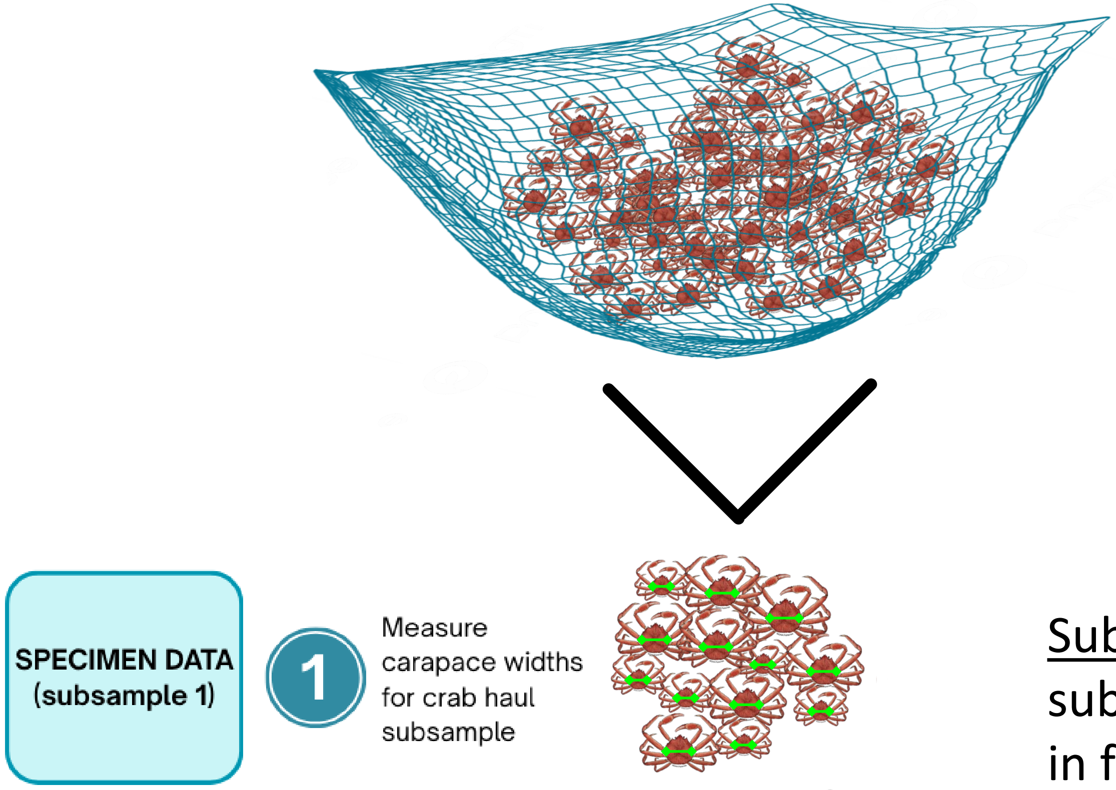
Updated *Chionoecetes* maturity workflow

Emily Ryznar
NMFS-AFSC Kodiak

Two-stage sampling designs are advantageous for population metrics that cannot be measured directly

