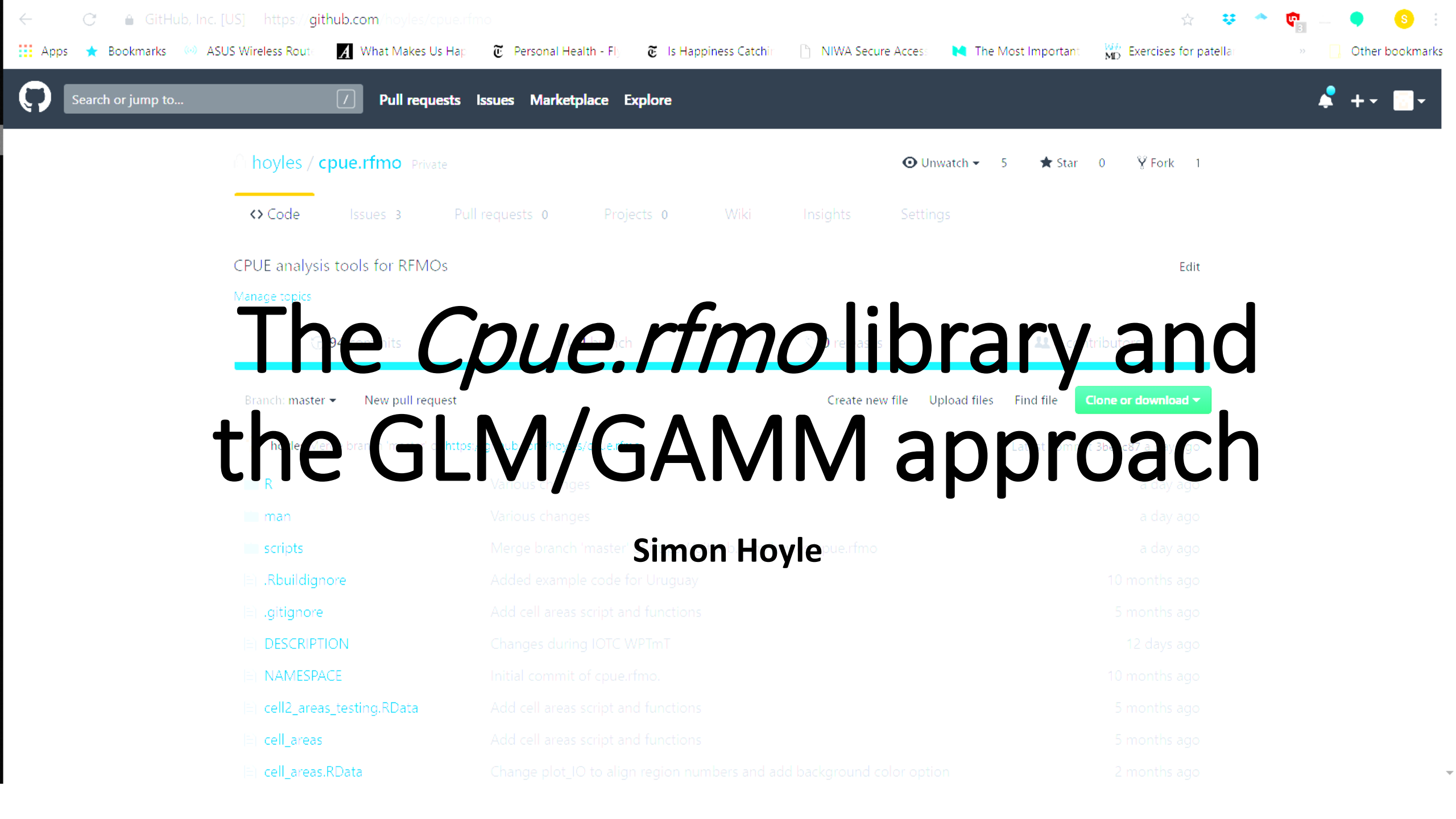


The *Cpue.rfmo* library and the GLM/GAMM approach

Simon Hoyle

Introduction

- Cpue.rfmo: what is it?
- Purpose of library
- Structure
- Development needs
- What next?

GitHub, Inc. [US] | https://github.com/hoyles/cpue.rfmo

Apps Bookmarks ASUS Wireless Route What Makes Us Happ Personal Health - Flyi Is Happiness Catchin NIWA Secure Access The Most Important Exercises for patellar Other bookmarks

Search or jump to... / Pull requests Issues Marketplace Explore

hoyles / cpue.rfmo Private

Unwatch 5 Star 0 Fork 1

Code Issues 3 Pull requests 0 Projects 0 Wiki Insights Settings

CPUE analysis tools for RFMOs

Edit

Manage topics

94 commits 1 branch 0 releases 4 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

hoyles Merge branch 'master' of https://github.com/hoyles/cpue.rfmo Latest commit 3becc87 a day ago

R	Various changes	a day ago
man	Various changes	a day ago
scripts	Merge branch 'master' of https://github.com/hoyles/cpue.rfmo	a day ago
.Rbuildignore	Added example code for Uruguay	10 months ago
.gitignore	Add cell areas script and functions	5 months ago
DESCRIPTION	Changes during IOTC WPTmT	12 days ago
NAMESPACE	Initial commit of cpue.rfmo.	10 months ago
cell2_areas_testing.RData	Add cell areas script and functions	5 months ago
cell_areas	Add cell areas script and functions	5 months ago
cell_areas.RData	Change plot_IO to align region numbers and add background color option	2 months ago

Purpose

- R library of scripts and functions
- Developed to standardize RFMO tuna catch and effort data efficiently
- A tool for sharing methods among scientists and RFMOs
- Scripts to handle the processes of
  - Importing & cleaning
  - Characterising
  - Clustering
  - Standardizing

Structure

- Scripts for each RFMO
  - Analysis script for each fleet
  - Figures for each fleet
- Functions
  - Grouped by category

Development needs

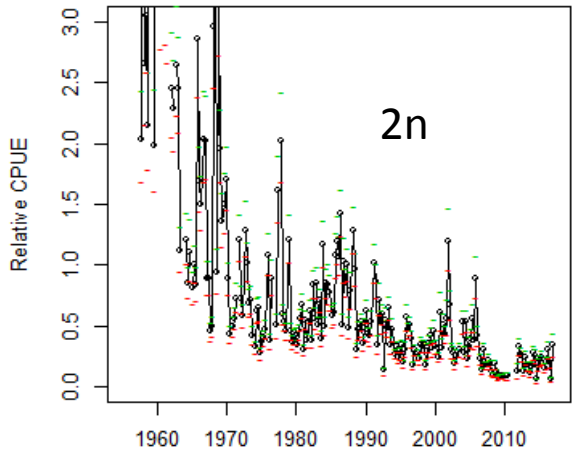
- Many! Such as…
- Package-specific
  - Improve documentation
    - Wiki and ‘how-to’ on front page
    - Improve help files
  - Fix package so it can be installed using devtools ‘install_github()’
  - Rationalize scripts and reduce number
- Statistical methods
  - Add scripts and functions for new methods
  - e.g. Okamura 2018, VAST
- Public version

