FISH 621
Estimation of Fish Abundance:
12: CPUE GAMs

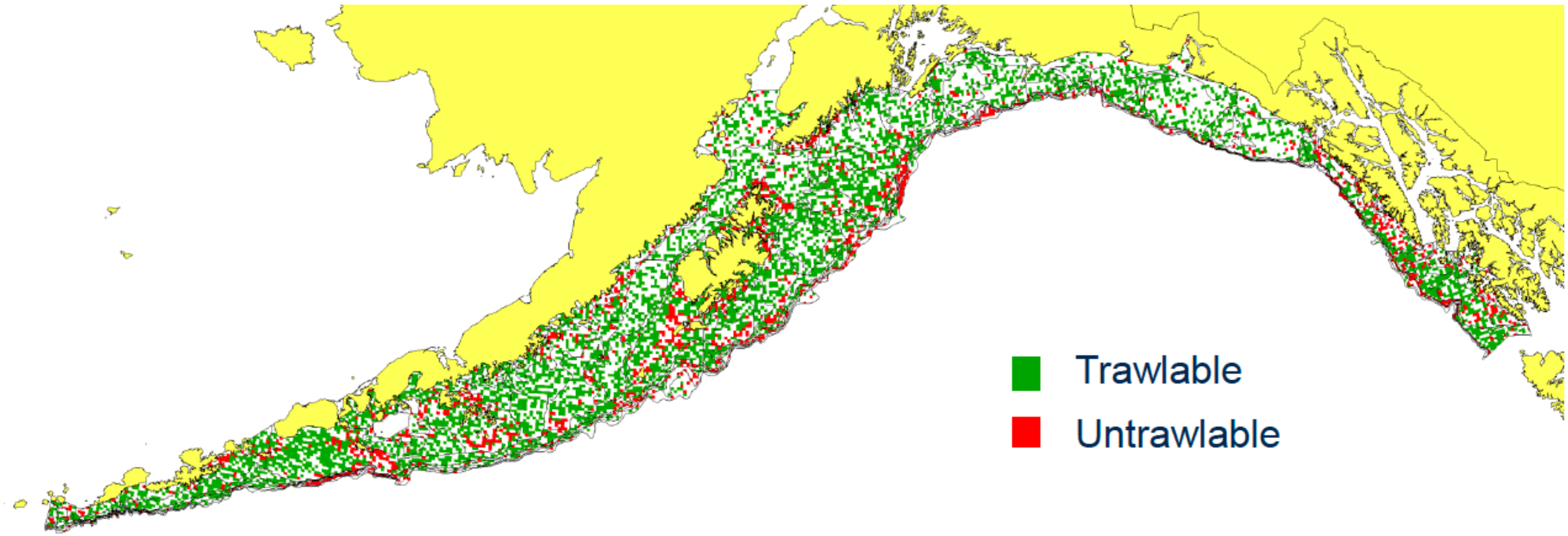
Dr. Curry Cunningham: cjcunningham@alaska.edu

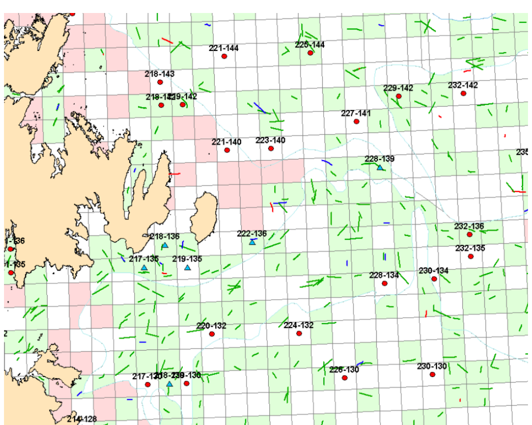
NMFS Bottom Trawl Survey Data: Stratified Random Design

- 59 strata
  - Based on geography, habitat, and depth
- 15 minute trawl
  - ~ 1.5km distance
- Estimate catch per unit effort (kg/km^2)
  - Collect length, age and other biological samples