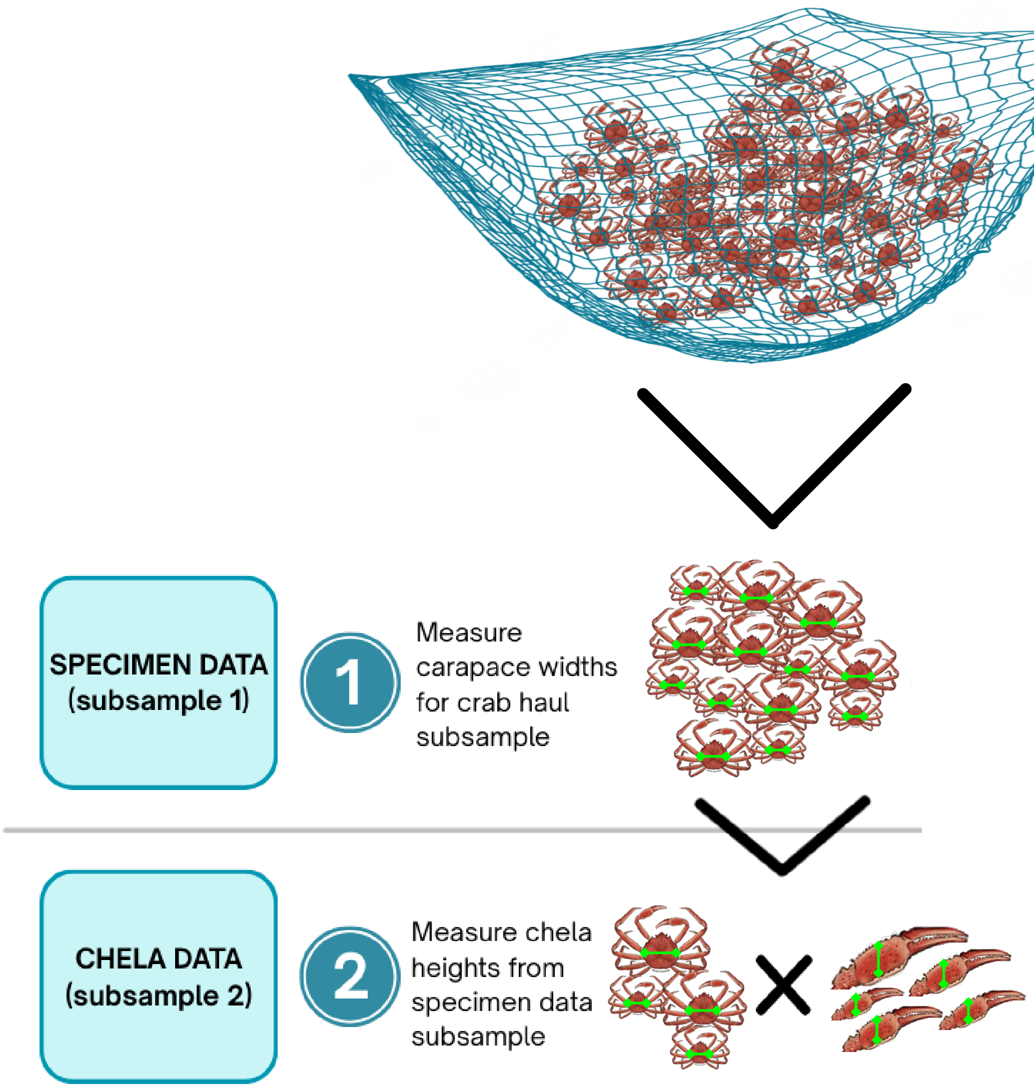


SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

<u>Subsample 1 (“specimen data”)</u>: crab subsampled for size measurements (or lengths in fish)

Two-stage sampling designs are advantageous for population metrics that cannot be measured directly



SPECIMEN DATA (subsample 1)

