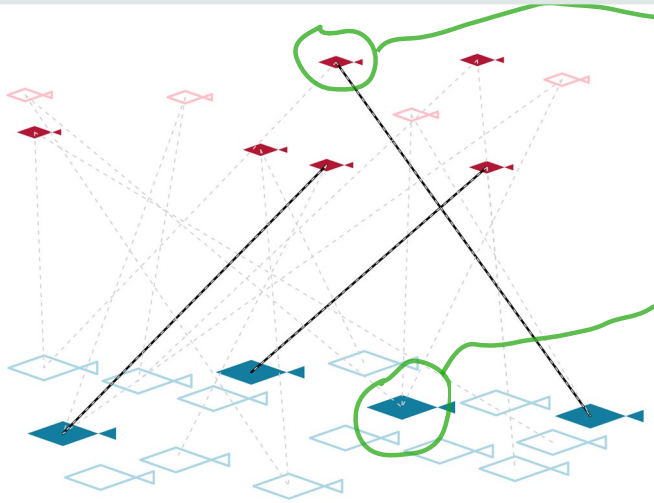
CKMR:
models and designs
for fisheries and bycatch

Mark Bravington, CSIRO: June 2021

CKMR is...

Biopsies from juves & adults (dead OK) over a few years

Some idea about age/sizes



You have 1 mother

Chance it's *this* ♀ = $1/N_{♀}$

This formula is *too simple* for realistic situations...

$$\hat{N}_{\text{ad}} = \frac{2 \times \#\text{comps}}{\#\text{POPs}}$$

Parent-Offspring Pair (MOP or FOP)

Also: Half-Sibling Pair, HSP

CKMR is...

Biopsies from juves & adults (dead is OK) over a few years

Some idea about age/sizes

You have 1 mother

Chance it's *this* ♀ = $1/N_{♀}$

Check genetics...

... repeat for all pairs in sample

... estimate adult ♀ abundance!

... ditto for ♂ adults

This formula is too simple for realistic situations, but can be adjusted eg time, age, size, mortality

$$\hat{N}_{ad} = \frac{2 \times \#\text{comps}}{\#\text{POPs}}$$

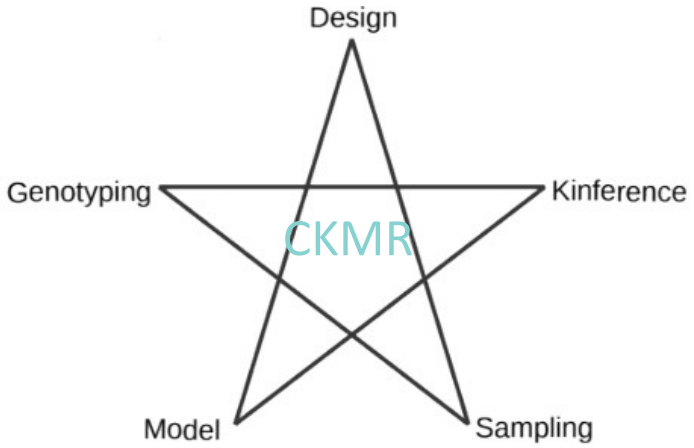
CKMR models/designs: fisheries 2021

1. Intro
    - ideas / projects / history
    - kinship probs and mind-check
    - putting them together--- “The Framework”
    - sample size & scope
2. Realistic POP examples--- mammal
    - can we break it?
    - different sampling issues
3. Super-simple POPs
4. HSP examples--- mammalesque sharks
    - genetics in a nutshell
5. Fish time! POPs + HSPs
    - what does CKMR *really* tell us?
6. Spatial stuff
7. Design

By the end, you should...

CKMR is a 4-letter word...

easy



MAGICAL RESULTS ...

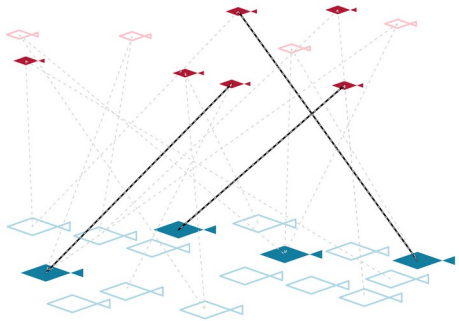
come from CAREFUL ENGINEERING !