Wayne Palsson
NOAA-NMFS

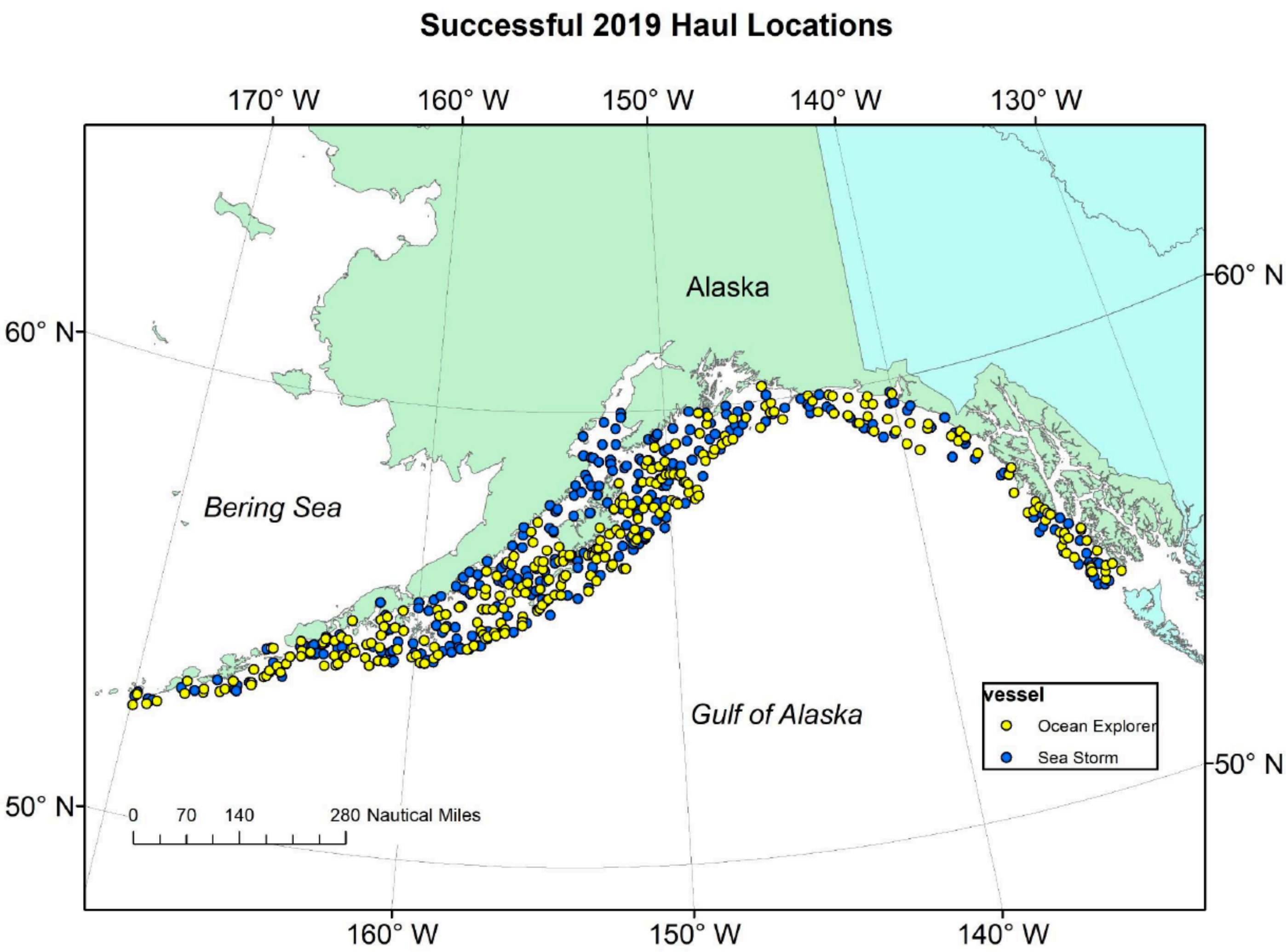
NMFS Bottom Trawl Survey Data: Stratified Random Design