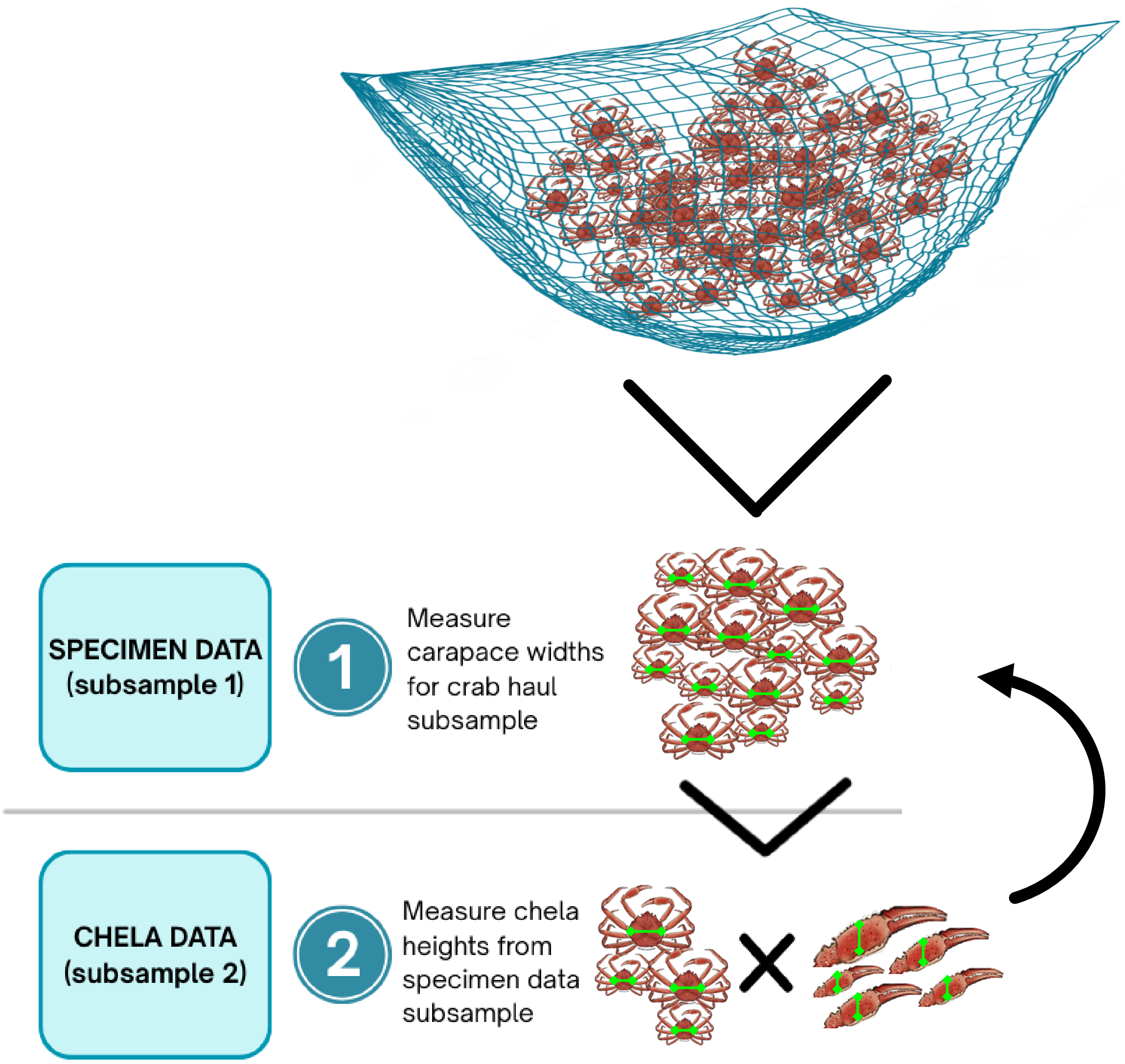
1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

Subsample 2 ("chela data"): crab subsampled from the first subsample for chela maturity measurements (or otoliths in fish for age), with sampling targets

Two-stage sampling designs are advantageous for population metrics that cannot be measured directly



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

Many approaches do not account for uneven sampling or spatiotemporal variability in this step

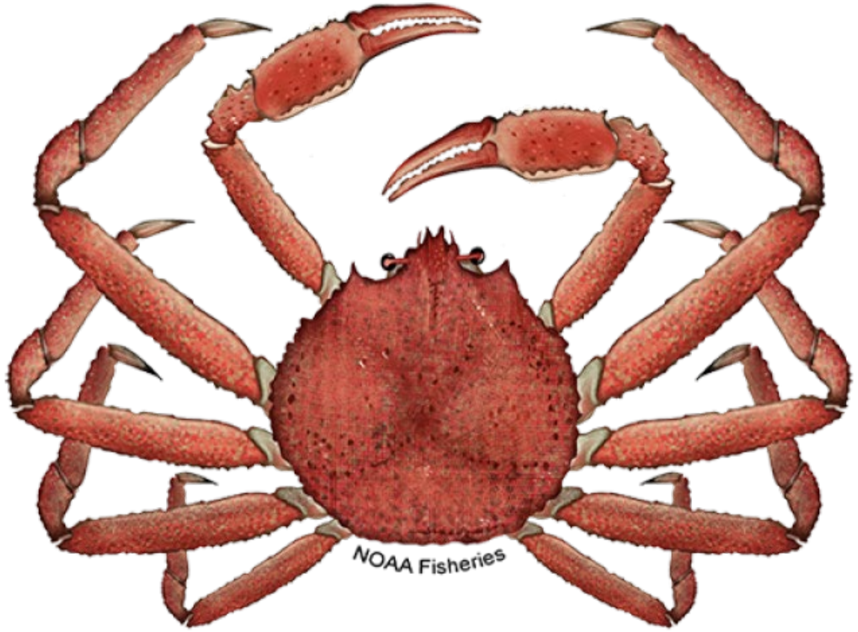
Metric of interest (maturity-at-size, age-at-length) is calculated using subsample 2 data then applied to subsample 1 via age-length-keys, models, etc.

