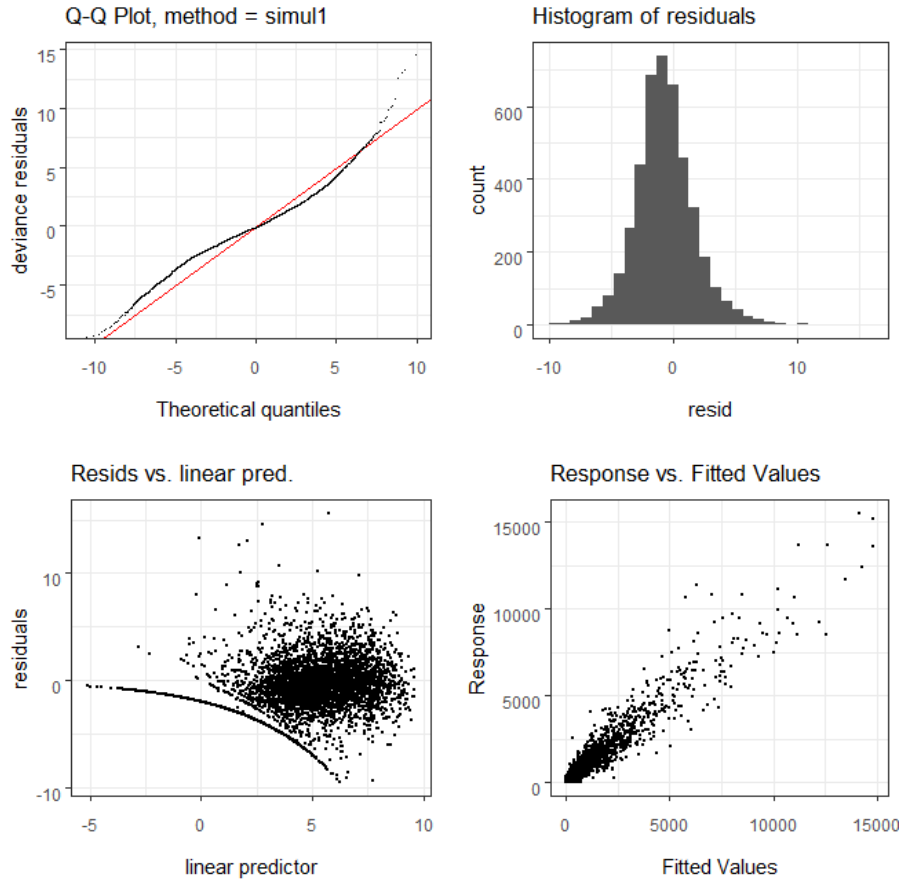
Variance is inversely related to CPUE



# Diagnostics for aggregated data: lognormal positive



# Diagnostics for aggregated data: Tweedie



# 5. Target change and covariates

- Target change is a particular challenge
  - Big impacts on indices in some cases, and important to address in multispecies fisheries
  - Almost all methods require operational data
- Various approaches in use to identify métiers / targeting practices
  - Data subsetting based on covariates (season, location, gear use)
  - Catches of other species – many approaches
    - E.g., Clustering on species composition (He et al 1997)
  - Success of methods vary in different situations - simulations.
  - Need to understand the fisheries, & ensure métiers are consistent with known targeting strategies.

# Data aggregation - overview

- Effects of aggregation
  - Less ability to understand the structure & behavior of fleet and target species.
  - Missing variables (vessel, set characteristics) that may affect catchability.
  - Loss of spatial and temporal resolution.
  - Loss of information to identify targeting.
  - Changed variance structure which degrades fit.
- Overall – more bias and uncertainty in indices, and less appropriate model structure.
  - These reduce assessment reliability and increase risk.

# Data aggregation - CCSBT example (2022)

- Old CPUE analysis method using GLM with aggregated data broke down (model failure) due to increasing effort concentration and sparse data.
- Developed new method using GAMs (spatiotemporal smoothers) and aggregated data which was more stable, much less biased, and much more precise.
- GAM aggregated and operational data versions, with the same model structure, produced indices that were slightly (but significantly) different, in ways that mattered for the assessment.
- GAM model chosen for SBT assessment uses operational data.