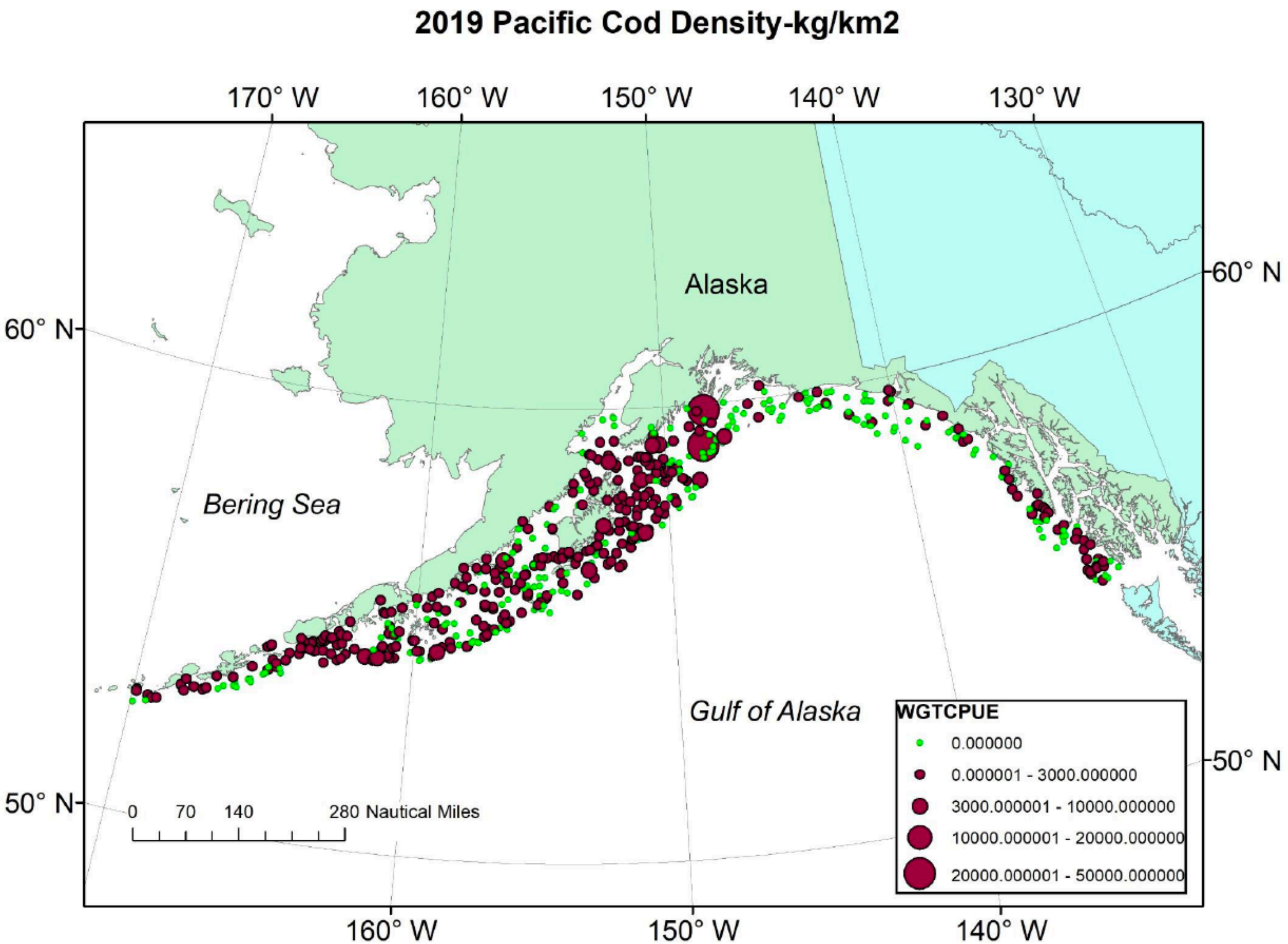
Wayne Palsson
NOAA-NMFS

NMFS Bottom Trawl Survey Data: Stratified Random Design



Wayne Palsson
NOAA-NMFS

NMFS Bottom Trawl Survey Data: Stratified Random Design



Wayne Palsson
NOAA-NMFS

Fishery-dependent CPUE

- Allen and Punsley (1984) - Tuna

CATCH RATES OF YELLOWFIN TUNA 325

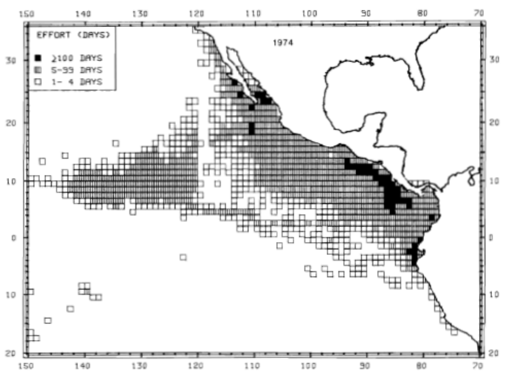


FIGURE 5. Distribution of fishing effort by seiners in the eastern Pacific Ocean in 1974 for all trips for which usable logbook data were obtained.