Why is this important for Bering Sea *Chionoecetes*?

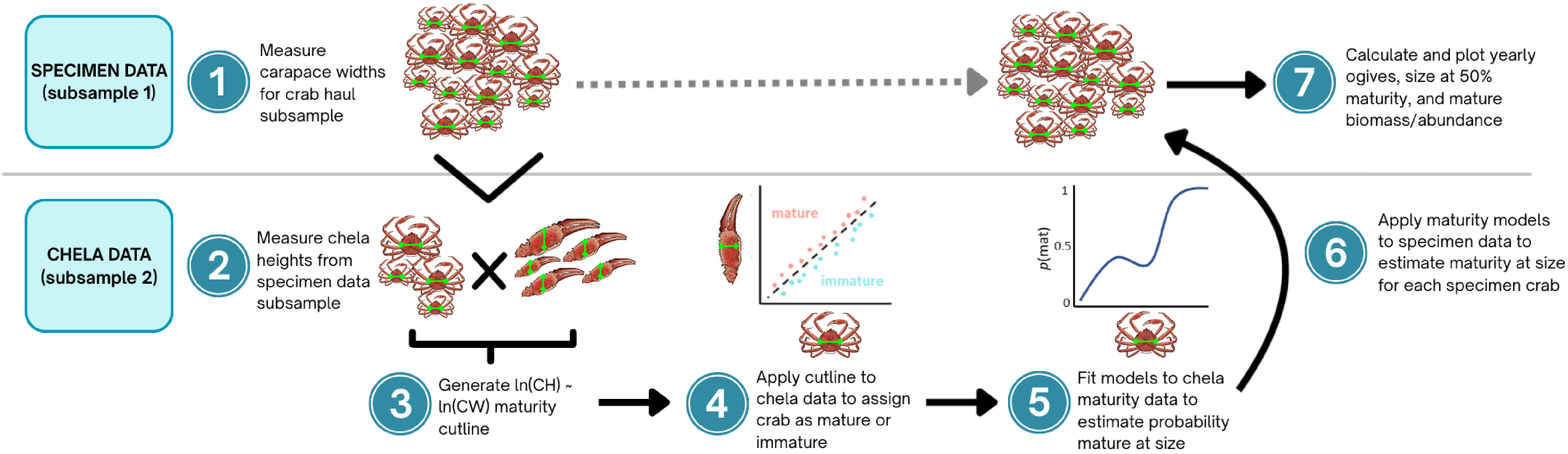
- *Chionoecetes* undergo a terminal molt to maturity
- Size-at-maturity (SAM) is an important metric for biological fitness and management
- There is evidence that SAM is declining for both snow crab and Tanner crab in the Bering Sea
- *Chionoecetes* maturity dynamics can exhibit spatiotemporal variation
- Reliable and unbiased maturity estimates are needed that account for spatiotemporal and sampling uncertainty

Goals

1. Evaluate the AFSC Shellfish Assessment Program's legacy workflow for processing *Chionoecetes* maturity data from the EBS bottom trawl survey
2. If necessary, update the workflow to reflect potential spatiotemporal variation in maturity and sampling uncertainty