# My experience with CPUE analyses of DWFN operational logbook data

1. WCPFC (SPC) – catch-effort logbooks for albacore recorded at Pago Pago (2008-2013)
2. WCPFC (SPC) – logbooks submitted to countries on fishing in EEZ (2008-2013)
3. WCPFC (SPC) – collaborative work with Japanese scientists (2010-2013)
4. IOTC – joint analyses with JPN, KOR, TWN, SYC scientists (2015-2019)
5. ICCAT – joint analyses with JPN, KOR, TWN, BRA, CHN, USA scientists (2018-2019)
6. WCPFC (NIWA) – collaborative analyses of porbeagle shark CPUE by country with JPN, ARG, URY, CHL, NZL (2016-2017)

# Processes for developing joint analyses

- IOTC and ICCAT, 2015 to 2019
  - Analyses led by consultant, working with national scientists.
  - All participants meet for a period of 1 to 3 weeks.
    - Clean and prepare datasets, develop code, run analyses, troubleshoot, prepare indices & diagnostics.
  - All datasets shared with everyone present – high level of trust involved.
  - Datasets deleted at end of meeting.
- IOTC and ICCAT, 2020 to present
  - Similar process but analyses led by Japanese scientist, collaboration among DWFNs.
  - Covid prevented meetings 2020-2022, so analyses used data aggregated 1 x 1 x vessel x year x month.
  - Some resulting indices were different from 2018-19 versions, possibly due to the aggregation.
- Alternative model: WCPFC
  - National datasets held on a secure server at SPC with restricted access but more time.
  - More limited range of covariates.
  - Some aggregation? Unsure.

# Joint analyses

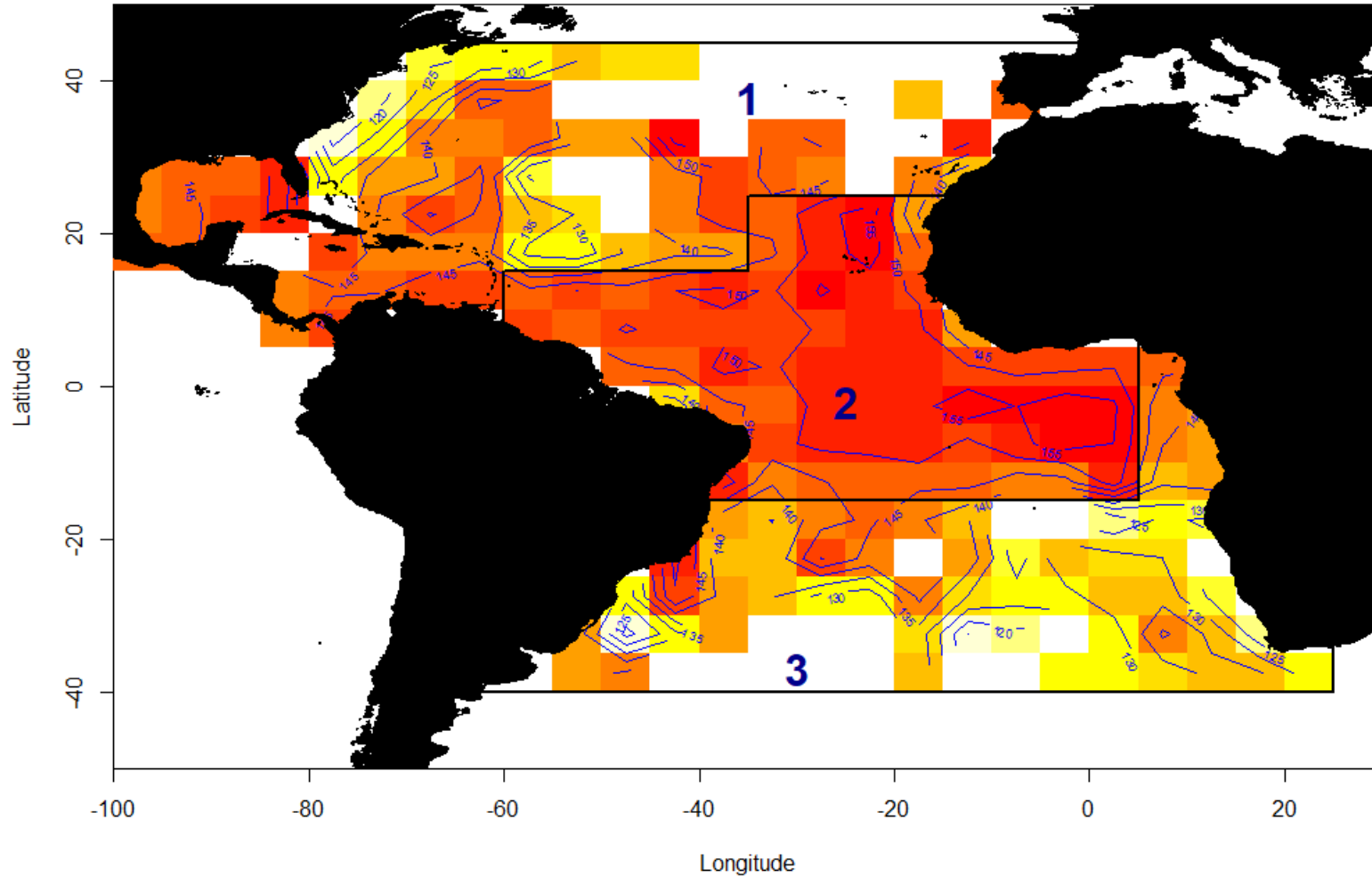
- Advantages
  - Expertise of national scientists is vital for understanding the datasets and fisheries.
  - Comparisons between datasets help to identify problems and solutions.
  - Joint datasets have better spatial and temporal coverage, although variables differ.
  - Collaboration builds understanding, consensus and acceptance of results.
  - Independence of lead analyst supports acceptance of assessment results.
- Problems
  - Not enough time
    - Little time to think, develop understanding, do research.
    - Problem solving is rushed and new approaches unlikely.
    - Hard to produce alternative models for risk analysis.
    - Promotes recycling of the same approaches every time - minor changes only.
    - Can't run slower models such as VAST
    - Risk of error when rushing to finish analyses.
  - Risk of 'groupthink' – not trying different approaches



