THE ATLANTIC-WIDE RESEARCH PROGRAMME FOR BLUEFIN TUNA (GBYP Phase 12)

SHORT TERM CONTRACT FOR THE ADVICE ON CLOSE-KIN MARK-RECAPTURE WORKSHOP (ICCAT GBYP 01/2023)

Final Report (Deliverable 2)

Madrid, 13 March 2023

Estimark Research, Australia:

Dr. Mark Bravington

This project is co-funded by the European Union

The material included here formed a set of presentations, given to stimulate discussion at an ICCAT/GBYP workshop on Close-Kin Mark-Recapture for Eastern Atlantic Bluefin Tuna (Madrid, March 2023). The report of the workshop itself may be found on the ICCAT website.

1 CKMR general

Slide 1.1

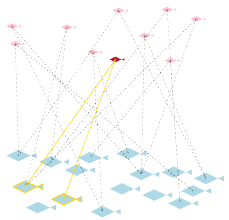
Contents of these slides

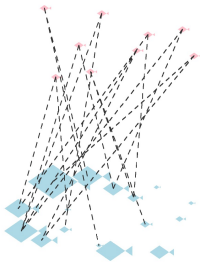
- Recap of CKMR
  - General ideas — not spatial
  - Basic requirements for typical fish
  - Examples of real CKMR
- Spatial CKMR for discrete sites
- EABT : some options

Slide 1.2

What is CKMR?

but what if...?



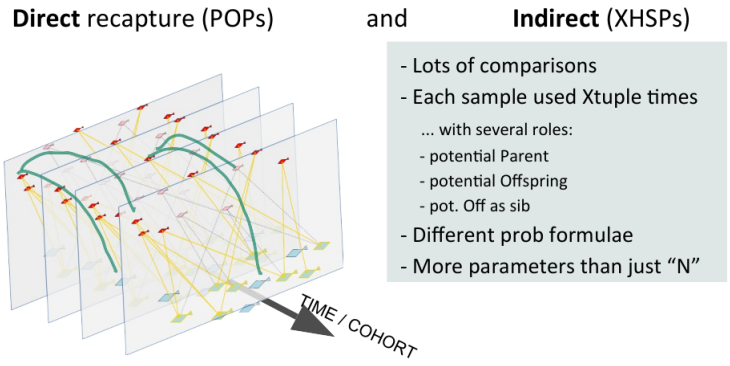


- A *data source* and *modelling framework* for stock assessment
- Uses only tissue samples from the catch
- Direct estimate of SSB and other demographic params
  - e.g. natural mortality; fecundity-at-age schedule
- Reveals *biological* info, as well as hardcore stock asst
  - so you don't need to know *everything* before you start

Slide 1.3

The basic idea

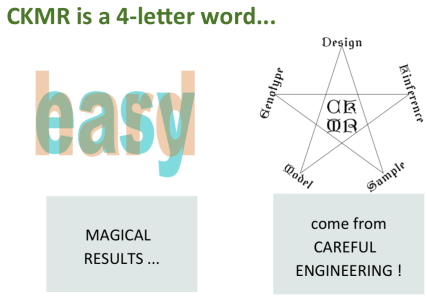
- Each offspring "tags" its mother and father genetically
- Can find "recaptured" tags using genetics
- Only needs samples from *dead* animals
  - plus measurements, eg date/size/sex/age/place



- "Response data" are genetic outcomes from *lots* of pairwise comparisons— almost all will be "Unrelated"

Slide 1.4

CKMR: general requirements



For *fish*: tissue samples from

- multiple cohorts of "juveniles" (potential offspring)
  - with pretty good age info
- full age/size range of adults (potential parents)
  - with *adequate* age/size info, and sex info
  - e.g. *epigenetic age* (just from tissue samples)
- *Enough* samples!!! and adequate composition (big/small/etc)

Slide 1.5

What does CKMR tell you, and how?