Maturity data workflow



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

3 Generate ln(CH) ~ ln(CW) maturity cutline

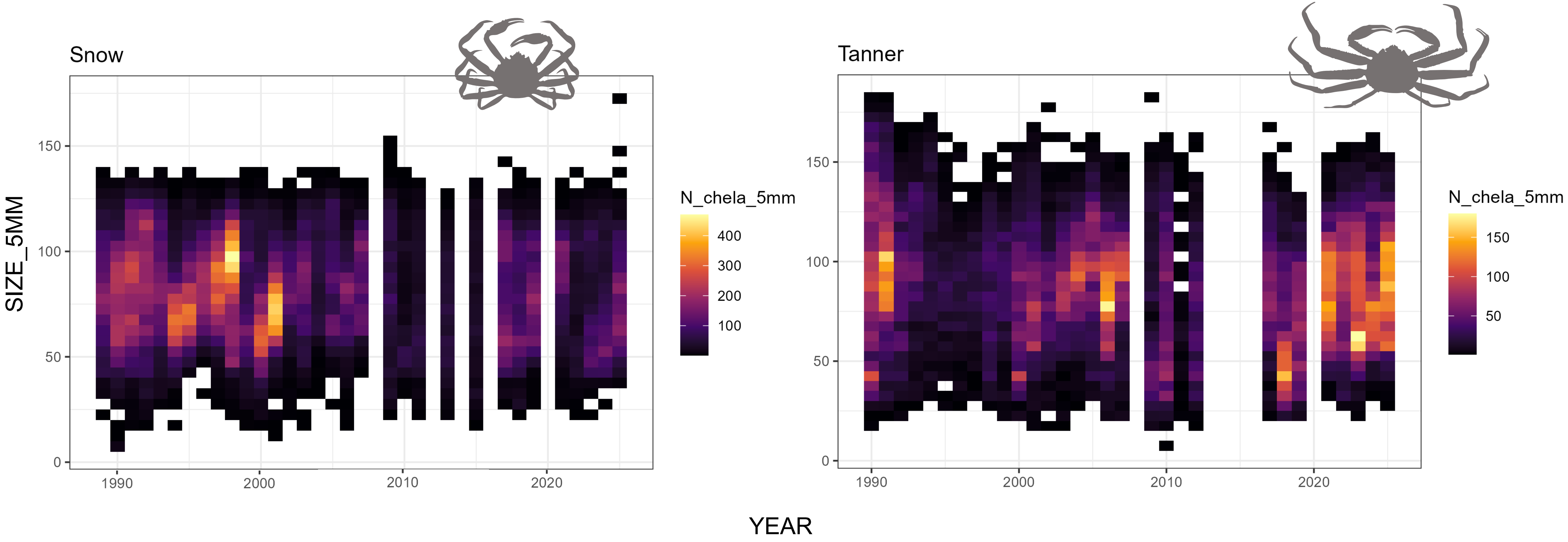
4 Apply cutline to chela data to assign crab as mature or immature

