Issues highlighted by applications

Toshihide KITAKADO

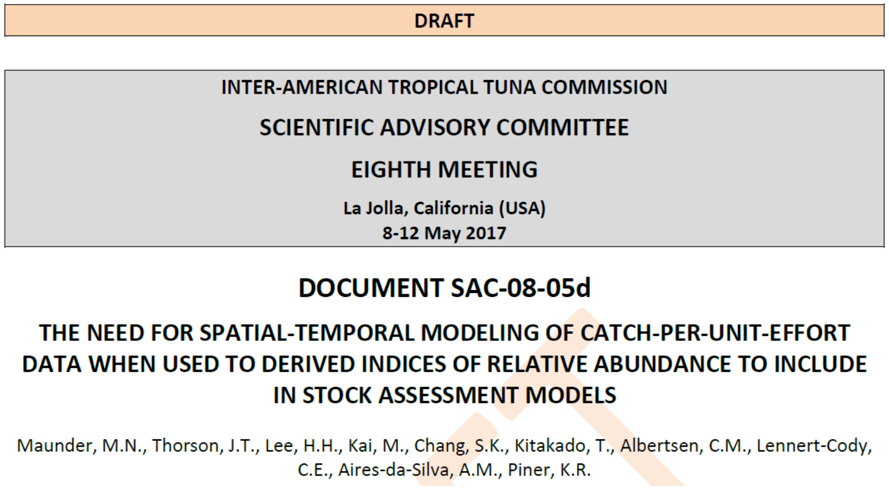
Tokyo University of Marine Science and Technology

Development of spatio-temporal models of fishery catch-per-unit-effort data to derive indices of relative abundance

Feb 27, 2018@CAPAM

Apology…

- This time, I was intending to show some lessons learnt from my on-going analysis on “Spatial and temporal modelling for CPUE STD for the Atlantic bigeye”



DRAFT

INTER-AMERICAN TROPICAL TUNA COMMISSION

SCIENTIFIC ADVISORY COMMITTEE

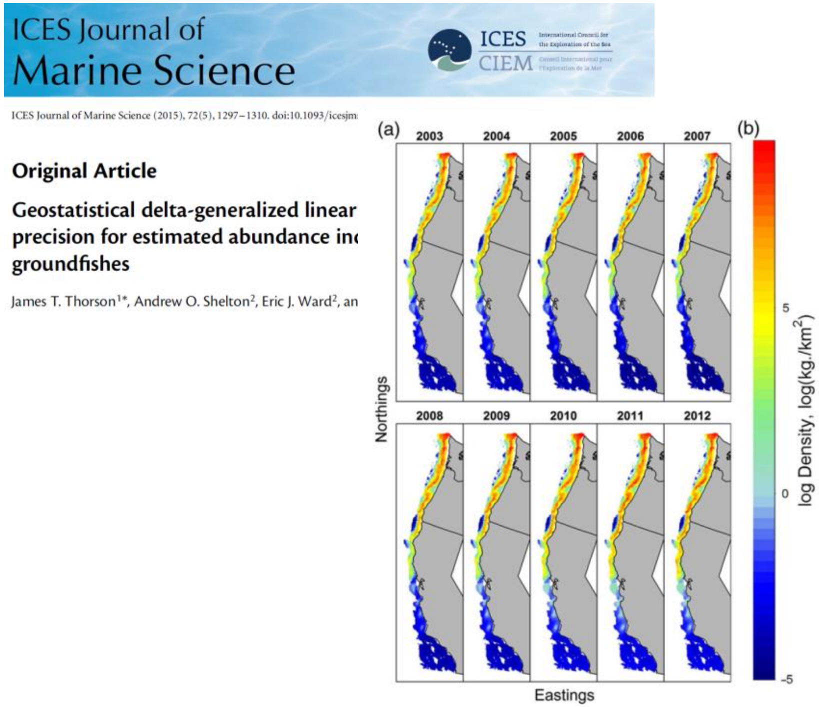
EIGHTH MEETING

La Jolla, California (USA)
8-12 May 2017

DOCUMENT SAC-08-05d

THE NEED FOR SPATIAL-TEMPORAL MODELING OF CATCH-PER-UNIT-EFFORT DATA WHEN USED TO DERIVED INDICES OF RELATIVE ABUNDANCE TO INCLUDE IN STOCK ASSESSMENT MODELS

Maunder, M.N., Thorson, J.T., Lee, H.H., Kai, M., Chang, S.K., Kitakado, T., Albertsen, C.M., Lennert-Cody, C.E., Aires-da-Silva, A.M., Piner, K.R.



ICES Journal of Marine Science

ICES Journal of Marine Science (2015), 72(5), 1297–1310. doi:10.1093/icesjm

Original Article

Geostatistical delta-generalized linear precision for estimated abundance in groundfishes

James T. Thorson1*, Andrew O. Shelton2, Eric J. Ward2, an

- But it has not been finished yet unfortunately, so …

Contents of my presentation

- A simple example motivating me to address spatial and temporal aspects (not as examples of extracting abundance indices)