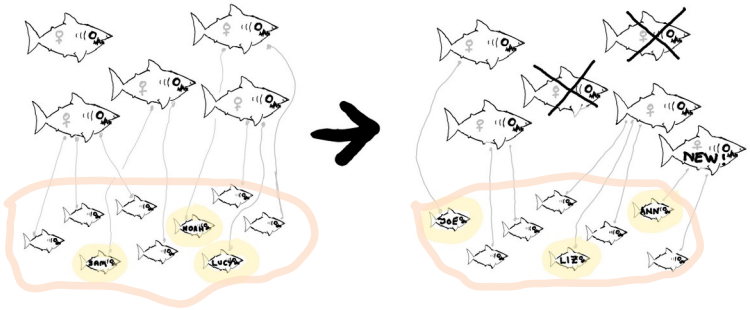
CKMR is...

cross-cohort half-sibs

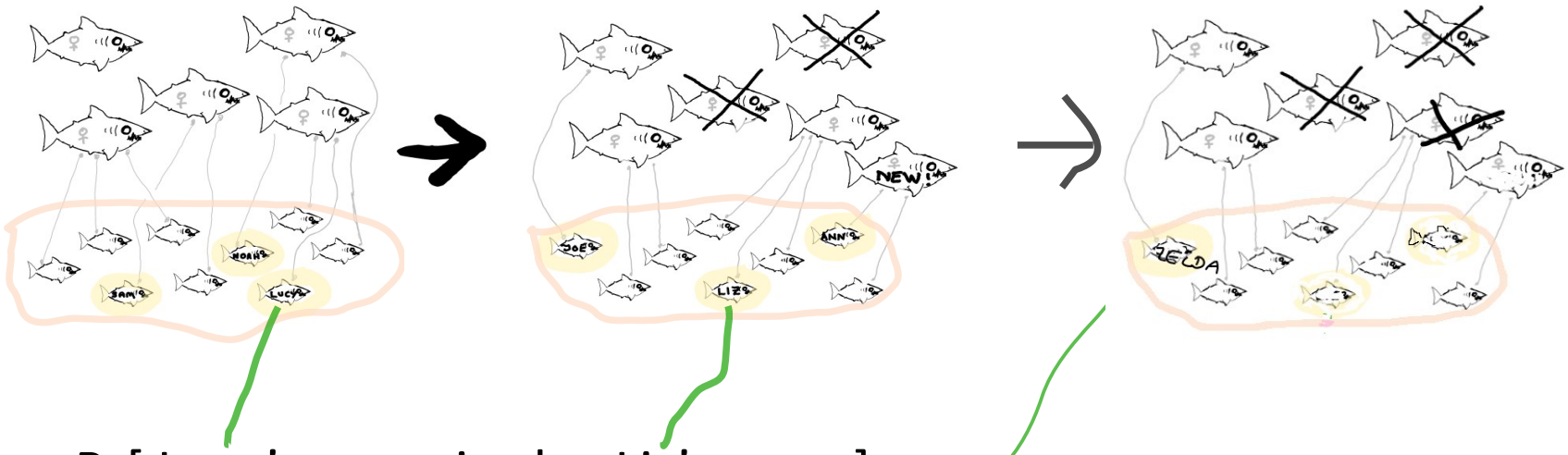
Parents are “marked” by their sampled offspring

Direct recapture (POPs) and **Indirect** (XHSPs)





HSPs: abundance* and mortality* info



Pr[Lucy's mum is also Liz's mum]