FIGURA 5. Distribución del esfuerzo de pesca por cerqueros en el Océano Pacífico oriental en 1974, de todos los viajes en los que se obtuvieron datos utilizables.

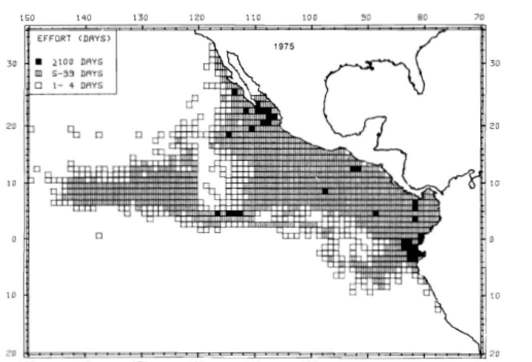


FIGURE 6. Distribution of fishing effort by seiners in the eastern Pacific Ocean in 1975 for all trips for which usable logbook data were obtained.

FIGURA 6. Distribución del esfuerzo de pesca por cerqueros en el Océano Pacífico oriental en 1975, de todos los viajes en los que se obtuvieron datos utilizables.

CATCH RATES OF YELLOWFIN TUNA 327

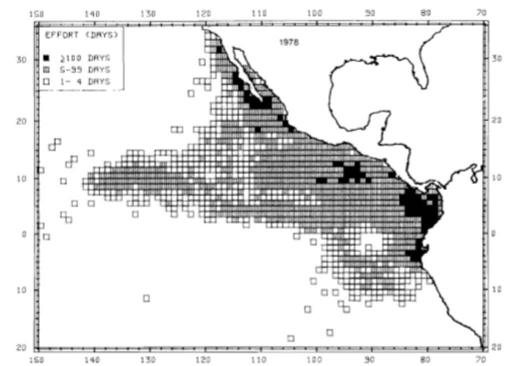


FIGURE 9. Distribution of fishing effort by seiners in the eastern Pacific Ocean in 1978 for all trips for which usable logbook data were obtained.

FIGURA 9. Distribución del esfuerzo de pesca por cerqueros en el Océano Pacífico oriental en 1978, de todos los viajes en los que se obtuvieron datos utilizables.

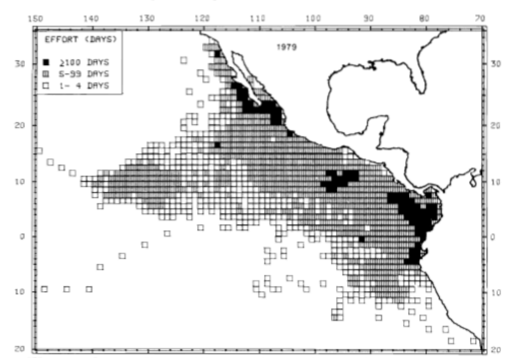
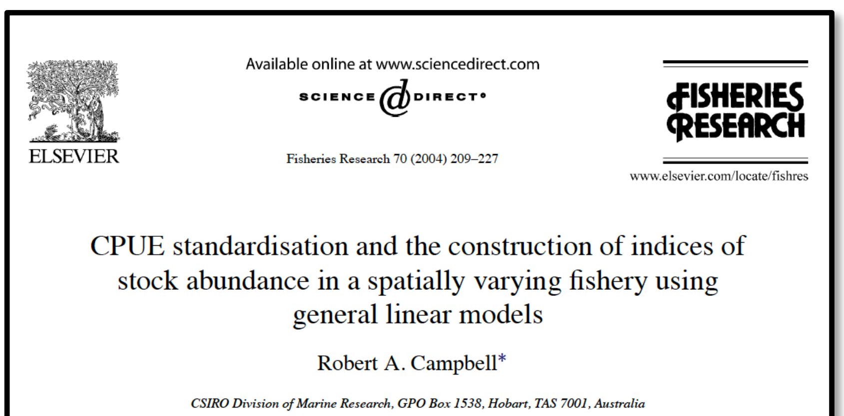


FIGURE 10. Distribution of fishing effort by seiners in the eastern Pacific Ocean in 1979 for all trips for which usable logbook data were obtained.