- Some issues including spatial and temporal modelling

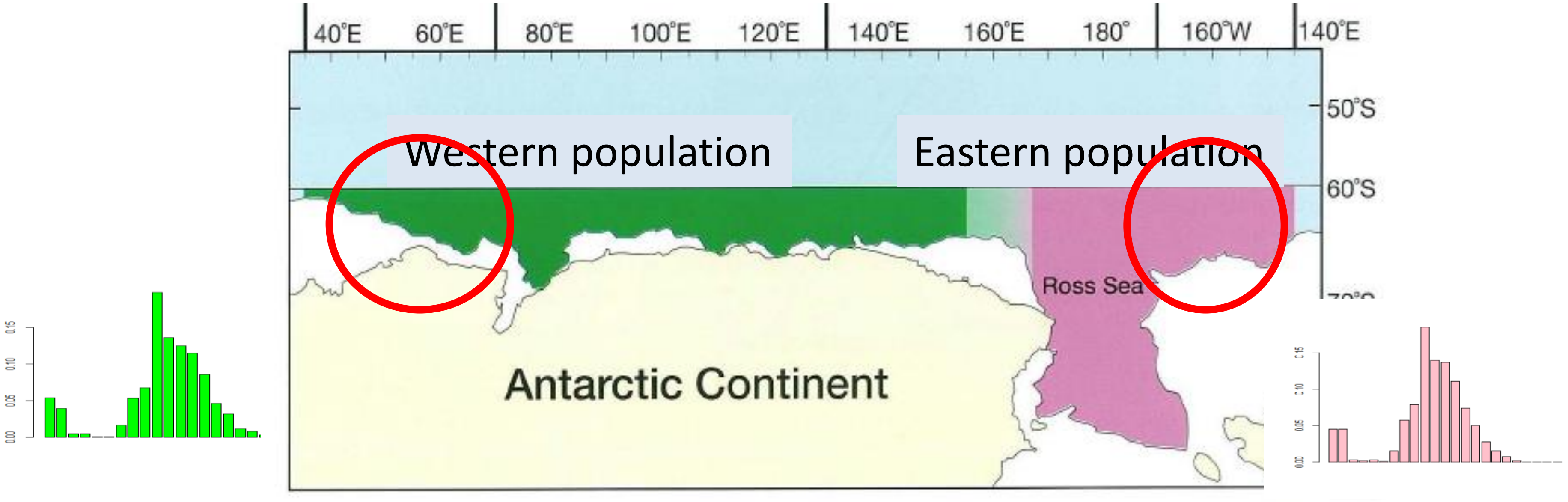
A “simple” example motivating me to address spatial and temporal aspects in general

(not as examples of extracting abundance indices)

Just my another on-going and unfinished works for spatial and temporal mixture modelling

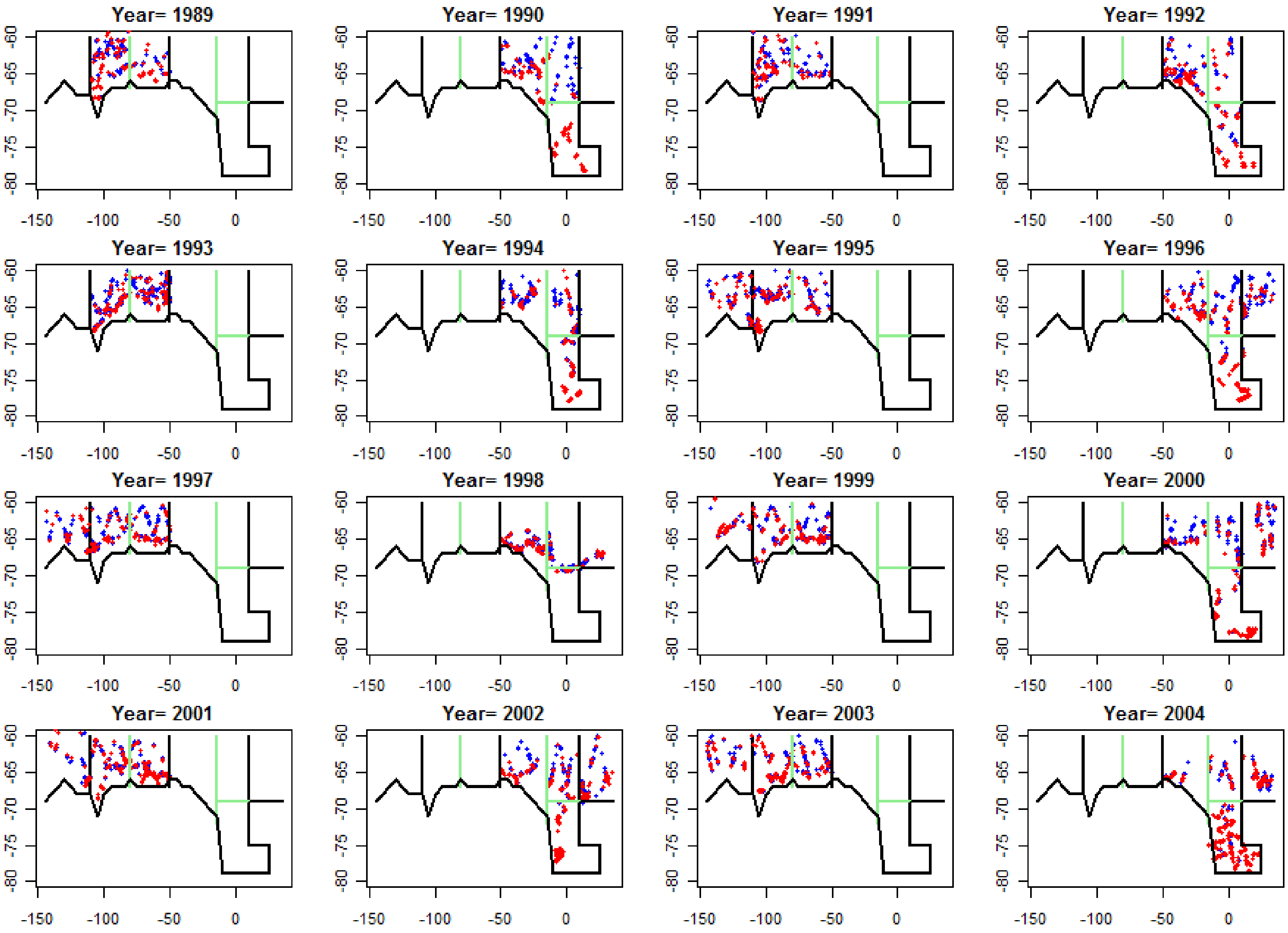
1) Two populations migrating to the Antarctic
2) There might be a mixing zone in between