>

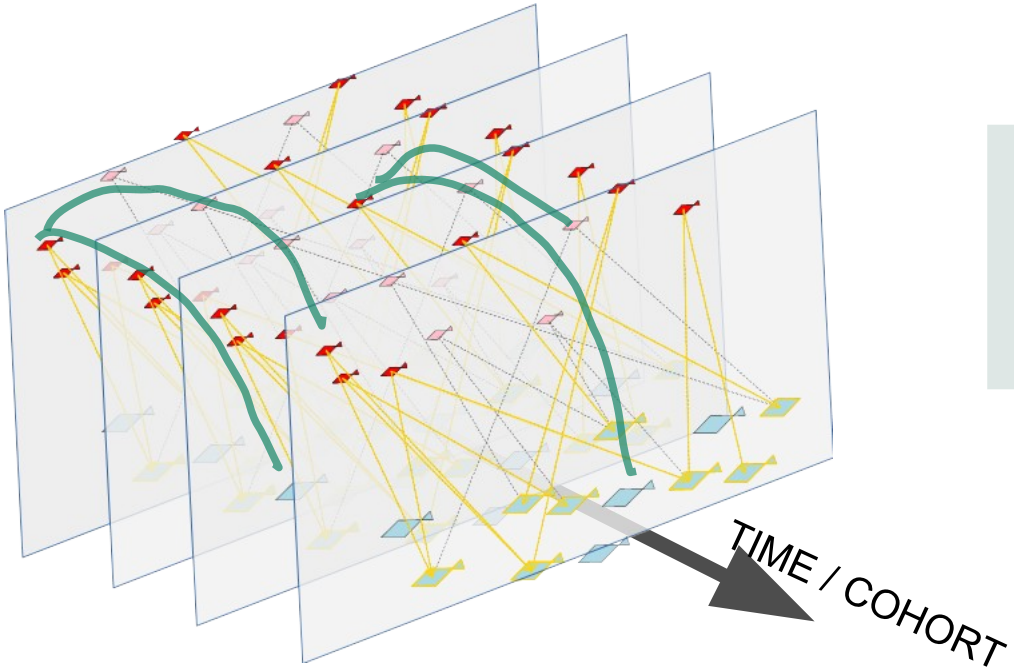
Pr[Lucy's mum is also Zelda's mum]

CKMR is...

cross-cohort half-sibs

Parents are “marked” by their sampled offspring

Direct recapture (POPs) and **Indirect** (XHSPs)



- Lots of comparisons
- Different prob formulae
- More parameters than just “N”

Don’t *have* to use *all* possible types of comparison

Fitting a CKMR model

Chassis: a standard-ish Age- and Sex- structured Pop Dyn model

PARAMETERS: "populate" the Pop Dyn Model

$$N_{a+1,t+1,s} = N_{ats}e^{-z} \;;\; a \geqslant 8,\; t \geqslant 2002$$
$$\log N_{8t} \sim N\left(\mu, \sigma^2\right)$$
$$\log N_{2002,a} \sim N\left(\mu e^{-z'(a-8)}, \sigma^2\right)$$

COVARIATES: for each pair of samples (eg "Mary" and "Simon"),

use ERRO to work out POP and HSP probabilities, e.g.

Pr[Mary is Simon's mother | Mary's covariates, and Simon's]

RESPONSE DATA: the kinship of each pair: POP, HSP, UP (unrelated)

inferred from their genotypes

LOG-LIKELIHOOD: *lots* of Bernoulli (yes-no) comparisons

Can put in other data too

CKMR is...

Biopsies from juves & adults (dead is OK) over a few years

Some idea about age/sizes

Two assumptions:

1. At birth, everything had 1 living mother and 1 living father
2. Reliably find *Parent-Offspring-Pairs* and *Half-Sibling-Pairs* with genetics

The rest is just *planning;* *logistics;* *maths;* and a lot of *FOR-loops*

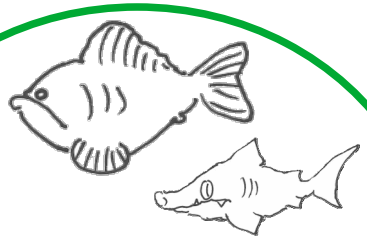
Absolute abundance --- and almost *entire* stock assessment for adults --- just from biopsying a few % of catch

- no $urvey$, no CPUE, no live-relea$e, no dodgy assumptions...