Fecundity-at-size (relative) because larger adults have measurably more detected offspring

Absolute adult abundance (really SSB time-series): $\mathbb{P}$[this adult ♀ is yr Mum] $\approx$ "1"/$N_{♀}$

Adult Z based on birth-gap between HSPs

and M too, if you know Catch

[Connectivity] in spatial case

- Each sample is compared to all (or most) others.
- Comparisons go back in time: for each adult you sample, you can look at its past history of spawning events.
- In practice, all parameters *must* be fitted inside an age-structured population-dynamics model, such as underpins most stock assessments. CKMR "rules" lead to formulae for $\mathbb{P}$[this pair will turn out to be kin|their covariates, demographic params]
- Parameters estimated by Max Likelihood / Bayes.
- Other "conventional" data can be added (but not essential).

Slide 1.6

(some) CKMR projects c. 2022

- Over 10 basically done (and mostly ongoing)
- SBTuna still the flagship
- CSIRO ones successful :) — though some could be better
- *Worst problems* from bad length/age measurements, and/or inadequate planning

Slide 1.7

Why has SBT been so successful?

- **Careful planning!!!**
- Complete control of sampling
  - Full size-range of adults
  - Known-age juveniles
- Excellent age and length data
- Spatially simple
  - at least for the samples we use
- Team with clear roles and full range of skills

Slide 1.8

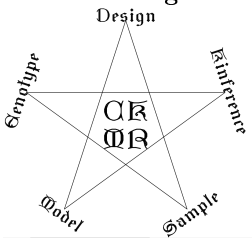
CKMR project lifecycle

1. **Design** to make sure sample sizes, composition, measurements will give adequately precise answers— *assuming* reality is as you expect!
2. Get enough samples to develop basic estimates and confirm/refute hypotheses (e.g. about connectivity)
   (a) Maybe realize that you need more samples (reality shock...)
3. *Maybe* e.g. 100 kin-pairs will give you adequate CV, and you have ground-truthed your assessment as a "one-off": and you might just **stop**.
4. Or, you can **redesign** and continue sampling: to update time-series for e.g. MP use; to improve precision (e.g. of $\hat{M}$); to refine hypothesis checks

The main difficulty with CKMR is getting to part 2. After the machinery is in place, it's quite cheap and simple.

Slide 1.9

CKMR design



If You...
- do not collect *enough* samples, or
- do not collect the right *types* of sample, or
- do not measure their stuff adequately

... **then** CKMR will **fail**

- Bottom line: *cannot* be precise with <50 POPs (and HSPs)
  - and may need substantially more than that
- Can't control *actual* number of kin, but can control *expected* number by sampling design and stratification
- In complex settings with many parameters and possibly inaccurate data, we need to do more:
  - build a "practice model" and explore variance of estimates (as a function of sample size, composition, ...)
  - There are mathematical tricks to help
- Also perfect for *learning* about each application!

Slide 1.10

CKMR vs individual (genetic) MR

Ignoring "conventional tags" here for well-known reasons

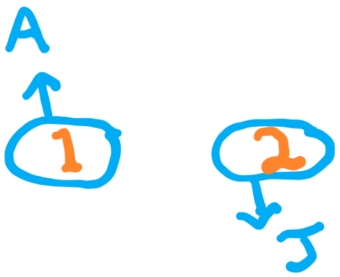
	CK	Indiv
Live-release?	N	Y
Abundance of..?	Adults	Tagged cohort(s)
Z / M ?	Y	(Y)
Fec@size	Y	N
Spatiality	"Lifetime"	Post-tagging
Tags per sample	2+	1
Genotyping	cheap	cheaper

"Unmodelled heterogeneity of capture probability":
could cause trouble for either (in different circumstances)
"Spatial" is a rich source of such problems...

2 Spatial CKMR

Slide 2.1

Spatial CKMR
Spatially-segregated "populations" need care with CKMR...
EG two separate stocks, adults & juves



- You *can* build valid spatial CKMR models
- and you *may* be able to sample so that such a model can be fitted
- but "any old sampling" probably *not OK*

Slide 2.2

Spatial CKMR: some terms

Assignable eg W/E ABT: can tell from genetics

Heritable not assignable, but *breeding* site (or feeding site...) "acquired" from parents

Faithful you pick a breeding site at maturity, and stick to it

Sticky you are mostly faithful, but might change

Random breed in different place every time

[Ontogenetic site changes with age]

Only the first is *necessarily* genetic. But, "heritable" matters a lot to management even if no genetics; and "faithful" matters a lot to CKMR, even if not (much) to management.