© National Museum of Nature and Science クロミンククジラ *Balaenoptera bonaerensis*



1. Develop a statistical model for estimating spatial population mixture by using not only genetic (mt DNA, microsatellite) data but also morphometric data (integration of the two sources of information)
2. Investigate how the spatial population mixture changes year by year and by sex for the Antarctic minke whales

Sampling locations (1989/90 - 2004/05)

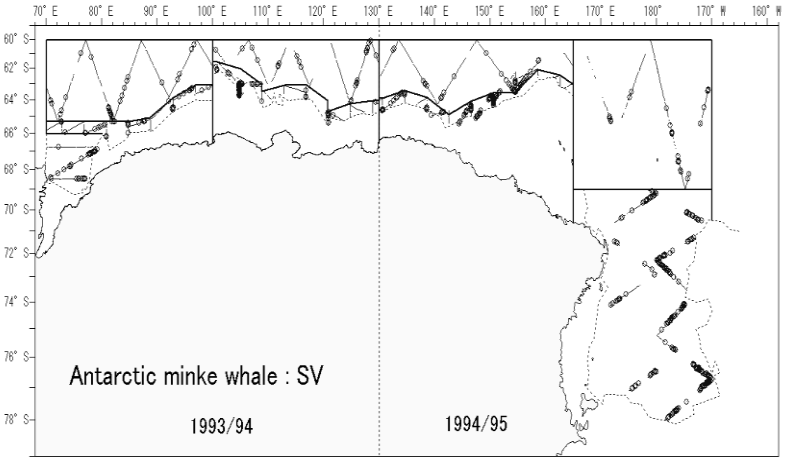


blue: male
red: female

Data employed in this study

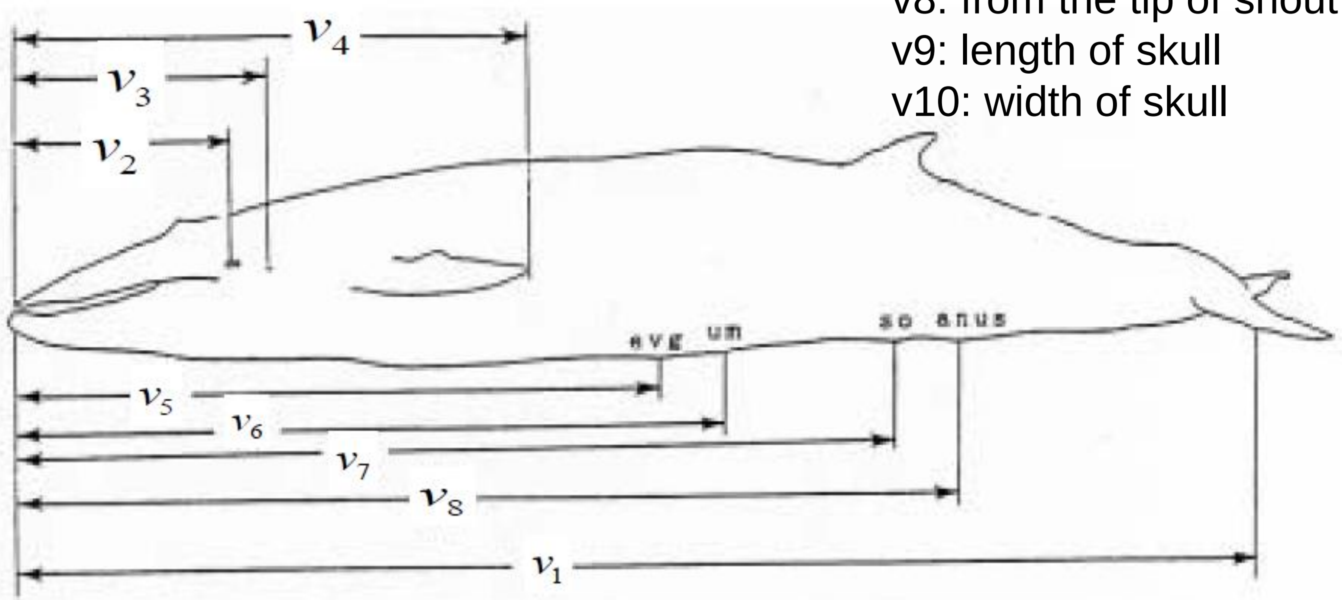
JARPA survey: sample size is ca. 6,000 for 16 years

- Individual sex and maturity information
- Individual genotypes at several microsatellite loci
- Individual haplotypes for mtDNA
- Individual 10 dimensional morphometric data