CLOSE-KIN TREE OF LIFE

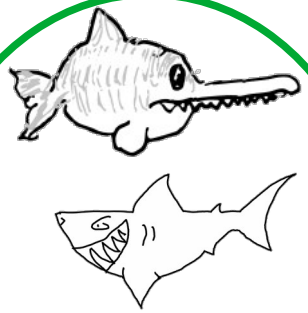
simplified version!

adults grow a lot



demanding!
only one good
way to sample

adult fecundity
does not change



more options
for sampling

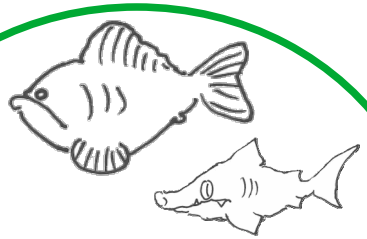
breed-and-die

forget it
(except if...)

CLOSE-KIN TREE OF LIFE

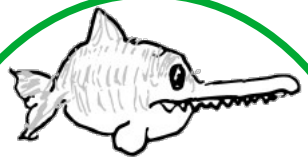
simplified version!

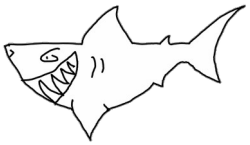
adults grow a lot



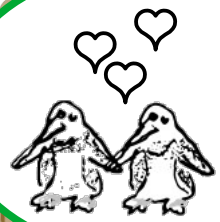
demanding!
only one good
way to sample

adult fecundity
does not change





more options
for sampling



breed-and-die

forget it
(except if...)

ETC ... ETC ....

Some myths, busted

“CKMR needs validation”

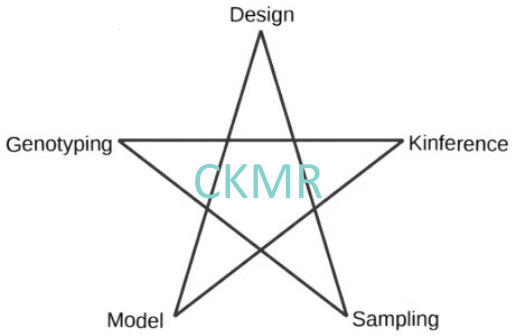
Some myths, busted

"CKMR needs validation"

That is NONSENSE! The basic assumptions are:

- everything had 1 mother and 1 father at birth;
- genetics lets you find POPs and HSPs

The rest follows by ironclad laws of mathematics...



That said:

of course people could *completely stuff up*

any particular CKMR study

So: *check* each study carefully!

can't "validate" directly-- against what???

But that has nothing to do with "validating CKMR" in general

Some myths, busted

~~CKMR needs validation~~

“It’s got to be done exactly like SBT c. 2014”

Some myths, busted

~~CKMR needs validation~~

"It's got to be done exactly like SBT c. 2014"

No it doesn't! CKMR is fairly flexible.

EG ideally you *don't* want to sample spawning grounds...

you *don't* absolutely need adults-*and*-juves well-mixed tho' it is best

you *do* need POPs and HSPs **for fish but not necessarily sharks**

- we had to use a special trick in SBT mk I POPs-only...
- ... which won't work in other species...
- ... and turned out not to be fully valid ...
- ... when we did mk II POPs+HSPs. But it's all good now !

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

"I've seen the cartoon. That 2/N formula will be biased if..."

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

"I've seen the cartoon. That 2/N formula will be biased if..."

Of course it won't work! *You're not meant to* use *2/N for real...*

CKMR theory tells you how to adapt the idea for real situations. Need a qualitative understanding of the biology and sampling, plus (a bit of) maths.

So, do that instead !!! And then you will *avoid bias*.

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

~~I've seen the cartoon. That 2/N formula will be biased if…~~

"It's expensive…"

* Unless you have to do something weird

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

~~I've seen the cartoon. That 2/N formula will be biased if...~~

"It's expensive..."

It is NOT* ! You sample a *tiny fraction* of the catch that happens anyway; each one is cheap to genotype (<$ [redacted]).

Much cheaper than dedicated surveys or normal MR if those even apply

Sure--- CPUE is free. Super! You totally get what you pay for there.

Plus, co$t of mi$management due to mi$leading conventional data.

May be uneconomic for small pop'ns of low-unit-value short-lived spp.