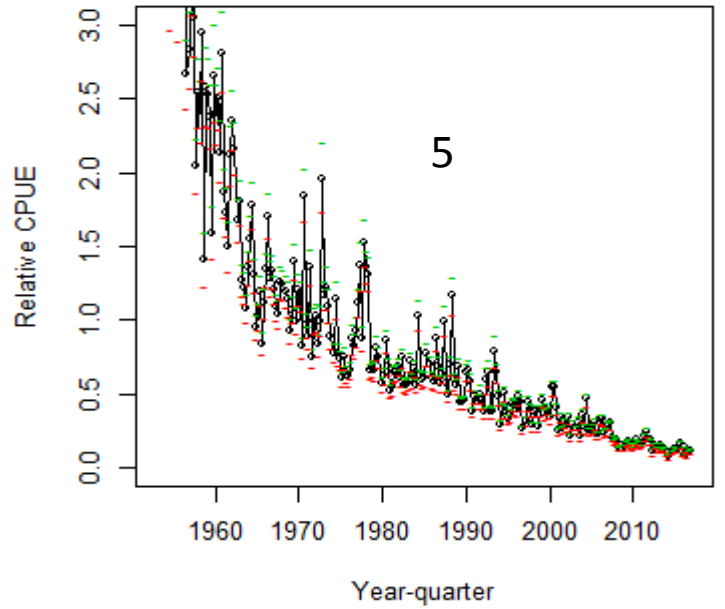
CPUE standardization methods, fundamentals 1

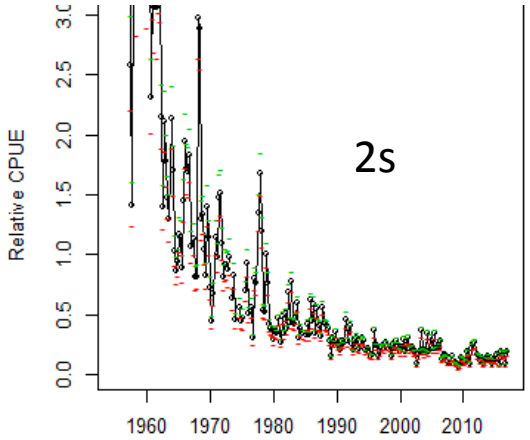
- Generalized linear models in R, modelling CPUE at the set level
  - Fixed effects for all parameters, such as vessel id.
- Lognormal constant
  - $\ln(CPUE_s + k) \sim yrqtr + vessid + latlong5 + f(hbf) + g(hooks) + cl + \epsilon$, where *k* = 10% of the mean CPUE.
- Delta lognormal
  - $(CPUE = 0) \sim yrqtr + vessid + latlong5 + f(hbf) + g(hooks) + cl + \epsilon$
  - $\log(CPUE) \sim yrqtr + vessid + latlong5 + f(hbf) + g(hooks) + cl + \epsilon$, for nonzero sets
- *HBF* and *hooks* parameters are cubic splines
- *yrqtr*, *vessel*, *latlong5*, and *cluster* are categorical variables

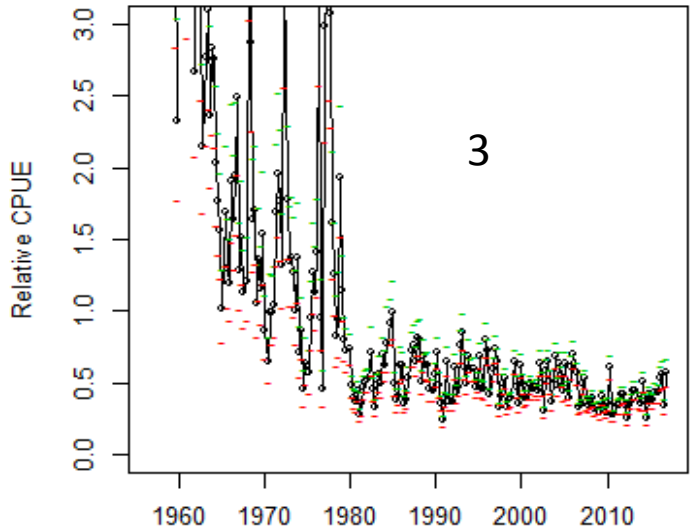
Approach to modelling spatial effects

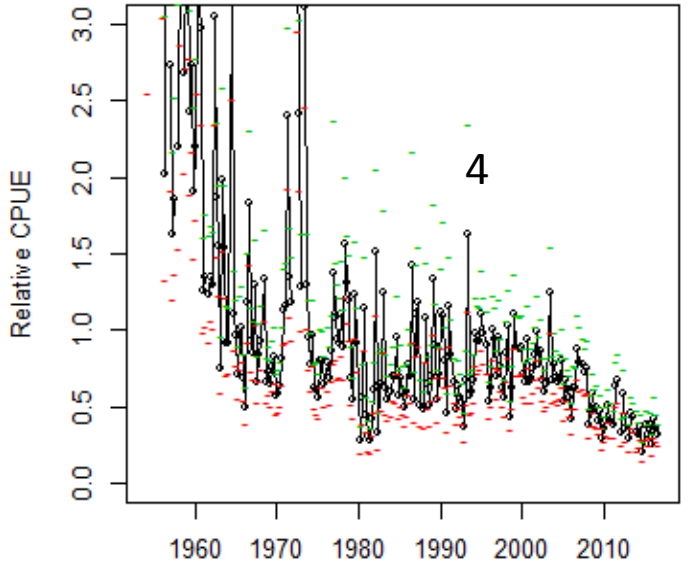
- Assessment regions are modelled independently, consistent with their treatment in the assessment.
- Within a region, 5° cells are modelled as independent categorical variables.