5 Fit models to chela maturity data to estimate probability mature at size

6 Apply maturity models to specimen data to estimate maturity at size for each specimen crab

7 Calculate and plot yearly ogives, size at 50% maturity, and mature biomass/abundance

Chela data nuances

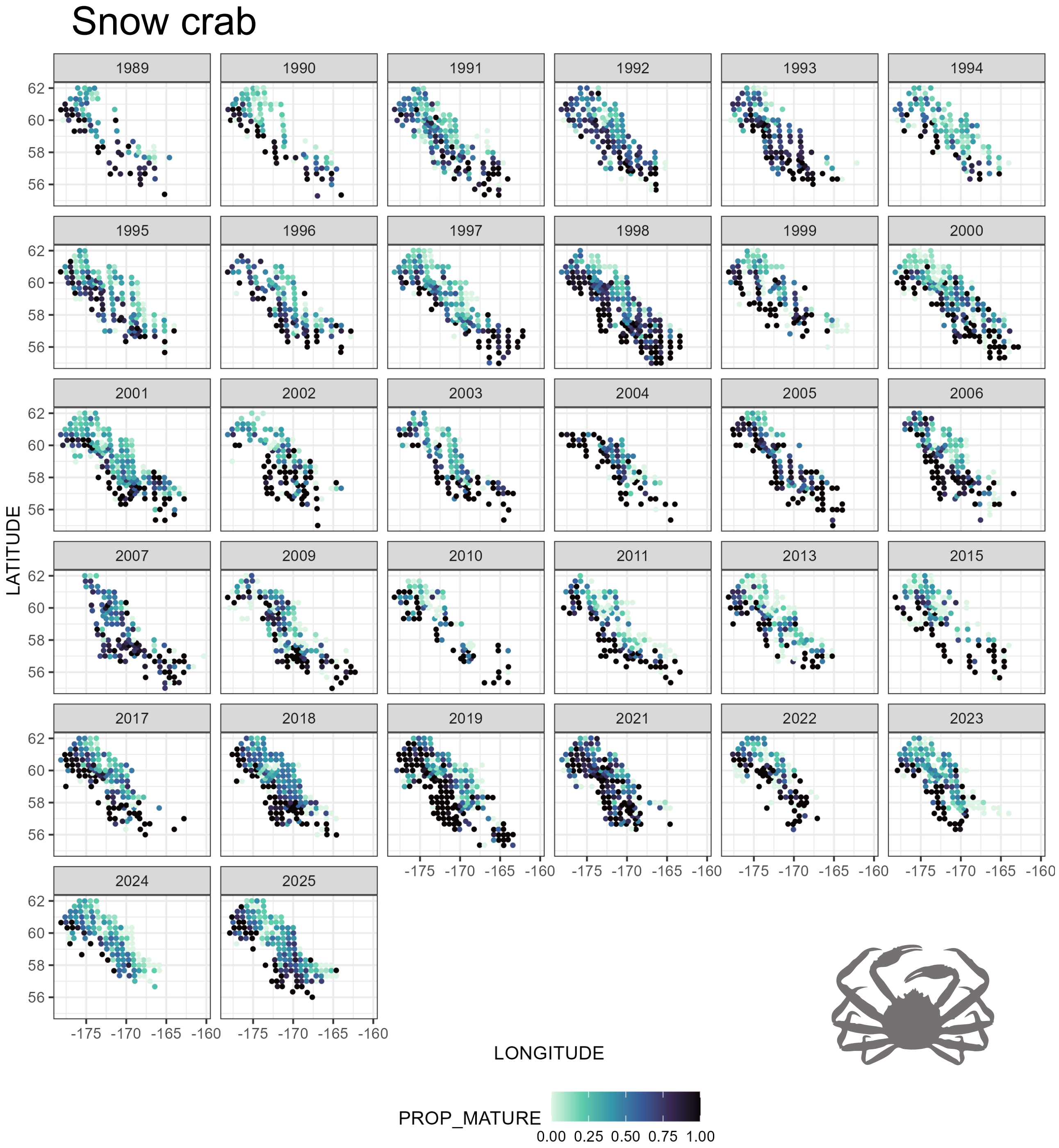


→ Sparse/uneven sampling across years and size bins