* Unless you have to do something weird

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

~~I've seen the cartoon. That 2/N formula will be biased if…~~

~~It's expensive…~~

"It needs really careful attention-to-detail and a lot of know-how"

Some myths, busted

~~CKMR needs validation~~

~~It's got to be done exactly like SBT c. 2014~~

~~I've seen the cartoon. That 2/N formula will be biased if...~~

~~It's expensive...~~

"It needs really careful attention-to-detail and a lot of know-how"

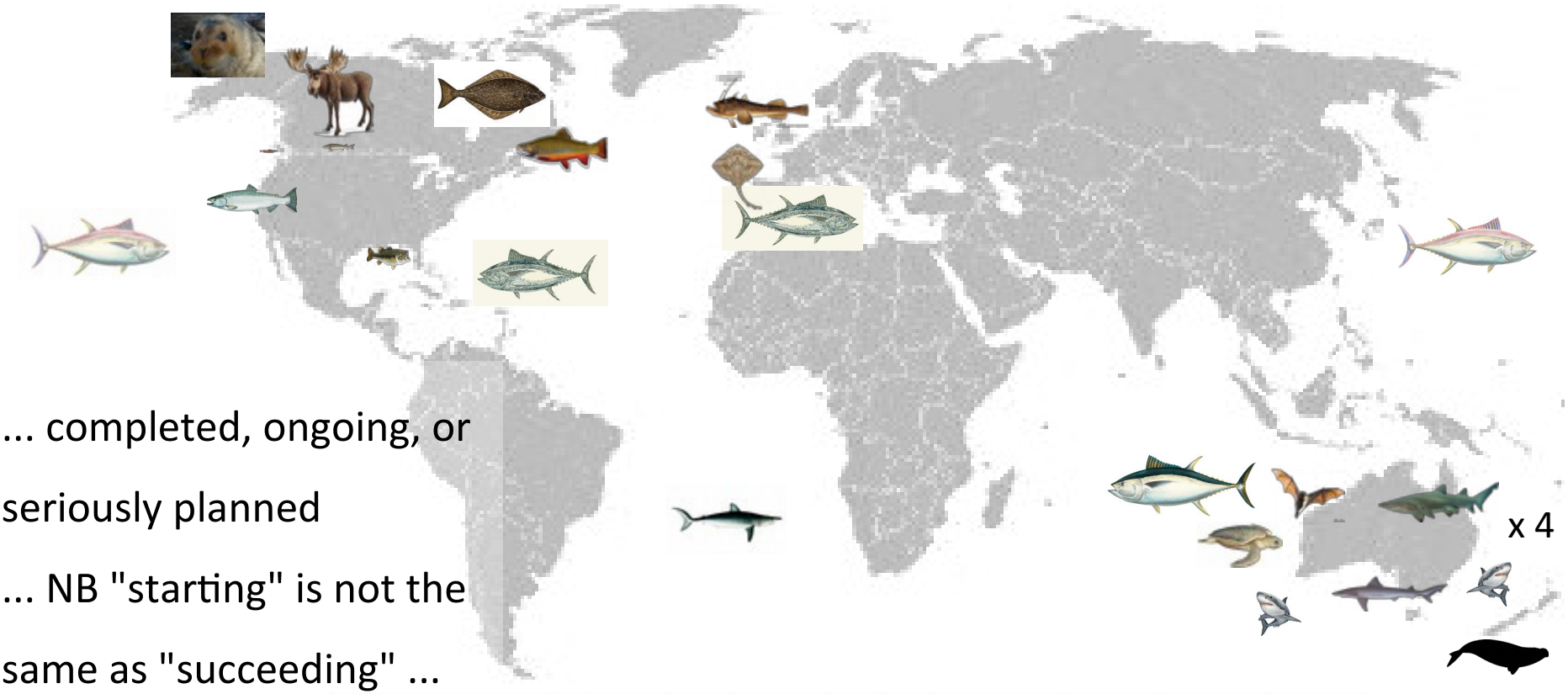
This one is **not** a myth! If you blunder into CKMR without a clear plan and without the right team, *things will end badly*. So: don't do that!