———

For a *sampling* site:

Well-mixed per capita prob of use is equal across "stocks"

Pure used only by one "stock"

Impure/partly-mixed stock-specific non-zero per capita usage prob

Slide 2.3

Spatial CKMR: the idea

Spatial CKMR is easy!

... from a lofty-enough viewpoint...

Non-spatial version:

$$\mathbb{P}\left[\text{Amy is Julian's mum} | \text{stuff about A \& J}\right] = \mathbb{E}\left[\frac{\text{\#A's J-like offspring (when J born)}}{\text{Total \# J-like offs (when J born)}} | \text{covariates of A,J}\right]$$

Slide 2.4

Spatial CKMR: the idea

Spatial CKMR is easy!

... from a lofty-enough viewpoint...

Non-spatial version:

$$\mathbb{P}\left[\text{Amy is Julian's mum} | \text{stuff about A \& J}\right] = \mathbb{E}\left[\frac{\text{\#A's J-like offspring (when J born)}}{\text{Total \# J-like offs (when J born)}} | \text{covariates of A,J}\right]$$

Spatial version:

$$\mathbb{P}\left[\text{Amy is Julian's mum} | \text{stuff about A \& J}\right] = \mathbb{E}\left[\frac{\text{\#A's J-like offspring (when \& where J born)}}{\text{Total \# J-like offs (when \& where J born)}} | \text{covariates of A,J}\right]$$

Slide 2.5

Spatial CKMR: POPs and HSPs

- HSPs as well as POPs are very important in CKMR
  - and will inform e.g. on site-faithfulness (fidelity)
- But, main abundance and connectivity signals will come from POPs
  - and POPs are complicated enough...
- So, here I'm just concentrating on "what we would learn/need from POPs"

Slide 2.6

Spatial CKMR: general formula

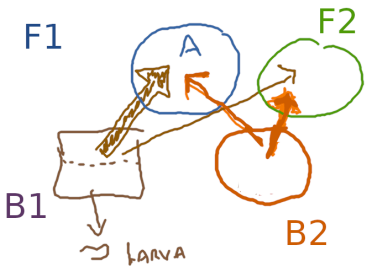
Animals A(dult) and J(uvenile) sampled at places p_A, p_J, at ages a_A, a_J, in years y_A, y_J. Let x be the (possibly unknown) birthplace of J. Then:

$$\begin{aligned}
&\mathbb{P}\left[K_{AJ} = \text{POP} | pay_A pay_J ...\right] \\
&= \sum_{\text{possible } x} \{\mathbb{P}\left[K_{AJ} = \text{POP} | x; ay_A ay_J\right] \times \\
&\quad \mathbb{P}\left[A \text{ at } x \text{ when } J \text{ born} | pay_A\right] \times \\
&\quad \mathbb{P}\left[J \text{ born at } x | pay_J\right]\}
\end{aligned}$$

First term: standard CKMR formula
2nd & 3rd: must be in right place, as well as right time
Example next...

Slide 2.7

Spatial formula: example 1
Two Breeding grounds; two adult Feeding grounds;
Pure larva J sampled at B1; impure adults at F1



$$\begin{aligned}&\mathbb{P}[K_{AJ} = \text{POP}|pay_A pay_J ...]\\ =&\sum_{\text{possible } x}\{\mathbb{P}[K_{AJ} = \text{POP}|x; ay_A ay_J] \times\\ &\mathbb{P}[A \text{ at } x \text{ when } J \text{ born}|pay_A] \times\\ &\mathbb{P}[J \text{ born at } x|pay_J]\}\end{aligned}$$

$\mathbb{P}[A@\text{F1 is from B1}] = \frac{\mu_1 N_1}{\mu_1 N_1 + \mu_2 N_2}$
μ: *mixing params*, usually unknown in advance
$\mu_1 = \mu_2$ if F1 *well-mixed*

$$\begin{aligned}&\mathbb{P}[K_{AJ} = \text{POP}]\\ =&\frac{\text{"1"}}{N_1} \times \frac{\mu_1 N_1}{\mu_1 N_1 + \mu_2 N_2} \times 1 \ + \ 0 \times \frac{\mu_2 N_2}{\mu_1 N_1 + \mu_2 N_2} \times 1\\ =&\text{"1"} \times \frac{\mu_1}{\mu_1 N_1 + \mu_2 N_2}\end{aligned}$$

"It's abundance, Jim, but not as we know it..."