Morphometric measurements

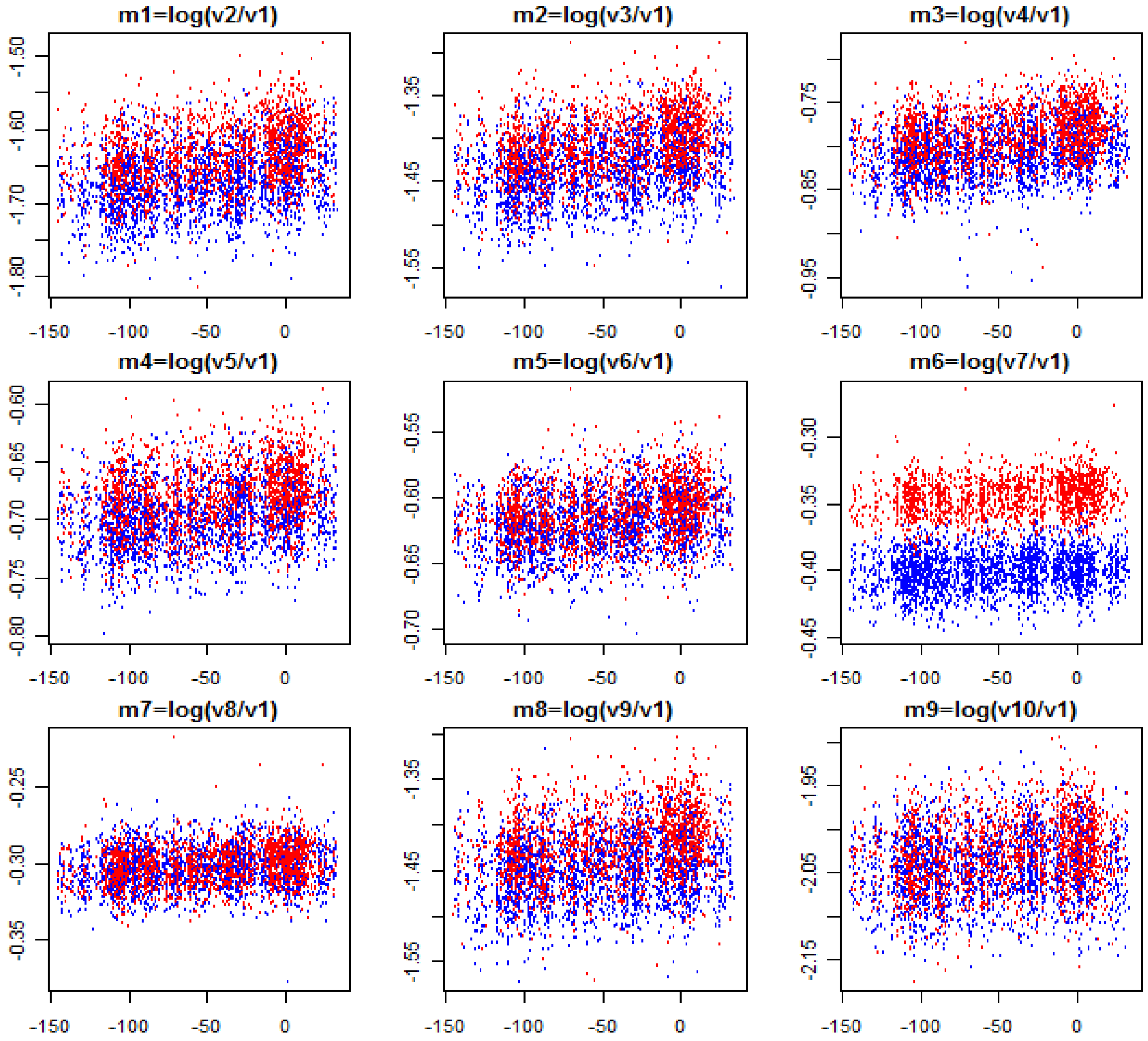
v1: Body length
v2: from the tip of snout to center of eye
v3: from the tip of snout to ear
v4: from the tip of snout to tip of flipper
v5: from the tip of snout to end of ventral gloves
v6: from the tip of snout to center of umbilicus
v7: from the tip of snout to sexual apparatus
v8: from the tip of snout to anus
v9: length of skull
v10: width of skull