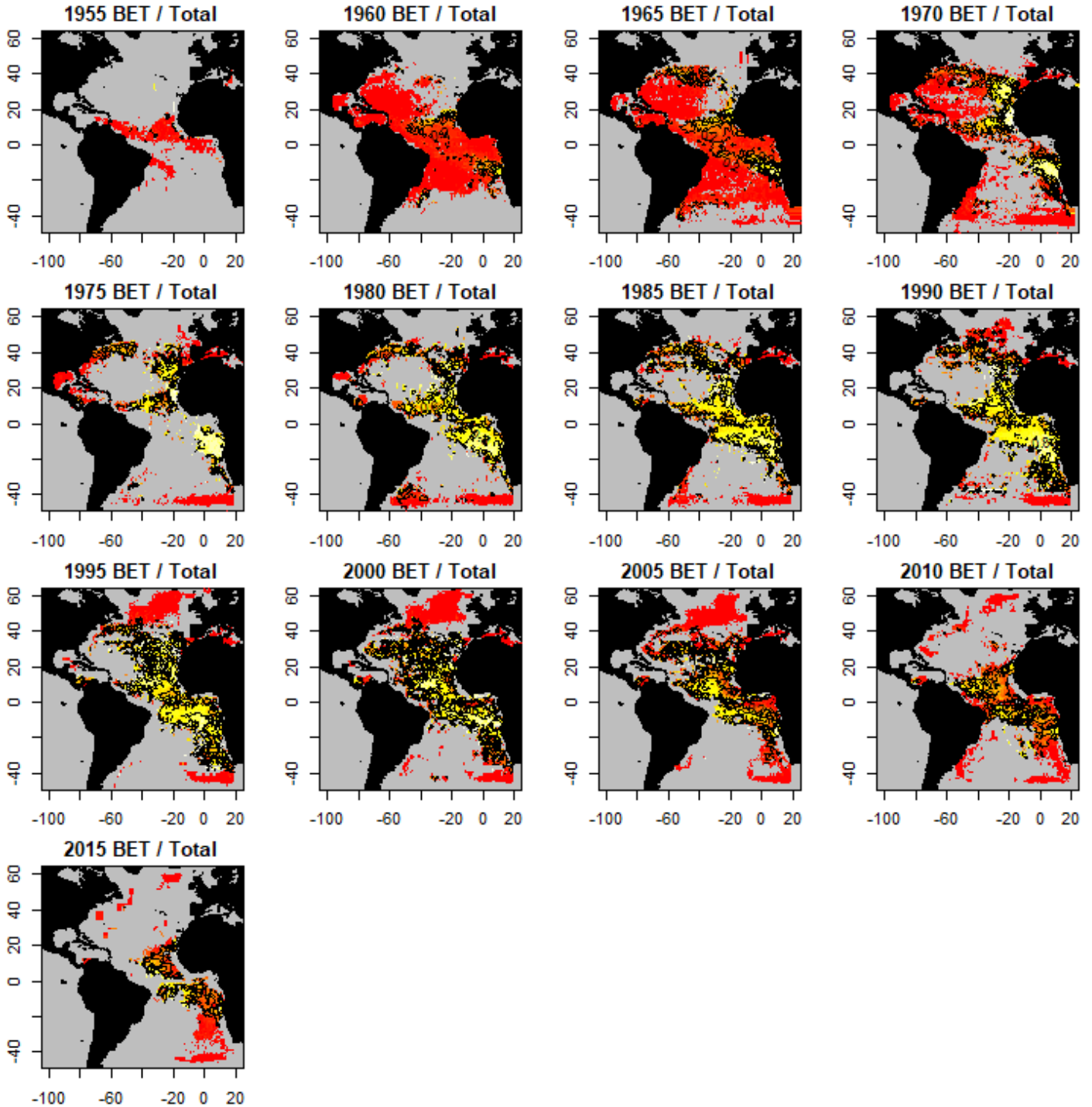






Changes in distribution & coverage

- In a 60 year dataset (since 1958), effort concentration moves around
  - Causes
    - Initial expansion into new areas
    - Area closures due to EEZs
    - Markets changing target preferences, e.g. sashimi market raising value of BET/YFT vs ALB
    - Fuel costs, competition
  - Effects
    - Areas without effort
    - Changing statistical weights among areas, biasing the indices



Within region: What to assume about areas without effort?

1. Time-area interactions often require spatial infilling, but this may be problematic (one size does not fit all)
    1. During expansion, unfished areas have high biomass & higher CPUE
        - Unfished areas never fished, so assume ~ initial CPUE in those areas
        - But catchability probably higher in the initial phase
    2. Later, when the index fleet leaves an area, what should we assume?
        - Fishing continues by other fleets (e.g. EEZ closed, outcompeted by other fleets)? B stays low.
        - Less fishing effort (e.g. piracy, or catches too low). B increases.
2. Within a region, model is CPUE ~ time + area, which avoids the need for infilling
    - Problematic to the extent that fish distributions change
3. Combined approach – explored
    - Time x area model (latlong5 + lat5 * qtr + lat5 * year)
    - Fill time-area ‘holes’ with estimates from time + area model

Biases due to changing effort distribution

- Shifting effort introduces bias. We do the following:
  1. Remove 5° cells with fewer than N1 sets across all years
  2. Randomly select N2 sets from each yq*cell stratum (applied when total # sets in dataset > limit of 60000)
  3. Adjust statistical weights to give each yq stratum the same influence (Punsly 1987, Campbell 2004)

     For set *j* in area *i* and year-qtr *t*, $w_{ijt} = \frac{log(h_{ijt}+1)}{\sum_{j=1}^{n} \log(h_{ijt}+1)}$

CPUE standardization methods, fundamentals 2

- Alternative model structures
  1. Data subset, Cluster + HBF + Hooks
     - $\ln(CPUE_s + k) \sim yrqtr + vessid + latlong5 + g(HBF) + f(hooks) + cl + \epsilon$
     - Data omits clusters catching very little of target species
  2. Data subset, Cluster + HBF
     - $\ln(CPUE_s + k) \sim yrqtr + vessid + latlong5 + g(HBF) + cl + \epsilon$
     - Data omits clusters catching very little of target species
  3. Data subset
     - $\ln(CPUE_s + k) \sim yrqtr + vessid + latlong5 + f(hooks) + \epsilon$
     - Data omits clusters catching very little of target species
  4. All data, HBF
     - $\ln(CPUE_s + k) \sim yrqtr + vessid + latlong5 + f(hooks) + HBF + \epsilon$
     - All clusters included in dataset