"DNAge" (methylation-based epigenetic ageing)

To date, **worst** problems are from **lousy age and/or length** data

- DNAge devel ongoing at CSIRO and elsewhere
    - prelim results 4 spp: from decent to amazing !
    - *relative* errors, ie better for younger
- **Cheap!** Under $[redacted] per sample and low setup cost
- A(nother) game-changer and useful beyond CKMR
- Needs calibration per species, then
- All you'll need: a quick jab and a length measurement

(some) CKMR projects c. 2020



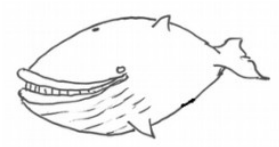
... completed, ongoing, or seriously planned

... NB "starting" is not the same as "succeeding" ...

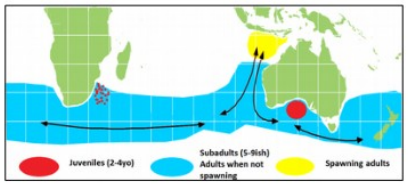
Species		Source	Geno	Motive	POP	HSP	Model	"Stocks"?	N_{adult}	#Samples
SBTuna	1	Dead	usat	$	✓	–	Full pop-dyn	–	■,000,000	15,000
	2	D	Cap		✓	✓				20,000
School Shark		D	Cap	$ (choke)	(–)	✓	"	–	■00,000	3,000
White Sh.	E	Live+D	Seq	!	–	✓	N, z, ρ	–	■00	200
	W	L+D	Seq		(–)	✓		–	■000	200
Grey Nurse Sh.		L+D	Seq	!	✓	✓	Stable-age	–	■000	400
Speartooth Sh.		L	Cap	!?	–	✓	N, z, ρ	(✓)	■000	300
Different speartooth Sh.		L	Seq	!?	–	✓	N, z, ρ	(✓)	■000	300

A long and winding road...

2000: Skaug pairwise comps; brilliant, but didn't work...



2005: Idea -- SBT ?
2007: FRDC project



sampling; usats; genotyping; POP-finding

2012: 13,000 samples
POP model
2013: into CCSBT asst

...more sampling...

CKMR for sharks:
no POPs -- how about HSPs?
But how? Time for SNPs !
Dart-- DartSeq

2015: DartCap -- high-volume SNPs to find HSPs

...more sampling...
2016: theory paper

2018: 20,000 samples
POP + HSP model!
Fully **F**ishery-**F**ree !
CCSBT now pay

A(nother) bit of history

1990s: Norway establishes DNA register of commercial minke whale catch

Schweder: Aha-- CK for abundance!

2001: Skaug publishes one method on Norway minkes

- right idea, wrong example, too early !
- genotypes inadequate, model not general enough
- and Nilssen notes in passing (but needs triads)

2005: CSIRO starts SBT CKMR

2011-13: ... starts other T (hreatened) EPS projects

2013: General demography framework (MVB & HJSkaug)

2017-18: framework for design

Other work to date: some stock structure

Individual MR (gene-tagging) now fairly common

Rawding et al 2014: "Transgenerational MR" (semelparous)

What you can get from CKMR …

Biopsies from juves and adults *eg from landings* ***plus*** *some size/age info:*

absolute abundance of adults

relative **fecundity-at-size** ♀ *and* ♂

And *if you also know catch-at-age, and have growth curve*

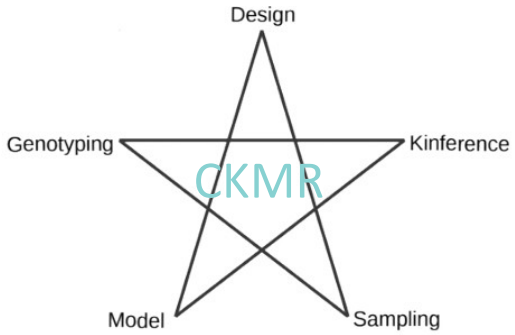
natural mortality averaged across adults

And

connectivity on management timescale (1 generation)

Provided that…

you do everything properly :)



Thus endeth the kinspiel...

... time to buckle up!