Allometric measurements

$$m_j = \log(v_{j+1} / v_1)$$

Morphometric data against the longitude where sample is taken



Allometric measurements

$$m_j = \log(v_{j+1} / v_1)$$

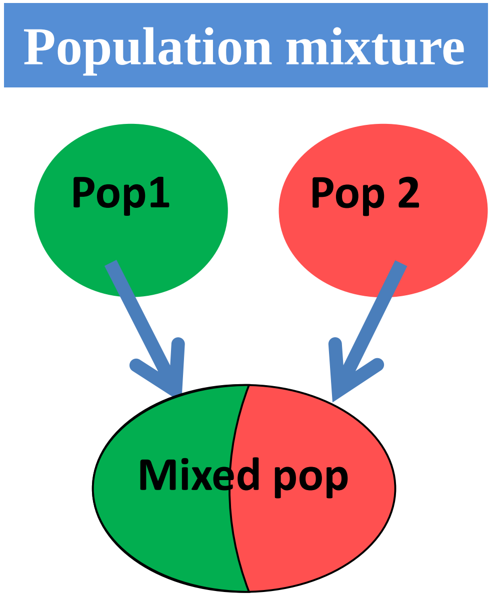
blue: male

red: female

This is a sign of difference between western and eastern populations

Statistical model for Mixture

Note: No baseline populations were assumed



Population mixture

Pop1

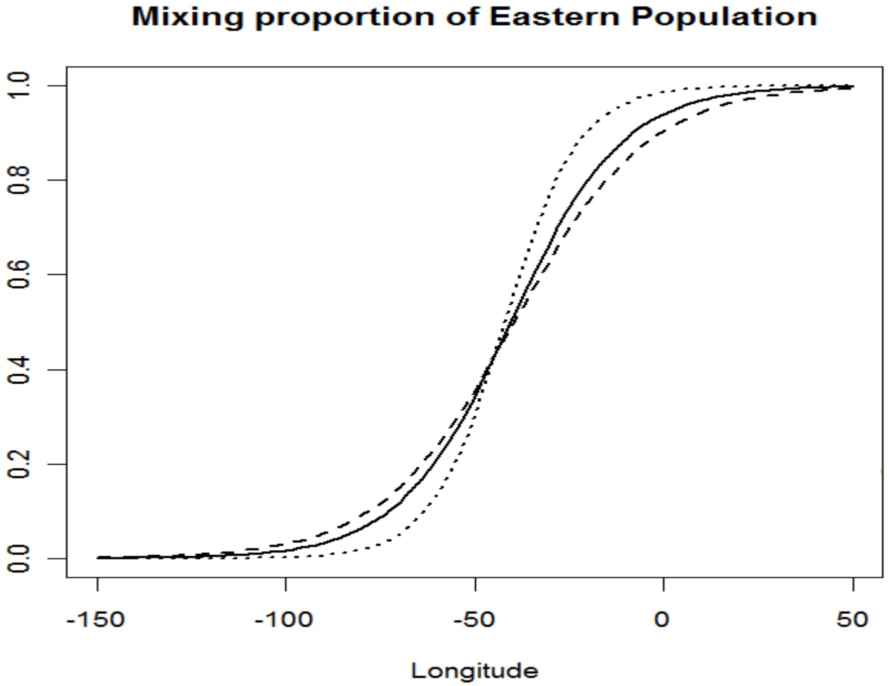
Pop 2

Mixed pop

Estimation of mixing proportion

Mixing proportion is assumed to be a function of longitude

$$P\left(I=E\right)=p=\frac{e^{\alpha_y+\beta x}}{1+e^{\alpha_y+\beta x}}.$$



Joint likelihood function

Joint likelihood function from the three different sources

$$L = \prod_{i=1}^{n} \left[(1 - p(x_i))L_i^W + p(x_i)L_i^E \right]$$

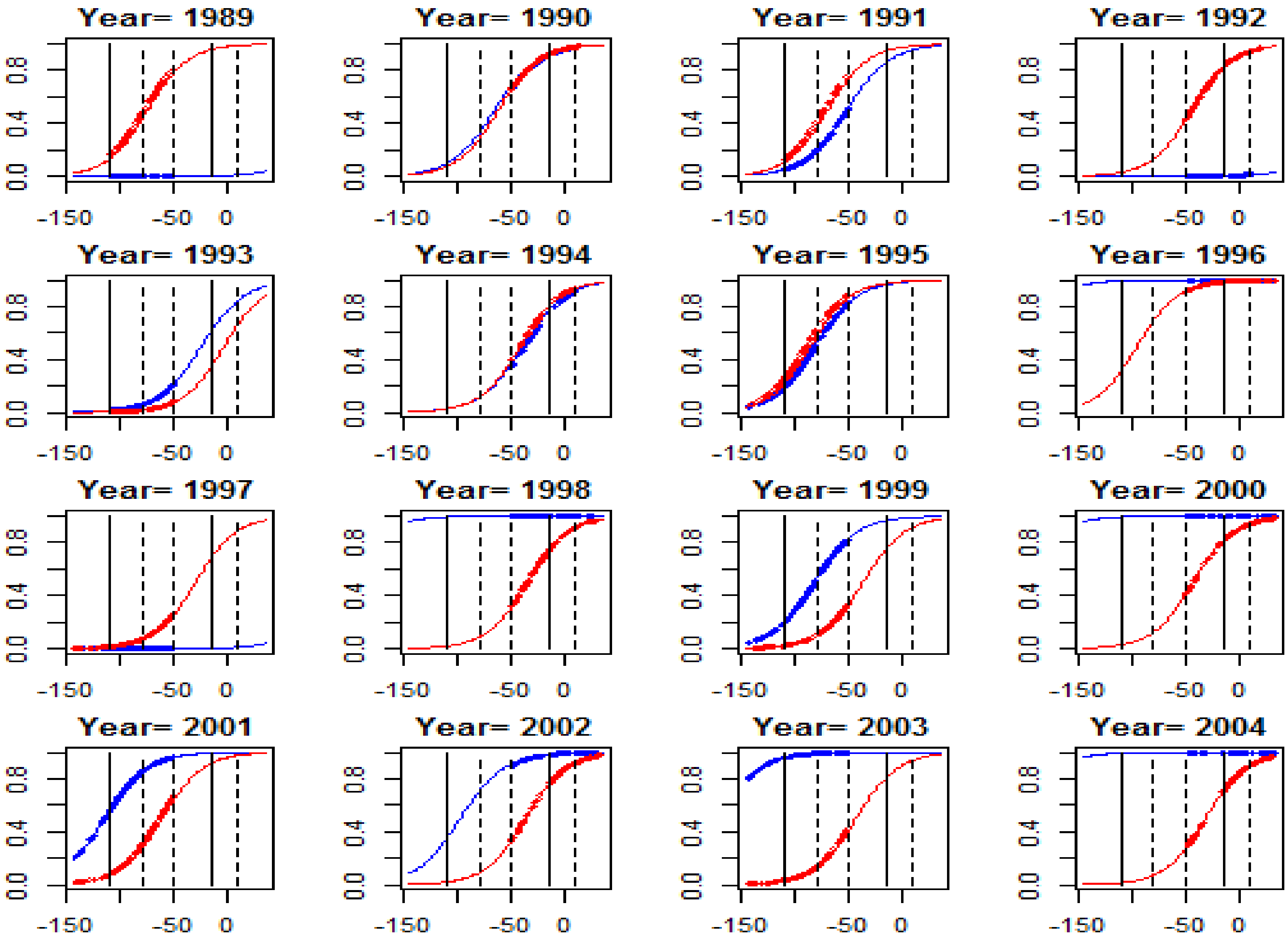
$$L_i^W = L_{micro}(\gamma^W)L_{mtDNA}(\delta^W)L_{morph}(\mu_{s(i)}^W, \Sigma_{s(i)})$$

$$L_i^E = L_{micro}(\gamma^E)L_{mtDNA}(\delta^E)L_{morph}(\mu_{s(i)}^E, \Sigma_{s(i)})$$