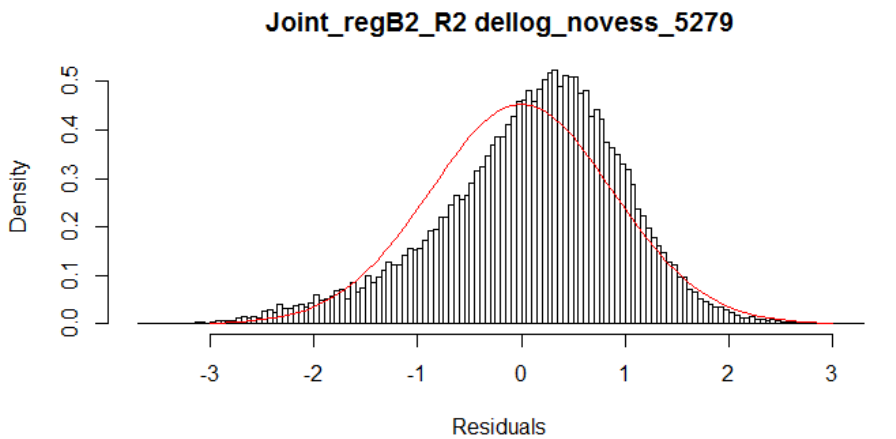
CPUE standardisation – some details

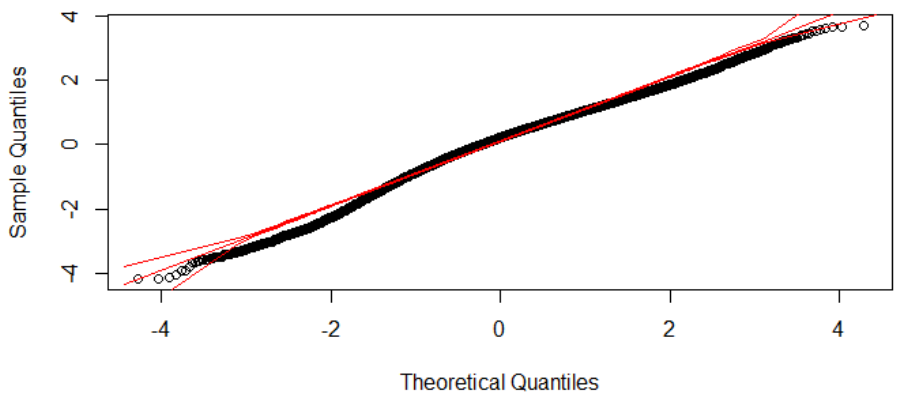
- Problems with large datasets
  - Very long runtimes with delta lognormal
  - Large memory use (> 16GB)
  - Hard to debug and fix problems
  - Limited benefit from extra precision – important sources of uncertainty are elsewhere
- Solutions
  - Reduce number of strata
    - Remove vessels fishing < N_1 qtrs
    - Remove cells, yr-qtrs, & vessels with < N_2 sets
  - Subsample data at random
    - Randomly sample (without replacement) N_3 sets from each year-qtr x cell stratum
    - Tested with WCPO data, indices stable with ~ 15 sets per stratum (Hoyle and Okamoto 2011)

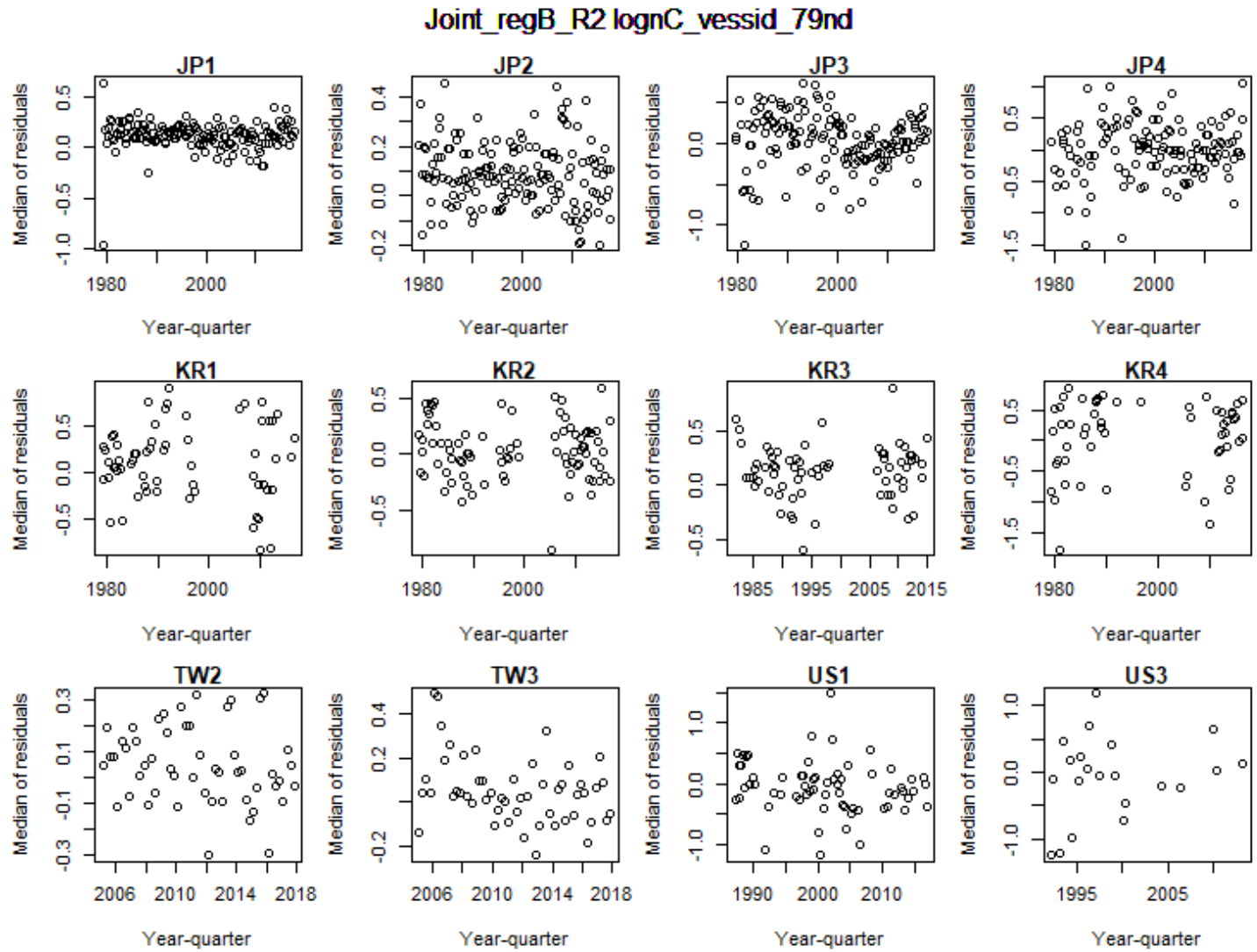
CPUE standardization methods 3

- 4 sets of indices
  - 1958 to recent, no vessel effects
  - 1958 to recent, with vessel effects
  - 1958 to 1979, no vessel effects
  - 1979 to recent, with vessel effects