FIGURA 10. Distribución del esfuerzo de pesca por cerqueros en el Océano Pacífico oriental en 1979, de todos los viajes en los que se obtuvieron datos utilizables.

CPUE Standardization

- In order to discern an informative index of abundance across time
  - Necessary to control for factors that may influence sampling efficiency or fishing power
    - Along with spatial considerations
- The objective is to remove the influence of potentially confounding factors
  - To more accurately describe trends across time
- Generalized Linear Models and GAMs
  - Have a long history of practice in this context
    - Allen and Punsly (1984) and Campbell (2004)



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT

Fisheries Research 70 (2004) 209–227

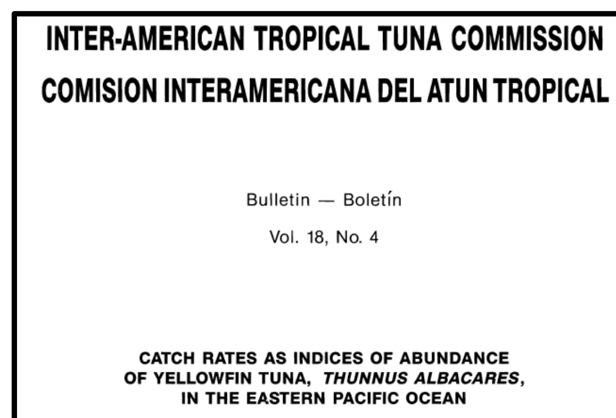
FISHERIES RESEARCH

www.elsevier.com/locate/fishres

CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models

Robert A. Campbell*

CSIRO Division of Marine Research, GPO Box 1538, Hobart, TAS 7001, Australia



INTER-AMERICAN TROPICAL TUNA COMMISSION

COMISION INTERAMERICANA DEL ATUN TROPICAL

Bulletin — Boletín

Vol. 18, No. 4

CATCH RATES AS INDICES OF ABUNDANCE OF YELLOWFIN TUNA, *THUNNUS ALBACARES*, IN THE EASTERN PACIFIC OCEAN

CPUE Standardization

- In order to discern an informative index of abundance across time
  - Necessary to control for factors that may influence sampling efficiency or fishing power
    - Along with spatial considerations
- The objective is to remove the influence of potentially confounding factors
  - To more accurately describe trends across time
- Generalized Linear Models and GAMs
  - Have a long history of practice in this context
    - Allen and Punsly (1984) and Campbell (2004)