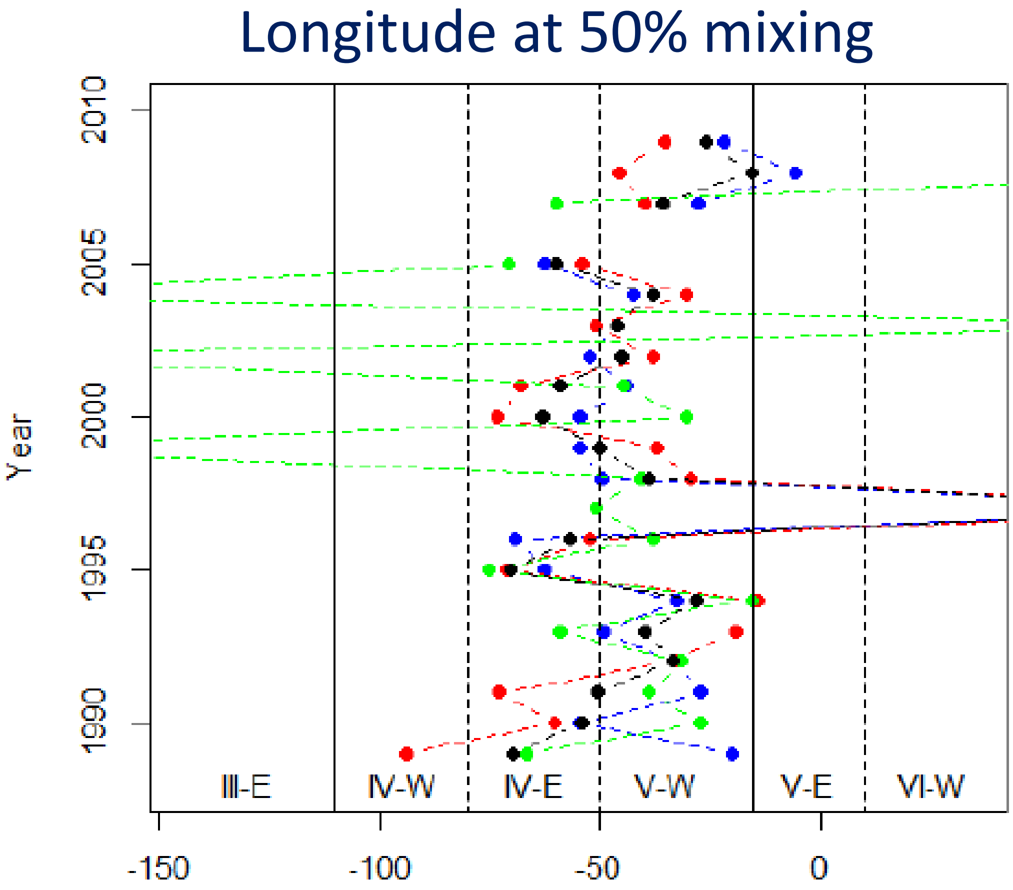
$$P\left(I_i = E\right) = p(x) = \frac{e^{\alpha_y + \beta x}}{1 + e^{\alpha_y + \beta x}}$$

Optimization was conducted by an optimization tool, ADMB

Result: sex-specific mixture



The most recent analysis (with additional recent data)



- Assume random effects for mixing proportions over years to borrow the strength from all the years
- More spatial and temporal smoothing aspects with TMB!

Some issues in CPUE standardization

Many issues in standardization of abundance indices

- Lots of “0” data (prob dist allowing 0; 2-stage distributions)
- Non-linearity
- Many possible interaction terms with “no” or few observation
- Different model selection criteria say differently
  （AIC, BIC, Cross-validation, …)
- Evaluation of uncertainty

- Extracting “Year” and “Year*Month” effects for stock assessment
- Area weighting
- Changes in spatial distribution over time
- Changes in selectivity patters over space and time
  (due to spatial and temporal changes in availability by age/size)
- Changes in target species
- Different or contradictory indices from different fisheries etc.

Objectives of CPUE STD

What do you want to do for standardizing CPUE, by the way?

- Mainly, to provide information on trend of relative exploitable biomass specific to fishery
- To provide information on patters of spatial habitat use in space and time etc.

Target quantity and observation

Target unknown quantity

$$\tilde{B}_y \propto \sum_a S_a \cdot W_a \cdot N_{y,a}$$

Observation

$$I_{y,i} = \sum_a q \cdot p(z_{y,i}) S_a \cdot W_a \cdot N_{y,a}$$

$$= q \cdot p(z_{y,i}) \sum_a S_a \cdot W_a \cdot N_{y,a}$$

$$= q \cdot p(z_{y,i}) \cdot \tilde{B}_y$$

Target of CPUE standardization

Changes in spatial distribution over time

Very basically…

CPUE <= Year + Month + SST + s(Lon, Lat) + ...

Typical habitat distribution

Not sure if migration is always same "timing"

CPUE <= Year + Month + Year*Month + SST + s(Lon, Lat) + ...