ERRO: the key to it all

Expected **R**elative **R**eproductive **O**utput

$$\mathbb{P}\left[\text{Amy is Julian's mum} \middle| \text{stuff about A \& J}\right]$$

$$= \mathbb{E}\left[\frac{\#\text{A's J-like offspring @ } b_J}{\text{Total \# J-like offs @ } b_J} \middle| \text{stuff about A}\right]$$

LEXIGRAM GOES HERE...

b: **b**irth-year

y: when **y**anked from the sea

a: **a**ge (usually at *y*)

t: **t**ime in general

Factors to always *consider* for ERRO

TIME

SEX of “adult”

AGE of “juve”

AGE and maybe SIZE of “adult”

PLACE

QUIRKS

“adult”: *potential parent*
(seen in POP, unseen in HSP)

“juve”: *potential offspring*
(in POP or in HSP)

“*I didn’t measure it*” **doesn’t** imply that “*it doesn’t matter*”

... see Examples!

ERRO for HSPs

Fred is born first; Lucy later

$$\begin{aligned}&\mathbb{P}\left[\text{Fred \& Lucy are MHSP} \mid \text{stuff about F \& L}\right]\\&\quad = \mathbb{P}\left[\text{Lucy has same Mum as Fred} \mid \text{stuff}\right]\end{aligned}$$

ERRO for HSPs

Fred is born first; Lucy later

$$\mathbb{P}\left[\text{Fred \& Lucy are MHSP} \mid \text{stuff about F \& L}\right]$$

$$= \mathbb{P}\left[\text{Lucy has same Mum as Fred} \mid \text{stuff}\right]$$

$$= \sum_{\text{"Mary"} \in \text{Fred's possible mothers}} \mathbb{P}\left[\text{Fred's Mum was "Mary"}\right] \times \ldots$$

$$\ldots \mathbb{P}\left[\text{"Mary" survived til } b_{\text{Lucy}}\right] \times \ldots$$

$$\mathbb{P}\left[\text{Lucy's Mum was "Mary"} \mid \text{"Mary" alive at } b_{\text{Lucy}}\right]$$

ERRO for HSPs

Fred is born first; Lucy later

$$\mathbb{P}\left[\text{Fred \& Lucy are MHSP} \mid \text{stuff about F \& L}\right]$$

$$= \mathbb{P}\left[\text{Lucy has same Mum as Fred} \mid \text{stuff}\right]$$

$$= \sum_{\text{"Mary"} \in \text{Fred's possible mothers}} \mathbb{P}\left[\text{Fred's Mum was "Mary"}\right] \times \ldots$$

$$\ldots \mathbb{P}\left[\text{"Mary" survived til } b_{\text{Lucy}}\right] \times \ldots$$

$$\mathbb{P}\left[\text{Lucy's Mum was "Mary"} \mid \text{"Mary" alive at } b_{\text{Lucy}}\right]$$

ERRO

ERRO

ie ERRO-weighted average of ERROs

ERRO for HSPs

Fred is born first; Lucy later

$$\mathbb{P}\left[\text{Fred \& Lucy are MHSP} \mid \text{stuff about F \& L}\right]$$

$$= \mathbb{P}\left[\text{Lucy has same Mum as Fred} \mid \text{stuff}\right]$$

$$= \sum_{\text{"Mary"} \in \text{Fred's possible mothers}} \mathbb{P}\left[\text{Fred's Mum was "Mary"}\right] \times \ldots$$

$$\ldots \mathbb{P}\left[\text{"Mary" survived til } b_{\text{Lucy}}\right] \times \ldots$$

$$\mathbb{P}\left[\text{Lucy's Mum was "Mary"} \mid \text{"Mary" alive at } b_{\text{Lucy}}\right]$$

ERRO

ERRO

ie ERRO-weighted average of ERROs

... breaks down if F & L in *same* cohort

"Will it be biased?" The Great Mind-Check

Stat theory says: as long as ERRO formula is OK, **no bias**

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

"Will it be biased?" The Great Mind-Check

Stat theory says: as long as ERRO formula is OK, **no bias**

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

OK, as long as there's no missing property of "Amys" that affects *both*:

(i) Amy's ERRO of J-likes among A-likes, *and*

(ii) Amy's sampling prob relative to other A-likes

after allowing for "Stuff" (which is already included in the model)

"Will it be biased?" The Great Mind-Check

Stat theory says: as long as ERRO formula is OK, **no bias**

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

OK, as long as there's no missing property of "Amys" that affects *both*:

(i) Amy's ERRO of J-likes among A-likes, *and*

(ii) Amy's sampling prob relative to other A-likes

after allowing for "Stuff" (which is already included in the model)

- "not measured" does *not* imply "I can ignore it" !
- heritable or non-heritable
- "Stuff" includes: fact of sampledness, and circumstances

"Will it be biased?" *HSP* Mind-Check

Stat theory says: as long as ERRO formula is OK, **no bias**

$$\begin{aligned}&\mathbb{P}\left[\text{Fred and Lucy are MHSPs} \,|\text{F-stuff, L-stuff}\right]\\ &= \sum_{\text{"Mary"} \in \text{poss F-mums}} \left(\text{Mary's ERRO of F-likes}\right)] \times\\ &\left(\text{Mary's future ERRO of L-likes}\right)\end{aligned}$$

OK, as long as there's no missing property of "Marys" that leads to correlated ERROs of F-likes and L-likes

- see previous

Fitting a CKMR model

Chassis: a standard-ish Age- and Sex- structured Pop Dyn model

PARAMETERS: "populate" the Pop Dyn Model

$$N_{a+1,t+1,s} = N_{ats}e^{-z} \;;\; a \geqslant 8,\; t \geqslant 2002$$
$$\log N_{8t} \sim N\left(\mu, \sigma^2\right)$$
$$\log N_{2002,a} \sim N\left(\mu e^{-z'(a-8)}, \sigma^2\right)$$

COVARIATES: for each pair of samples (eg "Mary" and "Simon"),

use ERRO to work out POP and HSP probabilities, e.g.

Pr[Mary is Simon's mother | Mary's covariates, and Simon's]

RESPONSE DATA: the kinship of each pair: POP, HSP, UP (unrelated)

inferred from their genotypes

LOG-LIKELIHOOD: *lots* of Bernoulli (yes-no) comparisons

Can put in other data too

Fitting a CKMR model

Chassis: a standard-ish Age- and Sex- structured Pop Dyn model

PARAMETERS: "populate" the Pop Dyn Model

$$N_{a+1,t+1,s} = N_{ats}e^{-z} \;;\; a \geqslant 8,\; t \geqslant 2002$$
$$\log N_{8t} \sim N\left(\mu, \sigma^2\right)$$
$$\log N_{2002,a} \sim N\left(\mu e^{-z'(a-8)}, \sigma^2\right)$$

COVARIATES: for each pair of samples (eg "Mary" and "Simon"),

use ERRO to work out POP and HSP probabilities, e.g.

Pr[Mary is Simon's mother | Mary's covariates, and Simon's]

RESPONSE DATA: the kinship of each pair: POP, HSP, UP (unrelated)

inferred from their genotypes

$$\text{LGLK+} = \begin{cases} \log p_{\text{Mary,Simon are xxP}} & \text{M\&S really are xxP} \\ \log 1 - p & \text{if not} \end{cases}$$

LOG-LIKELIHOOD: *lots* of Bernoulli (yes-no) comparisons

Can put in other data too

“The Framework”

$p \sim O(1/N)$
Binomial with tiny p = Poisson
Poissons add up, to give Poisson

```
lglk <- function( params) {
  # data prep already done & known about here...
  unpack(); # split and de-transform parameters
  popdyn();
  calc_pure_kinprobs(); # true kins, ideal covars eg true age
  calc_obs_kinprobs(); # distinguishable kins, available covars
  L := 0;
  for( K in distinct_kintypes) {
    L += sum( log_dpoisson( nkins[[K]], mean=ncomps[[K]]*kinprobs[[K]] ));
  };
  L += posthoc_pairs(); # extra info on known kinpairs, eg mtDNA
  L += other_lglks(); # age-at-length; etc; priors; CPUEeeeuggggh...
return( L)
}
```