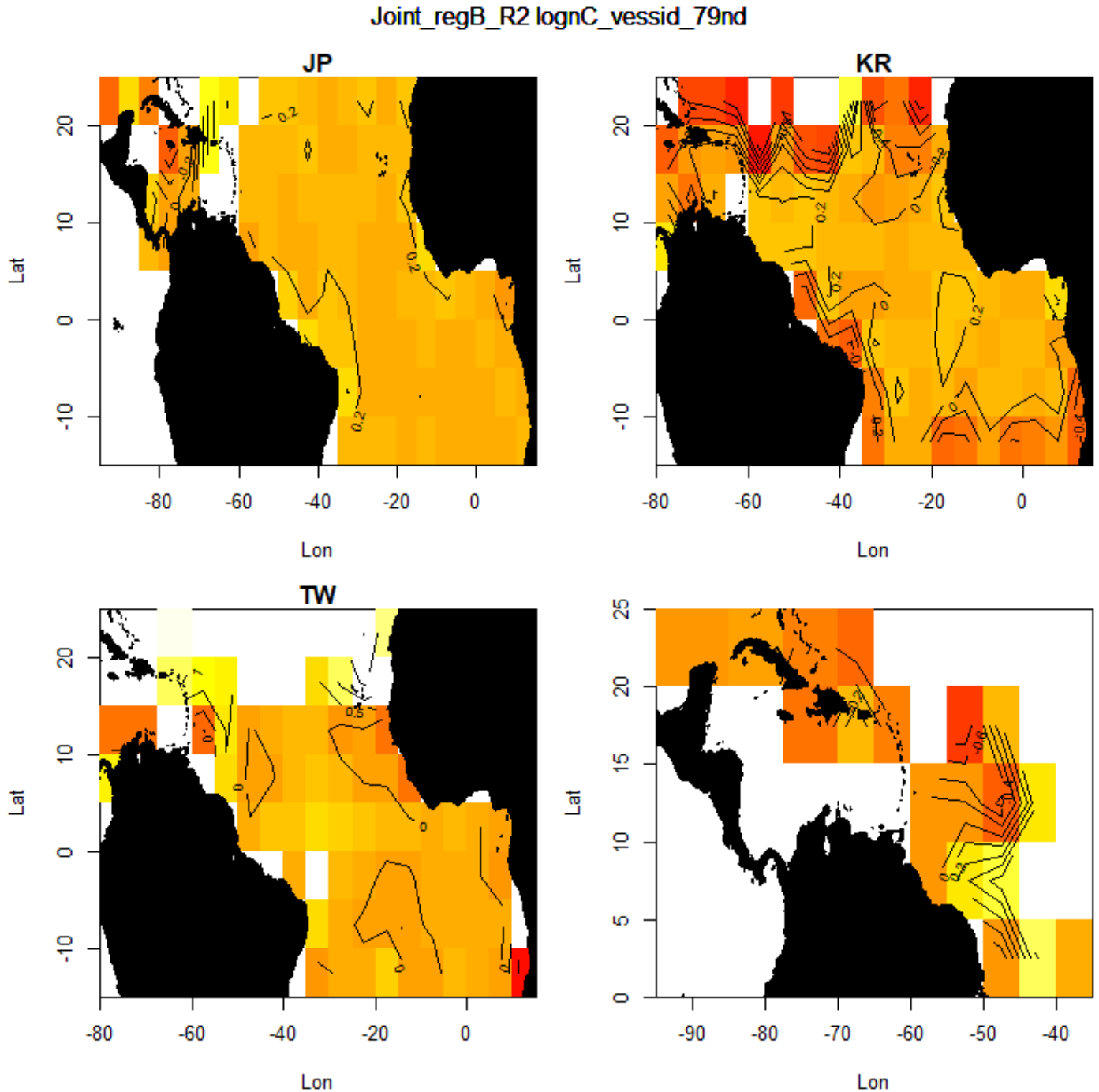
Diagnostic options 1



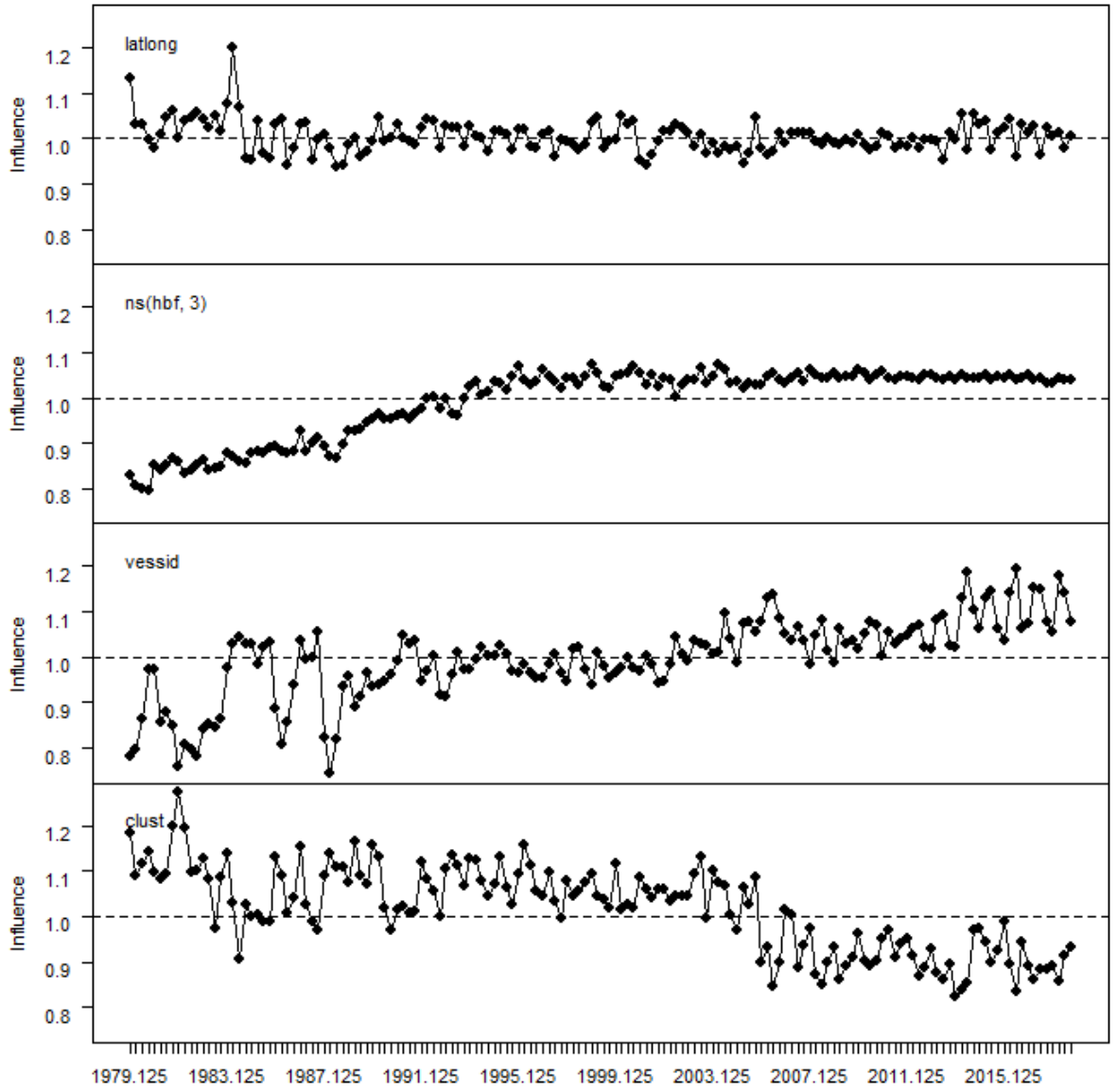




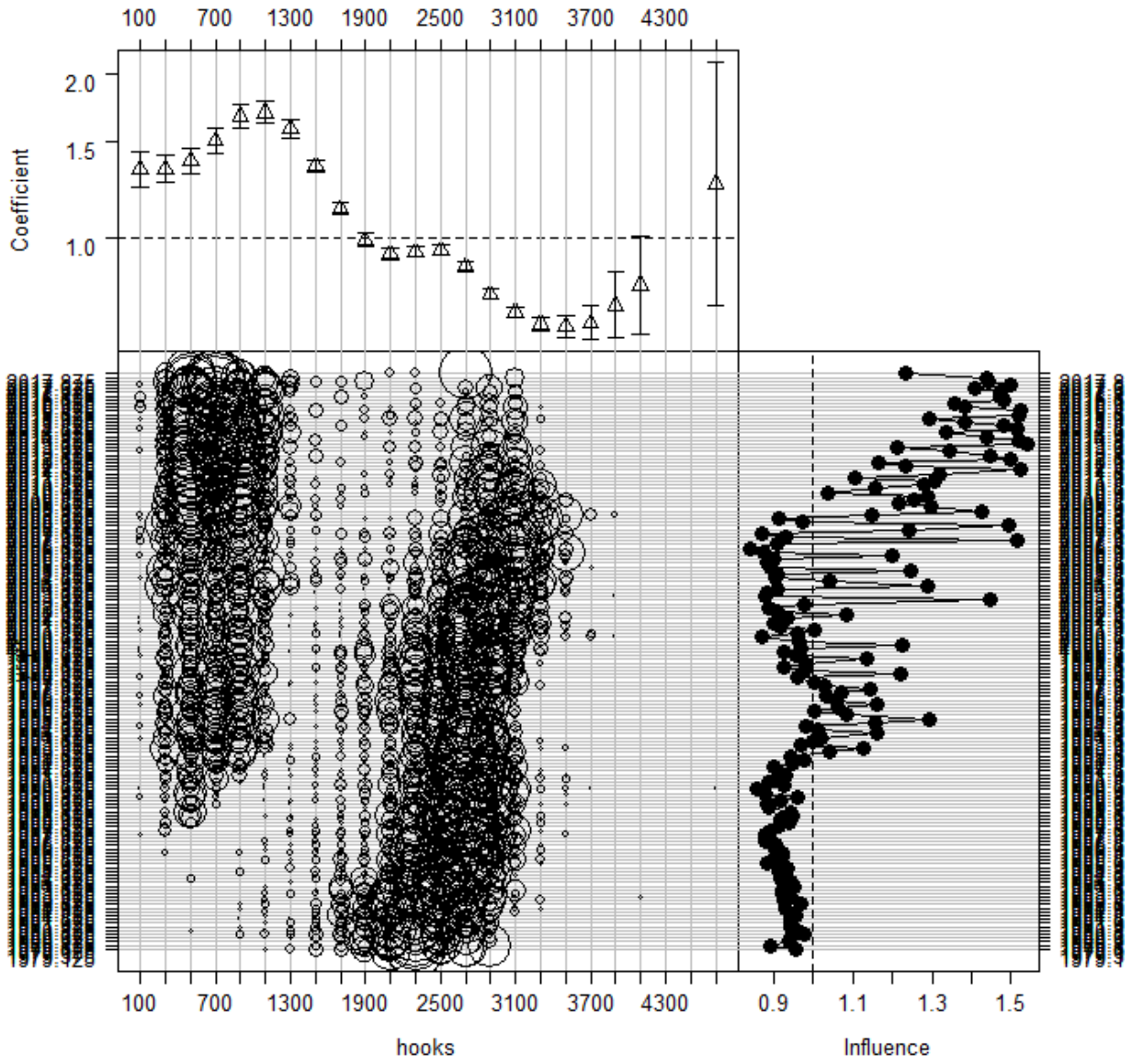
Diagnostic options 2



Diagnostic options 3 – Influence plots



R1 influence of hooks



Residual concerns

- Potential for differences between (& within) fleets
  - Factors not available for analysis
  - Different bait, gear configurations, reporting behaviour
  - Time series patterns in individual vessel behaviour
- Model issues
  - Assuming no interactions, e.g. between:
    - Targeting behaviour and vessel catchability
    - Season and spatial effects
- Possible future options
  - Random effects on e.g. vessel by target, to permit exploration of interactions
  - mgcv: as before, but add te(lat, lon, yr) + te(lat, lon, qtr)
  - VAST