As a biological effect

Not sure if migration is always same "place"

CPUE <= Year + Month + Year*Month

+ SST + s(Lon, Lat) +s(Lon, Lat) *Year + ...

Spatial-temporal interactions
(not accounted by SST, Lon, Lat)

Smooth changes over space and time

Staptio-temporal model as “State-model”

$$d(s,t) = \exp\left[d(t) + \gamma(s) + \theta(s,t) + \sum \beta_j x_j(s,t) \right]$$

Temporal effect

Spatial effect

Spatio-temporal interaction

Other covariates in (s,t)

- Temporal effect: fixed or smoothness
- Spatial effect: fixed, smoothness, Gaussian random field, …
- Interaction terms: fixed, random, GRF with temporal drift given area, …

=> Many random quantities … (need computational tools)

How to handle spatial and temporal components

- Spatial-temporal interactions are regarded as stochastic latent effects with some structures
  - In space, smoothing "Random field"
  - In time, Auto-regressive structure
  (e.g. Thorson et al. 2015, Kai et al. 2017, …)

- Estimation of spatial distribution with GAM and 3D-GAM
  (e.g. Augustin et al. 2013)

- Machine learning (e.g. Regression type random forests)
  - Spatial-temporal interactions with many dummy variables
  => Intentionally regard as n << p problem
  => Apply regression machine learning
  (Lennet-Cody et al. 2010, 2013)

- And more!

Extension: Thorson et al. (2015)

- Presence and absence (with Spatio-temporal component)?

$$d(s,t) = \exp\left[d(t) + \gamma(s) + \theta(s,t) + \sum \beta_j x_j(s,t) \right]$$

$$p(s,t) = \exp\left[d_0(t) + \gamma_0(s) + \theta_0(s,t) + \sum \beta_{0j} x_j(s,t) \right]$$

Again, objectives of CPUE STD

- Distribution pattern is same over growth ?
- Availability is same over years and size?
  (difference in catchability among different size categories)

Target unknown quantity

$$\tilde{B}_y \propto \sum_a S_a \cdot W_a \cdot N_{y,a}$$

$$\tilde{B}_y \propto \sum_a S_{y,a} \cdot W_a \cdot N_{y,a}$$

Perhaps time-varying option

Observation

$$I_{y,i} = \sum_a q \cdot p_a(z_{y,i}) S_a \cdot W_a \cdot N_{y,a}$$

$$= q \cdot p'(z_{y,i}) \sum_a p_a(z_{y,i}) \cdot S_a \cdot W_a \cdot N_{y,a}$$

$$I_{y,a,i} = q \cdot p_a(z_{y,i}) \cdot S_a \cdot W_a \cdot N_{y,a}$$

Extension: by Age/Size

$$I_{y,i} = q \cdot p'(z_{y,i}) \sum_{a} p_a(z_{y,i}) \cdot S_a \cdot W_a \cdot N_{y,a}$$

➡

$$I_{y,a,i} = q \cdot p_a(z_{y,i}) \cdot S_a \cdot W_a \cdot N_{y,a}$$
$$= q(z_{y,i}) \cdot p_a(z_{y,i}) \cdot W_a \cdot N_{y,a}$$

Not necessarily
for all age/size categories

$$d(s,t,l) = \exp\left[d(t) + \gamma(s) + \tau(l) + \theta(s,t,l) + \sum \beta_j x_j(s,t,l) \right]$$

Resource selectivity
+ gear selectivity

Different migration
pattern by year and size

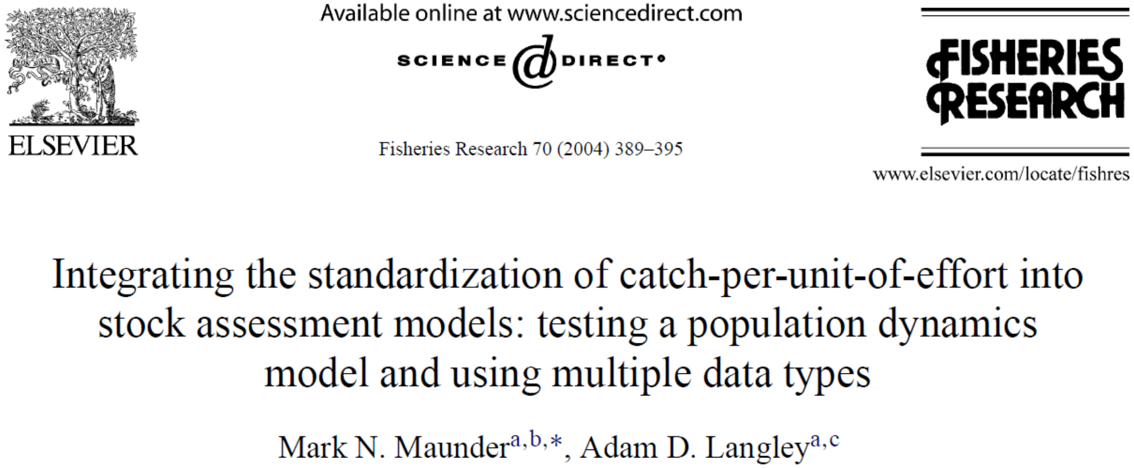
- Include size category information (for even some size classes)
- To account for difference in selectivity over different size categories (due to change in age/size composition over space and time)
- But, should not double use of length composition in the assessment
- Try to do joint estimation by CPUE and size for assessment?
  (not by “usual” two step approach)

Therefore…

Standardizing CPUE with more consideration of selectivity needed?