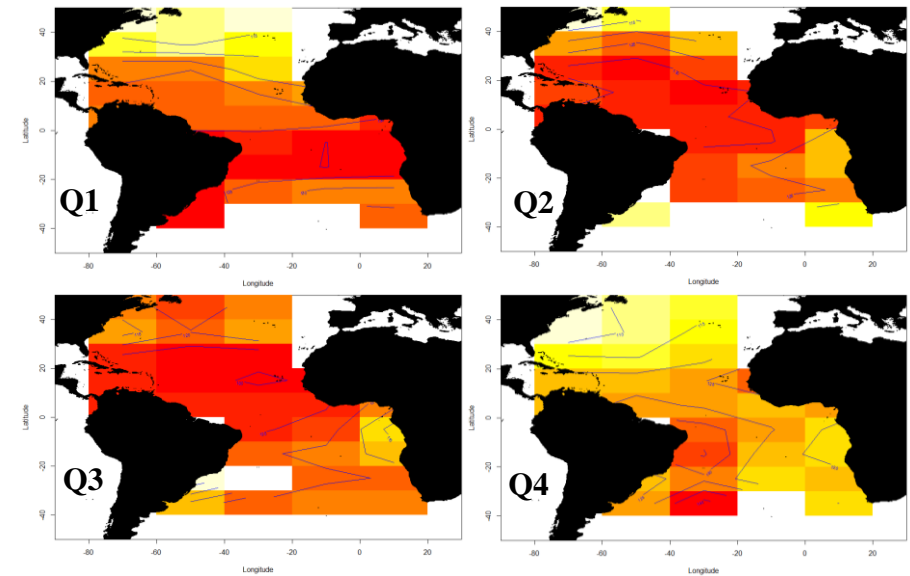
# Size variation in time and space

- Sizes of tuna and other species vary in time and space.
- These size patterns affect stock assessment results, and it is important to understand what causes them.
  - Currently they are not well described or understood.