Slide 2.8

Spatial formula: example 2
Adult site well-mixed, juve site impure:

$$\begin{aligned}&\mathbb{P}[K_{AJ} = \text{POP}]\\ =&\frac{\text{"1"}}{N_1} \times \frac{N_1}{N_1 + N_2} \times \frac{\mu_1 N_1}{\mu_1 N_1 + \mu_2 N_2} +\\ &\frac{\text{"1"}}{N_2} \times \frac{N_2}{N_1 + N_2} \times \frac{\mu_2 N_2}{\mu_1 N_1 + \mu_2 N_2}\\ =&\frac{\text{"1"}}{N_1 + N_2} \times \frac{\mu_1 N_1 + \mu_2 N_2}{\mu_1 N_1 + \mu_2 N_2}\\ =&\frac{\text{"1"}}{N_1 + N_2}\end{aligned}$$

TL;DR:
if one site WM, then estimate *total* abundance, regardless of other site
Note that "well-mixity" may only be true for some ages.

Slide 2.9

Spatial formula: example 3

Impure adults *and* impure juves:

$$\begin{aligned}
&\mathbb{P}\left[K_{AJ} = \text{POP}\right] \\
&= \frac{\text{"1"}}{N_1} \times \frac{\mu_{J1}N_1}{\mu_{J1}N_1 + \mu_{J2}N_2} \times \frac{\mu_{A1}N_1}{\mu_{A1}N_1 + \mu_{A2}N_2} + \\
&\frac{\text{"1"}}{N_2} \times \frac{\mu_{J2}N_2}{\mu_{J1}N_1 + \mu_{J2}N_2} \times \frac{\mu_{A2}N_2}{\mu_{A1}N_1 + \mu_{A2}N_2} \\
&= \text{"1"} \times \frac{\mu_{J1}\mu_{A1}N_1 + \mu_{J2}\mu_{A2}N_2}{\left(\mu_{J1}N_1 + \mu_{J2}N_2\right)\left(\mu_{A1}N_1 + \mu_{A2}N_2\right)}
\end{aligned}$$

Enjoy!

Slide 2.10

Spatial CKMR...

Formulae generalize to >2 stocks and sites. "Easy" when:

- site is well-mixed ($\mu_1 = \mu_2$)
- or pure (all-but-one $\mu_i = 0$)

  Otherwise, each site has its own mixing params, to be estimated from CKMR (or via a miracle...).

Sometimes you can estimate all params unambiguously (NB there's multiple cross-site comparisons):

- EG: two partly-mixed Feeding grounds, and two pure Breeding grounds: can estimate both mixing rates, and both N's

but sometimes you can't: too many parameters (e.g. just from a small number of impure sites).

In that case, a simple aggregate model (for $N_1 + N_2$) will usually be biased, *but* unless things are extreme, the bias *may not be large*. However, we could probably do a bit better with more modelling...

3 E-ABT Spatial CKMR

Slide 3.1

(E)ABT as seen from Australia

- Overlaps *genetically distinct* Western (GoMex) ABT in mid-Atlantic & Canada/NE USA
- 3(+) breeding grounds in Med
  - **B**alearic, **C**entral, and **E**(verything else!)
  - AFAIK, main *managed* fisheries nowadays on B & C fish

Slide 3.2

"Strawman" fisheries/programs

You do **NOT** have to sample *everywhere* in order to do CKMR.
But you do have to sample *enough* places to allow for (and test) spatial stuff, and to e.g. get all adult sizes, enough juvenile cohorts, etc...