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT

Fisheries Research 70 (2004) 209–227

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models

Robert A. Campbell*

CSIRO Division of Marine Research, GPO Box 1538, Hobart, TAS 7001, Australia

INTER-AMERICAN TROPICAL TUNA COMMISSION

COMISION INTERAMERICANA DEL ATUN TROPICAL

Bulletin — Boletín

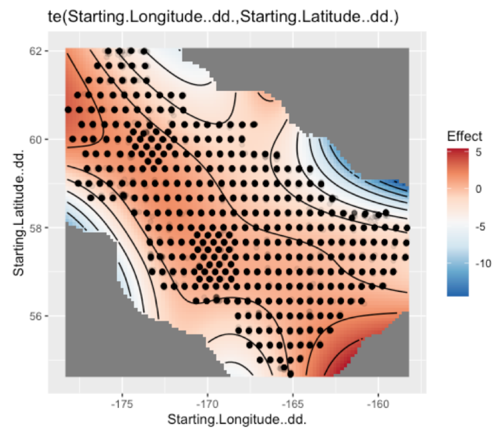
Vol. 18, No. 4

CATCH RATES AS INDICES OF ABUNDANCE OF YELLOWFIN TUNA, *THUNNUS ALBACARES*, IN THE EASTERN PACIFIC OCEAN

CPUE Standardization

- Errors are assumed lognormally distributed
  - Often by modelling the natural log of CPUE as the response
  - $E\left[log\left(CPUE_{i,k}\right)\right] = \mu_0 + year_i + region_k + \varepsilon rror_{i,k}$
    - $\varepsilon rror_{i,k} \sim Normal(0, \sigma^2)$ i.e normally distributed in log space
  - In R
    - ```
      glm( lnCPUE ~ factor(year) +
                 factor(region),
                 family=gaussian(link="identity) )
      ```
- In order to back calculate expected CPUE in normal space
  - "lognormal correction" must be applied
    - $E\left[CPUE_{i,k}\right] = \exp\left(\mu_0 + year_i + region_k + \frac{1}{2}\sigma^2\right)$
    - $E\left[CPUE_{i,k}\right] = \exp(\mu_0 + year_i + region_k) \exp\left(\frac{1}{2}\sigma^2\right)$
      ```

CPUE Standardization

- $E[CPUE_{i,k}] = \exp\left(\mu_0 + year_i + region_k + \frac{1}{2}\sigma^2\right)$
- Produces a ***relative*** index of abundance for the reference region
  - An ***absolute*** index by region can be found my multiplying by the area of each region (A_k)
    - $B_{i,k} = A_k E[CPUE_{i,k}]$
- The total absolute index of abundance for year i
  - Is found by summing across regions k
    - $I_i = \sum_{k=1}^{NR} B_{i,k}$

CPUE Standardization

- For fishery-dependent data seasons should be considered, if you suspect:
  - Catch rates vary among seasons
  - There is heterogeneity in fishing effort among seasons
    - Across years
- Luckily we can easily add season to our GLM or GAM
  - $E\left[log\left(CPUE_{i,k,j}\right)\right] = \mu_0 + year_i + region_k + season_j + \varepsilon rror_{i,k,j}$
  - In R
    - ```
      glm( lnCPUE ~ factor(year) +
                factor(region) + factor(season),
                      family=gaussian(link="identity))
      ```
    - ```
      gam( lnCPUE ~ factor(year) +
                factor(region) + s(day of year),
                      family=gaussian(link="identity))
      ```

$$E[CPUE_{i,k,j}] = \exp\left(\mu_0 + \cdots + \frac{1}{2}\sigma^2\right)$$

CPUE Standardization

- If controlling for season and area
  - The annual index of abundance found by
    - Summing area-weighted regional indices
    - Averaging across seasons (here assuming they are quarters)
  - Arithmetic mean
    - $I_i = \frac{1}{NS}\sum_{j=1}^{NS}\left[\sum_{k=1}^{NR} A_k E\left(CPUE_{i,k,j}\right)\right]$
  - Geometric mean (scale invariant)
    - $I_i = \sqrt[NS]{\prod_{j=1}^{NS}\left[\sum_{k=1}^{NR} A_k E\left(CPUE_{i,k,j}\right)\right]}$
- Finally, to calculate a relative index which
  - Relates the average abundance to a reference year i
    - $I_{i,ref} = \frac{I_i}{I_{ref}}$

GAMs for CPUE Analysis

- GAMs (mgcv or equivalent)
  - Offer a useful way to approach index standardization that allow for inclusion of smoothed effects
  - Benefits
    - Control for non factor-type or linear covariates (depth, space, time)
    - Estimate spatial fields (distribution)
    - Estimate space-time fields (changes in distribution across time)
- If depth is recorded by fishing event
  - And we believe catch rates vary as a function of depth
    - ```
      gam( lnCPUE ~ factor(year) +
                factor(region) + s(depth),
                family=gaussian(link="identity))
      ```
      ```