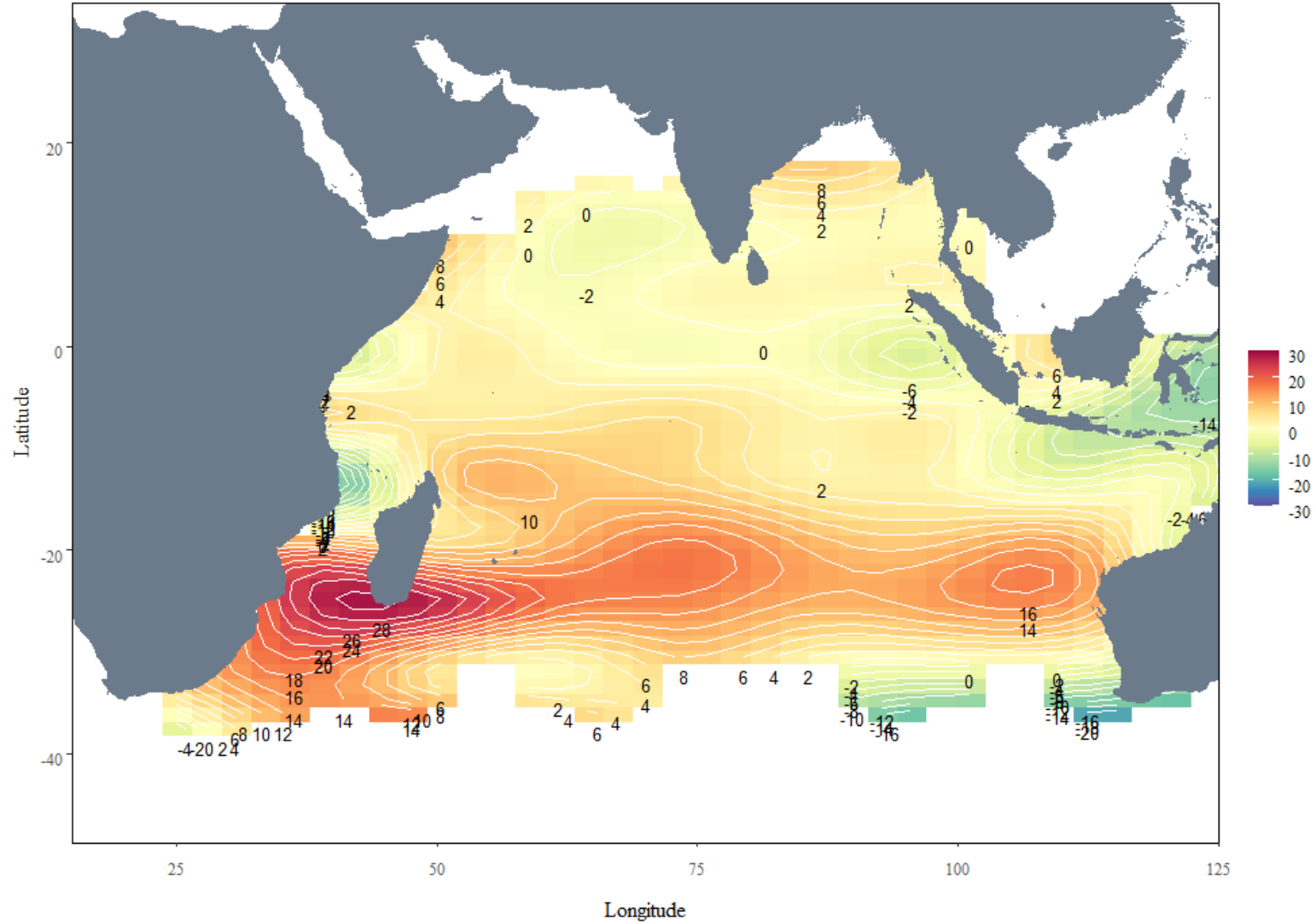
# Atlantic Ocean yellowfin length maps



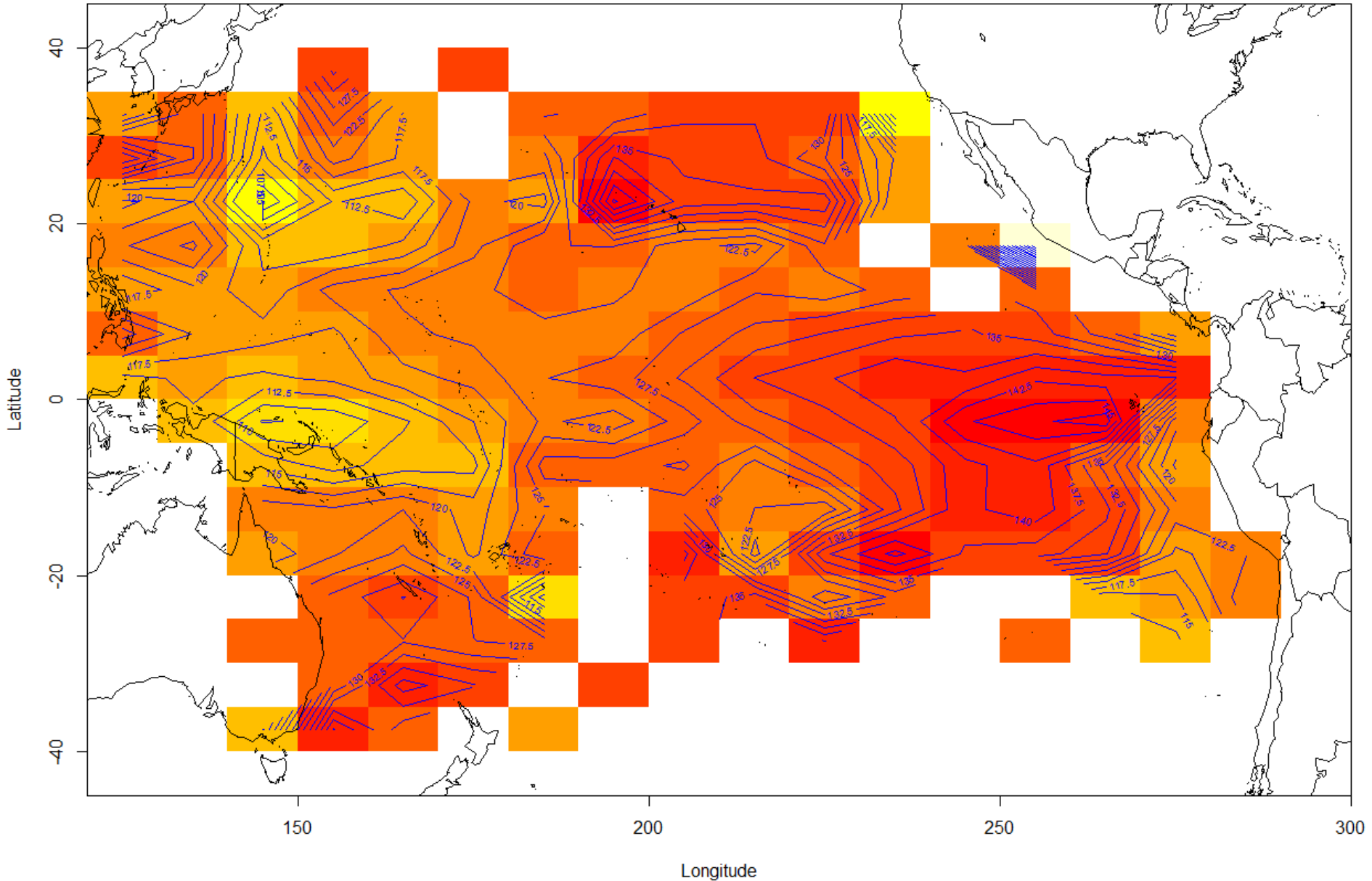
Size patterns change by season



# Indian Ocean yellowfin length map



# Pacific Ocean yellowfin length map



# Why do size patterns matter?

- Catch-at-length models use size data to provide information about  $Z$ 
  - Relationship with asymptotic length
  - Decreasing size suggests higher  $Z$
- Size data also standardized with VAST models as part of the index fishery approach.
  - Which covariates should be included in these models?
- What causes the size variation? This needs to be explored.
  - Age structure (as the assessment model assumes)
  - Growth variation (likely to some extent – would cause bias)
  - Selectivity variation (possible – would cause bias)

## Data needs:

- to explore size variation and include in models

- Logbook & observer data
  - Representative size data associated with catch
  - Gear configuration and set characteristics
  - Length, weight, sex, hook number, catch depth
- Biological studies
  - Ageing studies stratified by area and season

# Summary

- Operational data can provide significantly better understanding of fishery dynamics, substantially improve CPUE indices, and reduce assessment uncertainty.
- Joint CPUE analyses across multiple datasets have many advantages. Indices are too influential and the analyses too important to be done in a rush. Best results would come from long-term collaboration and data sharing.
- Size data are very influential in assessments, but factors driving size patterns are poorly understood. Better assessments require both time series of representative size data, and better understanding through research using detailed operational logbook and observer size data and focused biological studies.

Thank you

Climate, Freshwater & Ocean Science



**NIWA**

Taihoru Nukurangi