- Implicit assumption: B/C/E *faithful* (maybe *heritable*)
  - so, "mixing" in table means "relative to B/C/E stocks". (Especially B&C)
  - Size/age *always* matters; "well-mixed" maybe only some ages
- Main focus at start is B & C; can add more places later
- Next slides show how we could test and what we should learn

Place	Type	Mixing	Samp size	Limitations/ benefits
Bal larvae	J	pure	Large	- Intra-cohort sibs
Cro 2/3yo (+1yr)	J	impure C(+B+E)?	large?	+ extend juve cohorts
Prt traps	A*	well-mixed*	?medium	
Norway	A	well-mixed	?small	
Canada	A	well-mixed	medium	E & W together (∼50% E)
Cmed ad (Malta*)	A*	pure	medium?	¿- size range?
B ad (Spain)	A*	pure	medium?	¿- size range?

Slide 3.3

How could we..?

Place	Type	Mixing	Samp size	Limitations/ benefits
Bal larvae	J	pure	Large	- Intra-cohort sibs
Cro 2/3yo (+1yr)	J	impure C(+B+E)?	large?	+ extend juve cohorts
Prt traps	A*	well-mixed*	?medium	
Norway	A	well-mixed	?small	+2nd site in At1
Cmed ad (Malta*)	A*	pure	medium?	
B ad (Spain)	A*	pure	medium?	

HSPs useful too, but let's concentrate on POPs for clarity

Test heritability?	C/B"Ad" - C/BAd	Might take a while
Test faith? & Cro purity	CAd-Bju, BAd-CroJu vs BAd-Bju,CAd-CroJ	Can est. stickyness if not
Est *overall* abund & test well-mix	Bju - PrtAd; should be same as CroJu - PrtAd and Ju-NorAd	Everything is time/age adjusted...
Est. "stock-specific abund"	Bju - BAd ¿CroJu - CAd? ¿E by subtraction?	depends on faith & purity
Est fec-at-size	CroJu - Prt/Nor and - Cmed (check)	CroJu faster than BJu

These slides were produced in Beamer. They look ugly and it was hard. *I hate Beamer!!!*

Slide 3.4

What to do with all that..?

These "placewise thought-experiments" are interesting and broadly useful! But, there's too many and too ad hoc to investigate via zillions of "would such-and-such oversimplified model be biased?"
Rather, construct one or two overall "full models", having:

- site-specific mixing parameters
  - *possibly* age-specific...
- ... with strong "expert" priors
- ¿Models qualitatively differing WRTO faithfulness, which should be discovered "quickly"
- Concentrate on achieving
  - adequate overall variances (assuming priors OK)
  - ability to *test* over time: ie (in)validate priors
    - eg, need useful numbers of CroJu-Prt/NorAd *and* Bju-Prt/NorAd kin
  - some sensitivity tests (if priors badly wrong)
- This would be the hardest CKMR design yet!

Slide 3.5

Basic strategy..?

After doing a *proper design* as per last slide:

- Can maybe start some sampling earlier; only genotype later
- Start collecting data from *enough* core fisheries/programs to give basic answers fairly quickly
  - with some redundancy
- If there's a *massive* assumption failure, it should become clear...
  - via big differences between place-specific POP/HSP frequencies that were expected to be similar
- if there's not a *massive* difference, then there might be *subtler* biases (e.g. mixing somewhere is not *perfect*) that stay hidden, but are OK for short-term management...
- Can then go on collecting data, and adding new places, to refine things

Slide 3.6

Reality check

- Say we need 50,000+ samples overall, to generate 100–200 kin-pairs (e.g. 2017 rough calcs)
- Only $< 1\%$ of samples will be in *any* kin-pair!
- If a sample source has 1000 samples, it can only expect ~ 10 kin-pair-members...
  - and each new sample source requires new mixing param(s)...
- Can't get useful *quantitative* info from that few kin...
- ... so, only include *substantial* sample-sources
- [Big adults do give *somewhat* more kin per sample.]

Slide 3.7

Blueprint for preliminary work
Specific studies:

- Epigenetic age
- Genotyping details/costs
- [Large-scale Balearic sibship]
- Sampling protocols
- **Design**: build model(s) and explore realistic sampling options for core subset of fisheries/programs
- ...

In CKMR it's always better to **over-sample**: you won't necessarily have to spend the genotyping $ on *all* of them, but if you didn't collect them in the first place, then...