CPUE standardization methods 3

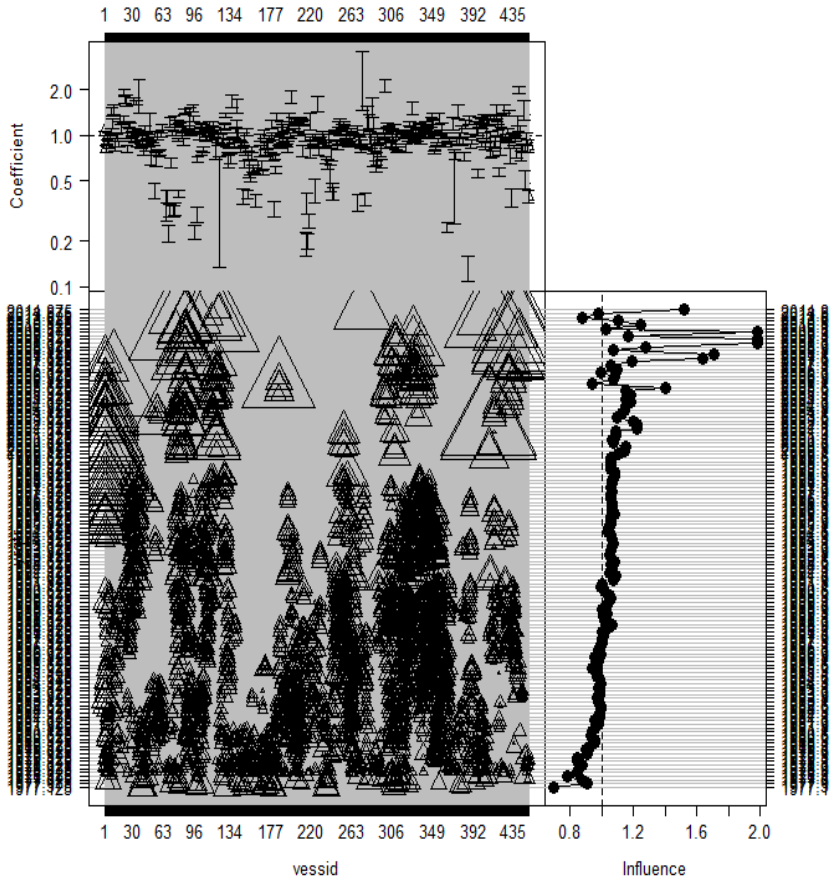
- 4 sets of indices
  - ‘novess_allyrs’ – 1958-2017 without vessel effects
  - ‘boat_allyrs’ – 1958-2017 with vessel effects
  - ‘novess_5879’ – 1958-1979 without vessel effects
  - ‘vessid_7914’ – 1979-2017 with vessel effects
- Possible option for assessments
  - Use 2 series: 1958-1979 (no vessel effects) and 1979-2017 (with vessel effects)
    - May be best to use just 1979-2017, due to uncertain consistency of pre-1979 indices

5. Assess fleet efficiency changes

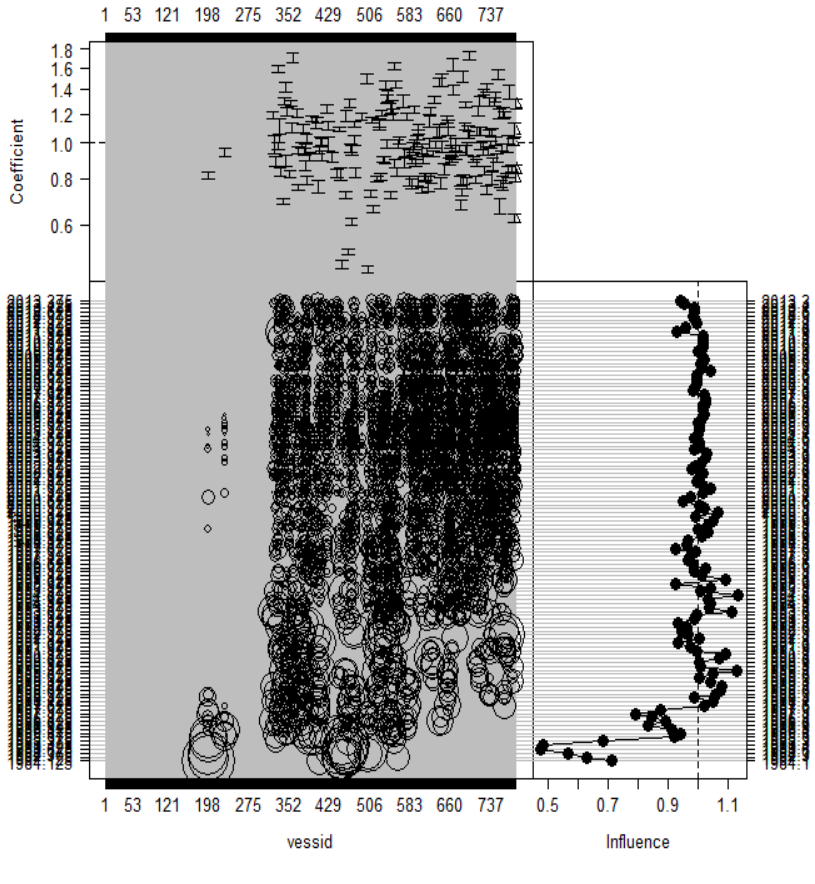
- Standardize CPUE with and without vessel id in the model, and compare trends
- This method lets us investigate changes in catchability associated with changes in the fleet
  - Represents effects of a) effort creep and b) changes in targeting.
  - i.e. the components of these associated with vessel turnover.

Vessel effects (bigeye models)