Normally : After standardizing CPUE without age/size effects,
and consider time-varying selectivity

Integration of CPUE STD and assessment?
=> Reincarnation of Maunder and Langley (2004)?



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Fisheries Research 70 (2004) 389–395

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types

Mark N. Maunder[a,b,*], Adam D. Langley[a,c]

Possible approaches for handling spatio-temporal model with size or age

- Integration of CPUE STD and assessment?
  - simultaneous estimation of all
  - avoid double use size composition in the assessment
  - have to estimate many parameters not directly relevant to stock status together

$$\tilde{B}_y \propto \sum_a S_{y,a} \cdot W_a \cdot N_{y,a} = \tilde{I}_y$$

$$L_{total}(\beta, S, N \mid CPUE, Size) = L(\beta, \tilde{I} \mid CPUE) L(S, N \mid \tilde{I}, Size)$$

$$L_{total}(\beta, S, N \mid CPUE\_noSize, CPUE\_withSize,\ Size)$$

$$= L(\beta, S, N \mid CPUE\_noSize)\, L(\beta, S, N \mid CPUE\_withSize,\ Size)$$

$$= L(\beta, \tilde{I} \mid CPUE\_noSize)\, L(\beta, S, N \mid \tilde{I}, CPUE\_withSize,\ Size)$$

$$L(\hat{\beta}, S, N \mid \tilde{I}, CPUE\_withSize,\ Size) \Rightarrow \max?$$

Time-varying getting smaller?

- Use and combine two different types of CPUE with/without age/size information

- Use CPUE with a part of size classes (as a covariate) to partially account for change in annual distribution by age/size ?

So, still many issues…

1. More size information for CPUE STD ?
2. Spatial/temporal difference in selectivity and target
3. Joint CPUE analysis ?
4. Way of extracting biomass index:
   area weighting, marginal mean
   (shift of efforts over time)

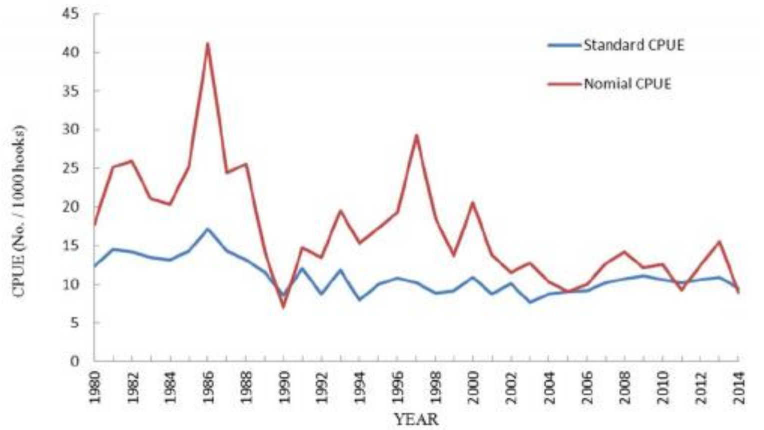
Issues in joint CPUE index

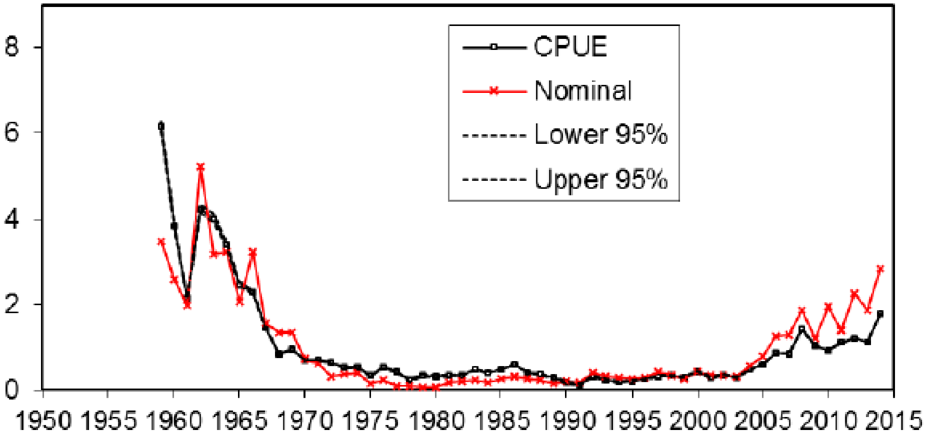
Joint CPUE index

- Common fishery (e.g. longline)
- But different patterns in standardized indices
- Need “trustable” indices for stock assessment

Possible reasons and solutions

- Different selectivity/target/…
- Lack of spatial coverage for some fisheries
- Try to produce “a joint index” by aggregating information available





Issues in joint CPUE index

Fishery 1

$$I_{y,f_1,i} = q_{f_1} \cdot p_{f_1}(x_{y,f_1,i}) \cdot r(z_{y,i}) \sum_a S_{f_1,a} \cdot W_a \cdot N_{y,a}$$

Same spatial and environmental effects (help each other to get information for less covered locations)

May not be same….

Fishery 2

$$I_{y,f_2,i} = q_{f_2} \cdot p_{f_2}(x_{y,f_2,i}) \cdot r(z_{y,i}) \sum_a S_{f_2,a} \cdot W_a \cdot N_{y,a}$$

If selectivity patters are similar, OK to combine to get a single abundance index

If not, perhaps again use age/size information (CPUE by some age/size classes?)

Any thoughts from the floor?

1. More size information for CPUE STD ?
2. Spatial/temporal difference in selectivity and target
3. Joint CPUE analysis ?
4. Way of extracting biomass index:
   area weighting, marginal mean
   (shift of efforts over time)