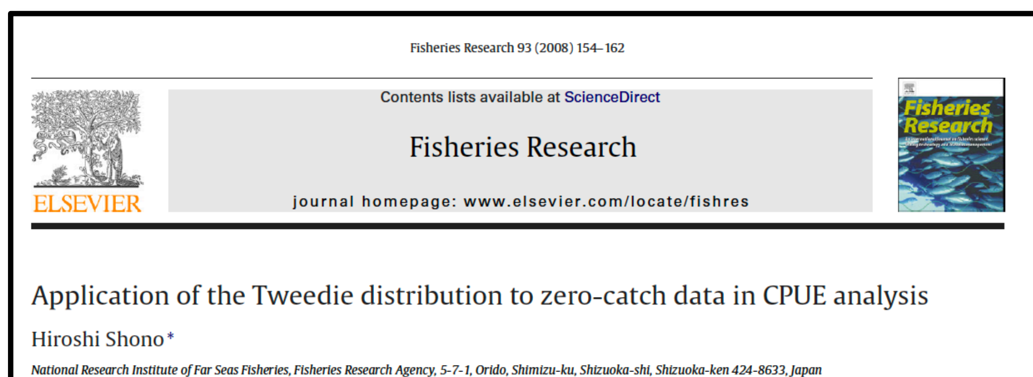
GAMs for CPUE Analysis



- Spatial model
  - Estimates average spatial CPUE field (distribution)
    - Useful for EFH designation!
  - Can account for spatial heterogeneity in fishing effort
    - ```
      gam( lnCPUE ~ factor(year) +
                + te(lat,lon),
              family=gaussian(link="identity"))
      ```
- Spatio-temporal model
  - Estimates average spatial field (distribution)
    - And deviations from the avg. field (changes in distribution)
  - Can account for spatial heterogeneity in fishing effort
    - ```
      gam( lnCPUE ~ factor(year) +
                + te(lat,lon) +
                    te(lat,lon, by=factor(year),

              family=gaussian(link="identity"))
      ```

Could be month, season, or other time reference

CPUE: Offsets for Effort

- If we are dealing with units of abundance
  - In the form of count data (55, 101, 10, 0, 1, …)
    - And we want to use a more appropriate error distribution (Poisson, negative binomial, ect.)
- We can treat log(effort) as an offset
  - But we need to update our link function
    - ```
      gam( Count ~ factor(year) +
                + te(lat,lon) + offset(log(effort)),

              family=Poisson(link="log"))
      ```
      ```

Dealing with Zeros

- CPUE data often have an excess of zeros
  - Which are problematic for many of the standard error distributions we
    - Assume within an index standardization model
- These represent no fish of a given species being caught in a given catch event
  - or sampled from the catch in the case of observer subsampling of retained catch
- Several methods are commonly employed to deal with zeros
  - See Shono et al. (2008)


Fisheries Research 93 (2008) 154–162

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Application of the Tweedie distribution to zero-catch data in CPUE analysis

Hiroshi Shono*

National Research Institute of Far Seas Fisheries, Fisheries Research Agency, 5-7-1, Orido, Shimizu-ku, Shizuoka-shi, Shizuoka-ken 424-8633, Japan

Dealing with Zeros

- Method #1   - Ad hoc
  - Add a small value to CPUE
  - $E\left[log\left(CPUE_{i,k} + \delta\right)\right] = \mu_0 + year_i + region_k + \varepsilon rror_{i,k}$
    - $\varepsilon rror_{i,k} \sim Normal(0, \sigma^2)$ i.e normally distributed in log space
  - In R
    - ```
      glm( log(CPUE+1) ~ factor(year) +
                factor(region),
                family=gaussian(link="identity) )
      ```
- Method #2 – Alternative error distributions
  - Tweedie distribution, zero-inflated Poisson, zero-inflated negative binomial
    - Allow for an excess of zeros, relative to base model assumptions
- Method #3 – Delta or hurdle model
  - Separately model encounter probability
    - And positive catch rate
  - $P(Abundance) = P(Encounter)P(Abundance|Encounter)$
      ```

Dealing with Zeros

- Method #3 – Delta or hurdle model
  - Add a binary 1/0 response variable column (`enc`)
  - Model encounter/non-encounter with a logistic regression
    - ```
      ep.mod <- glm(enc ~ factor(year) +
                          factor(region),
            family=binomial(link="logit")
      ```
  - Subset data for encounters only
    - data[data$enc==1,]
  - Model positive catches (enounters) only using appropriate error distribution
    - ```
      pc.mod <- glm(lnCPUE ~ factor(year) +
                             factor(region),
                    family=gaussian(link="identity"),
                    data=data[data$enc==1,])
      ```