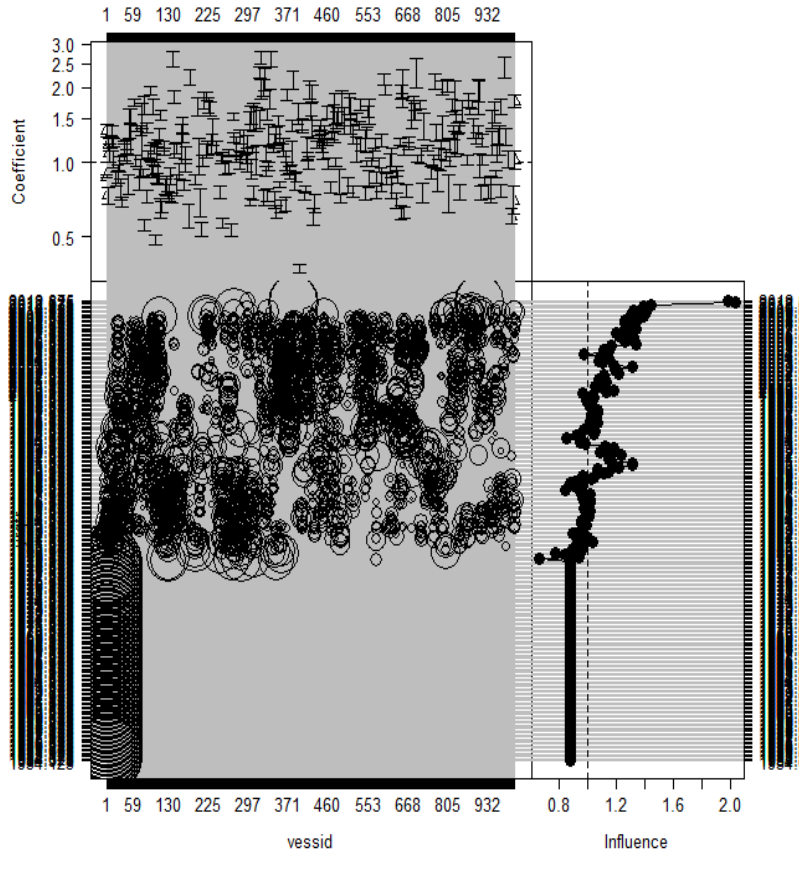
KR



TW

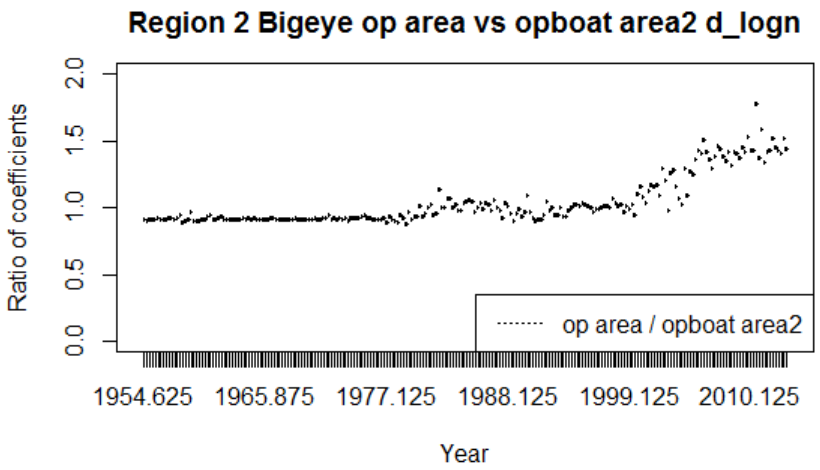


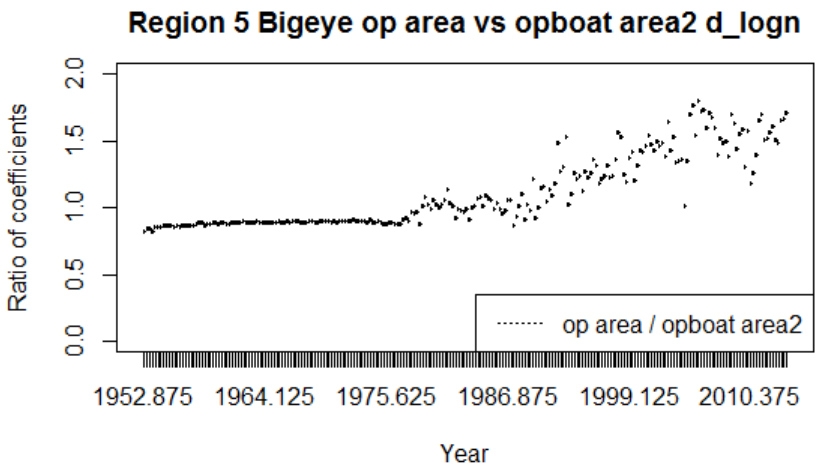
JP

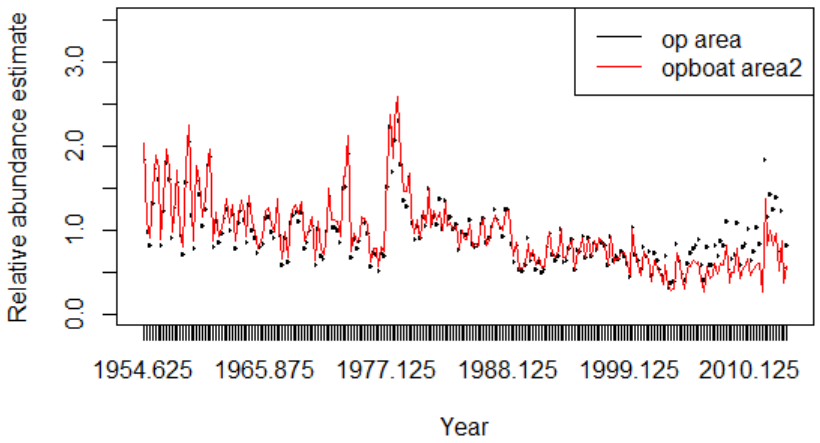


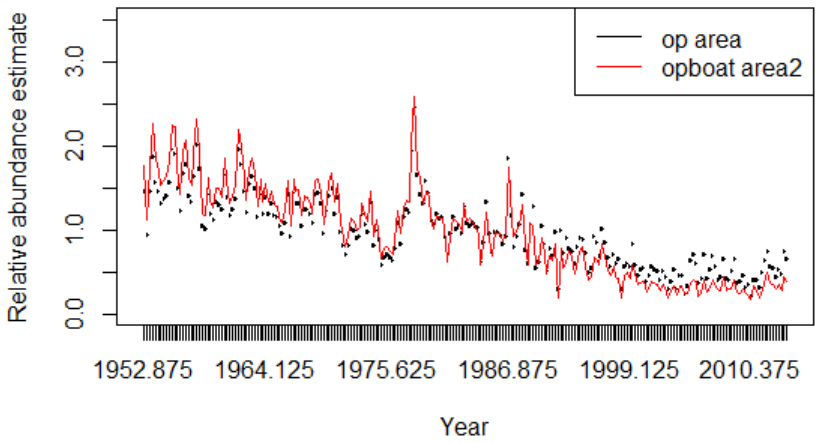
Tropical BET indices

- Each of the lower plots include the results of 2 similar analyses
- The red line includes vessel effects, the black dots have no vessel effects
- The upper plots show the ratio of the 2 approaches









Conclusion

- Cpue.rfmo
  - Allows me to run many analyses in a short time window
  - Allows national scientist colleagues to run the same analyses on their own datasets, using the same methods
  - Now want more contributors
- CPUE methods
  - Current methods are fairly robust
  - Room for improvement: both generic, and specific to individual analyses.