Chela data nuances

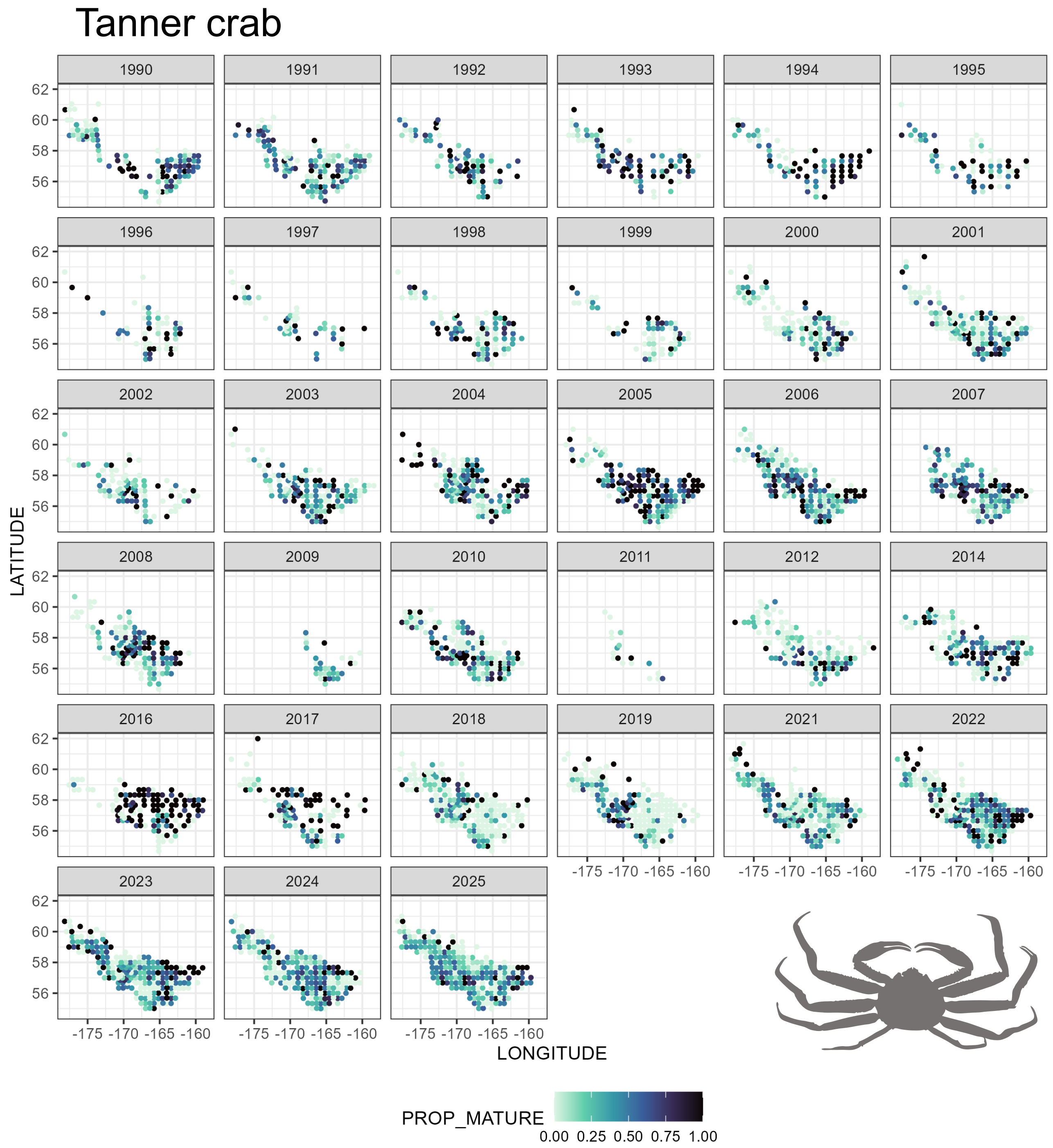
→ Uneven sampling across space

→ Evidence of spatiotemporal variability in maturity



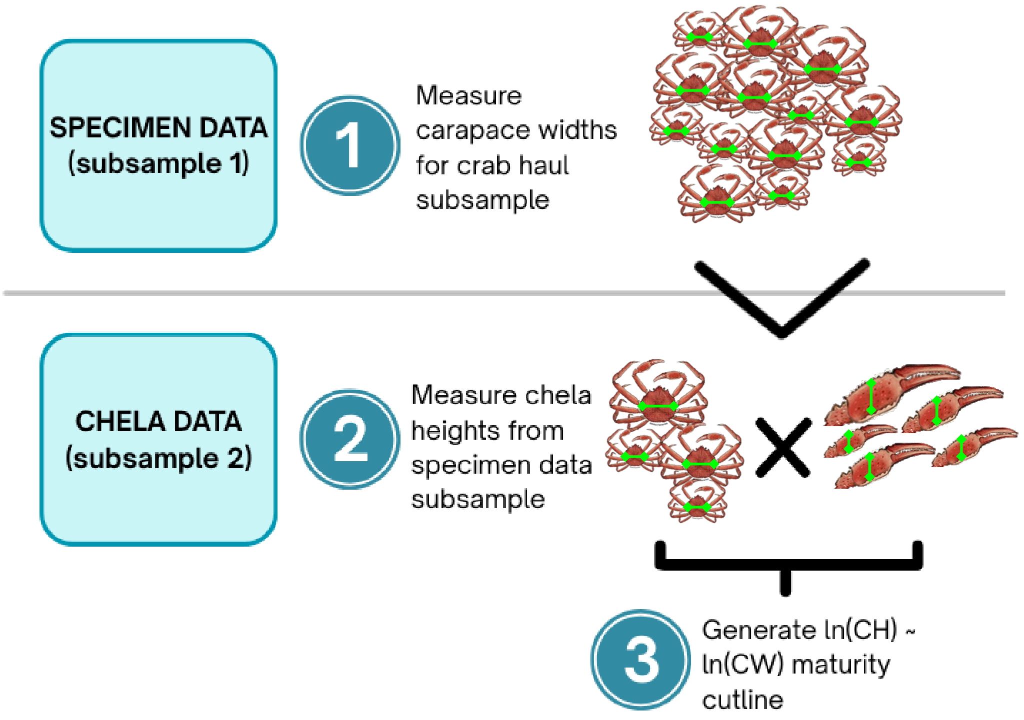
Chela data nuances

- → Uneven sampling across space
- → Evidence of spatiotemporal variability in maturity



3

Establishing maturity cutlines



SPECIMEN DATA (subsample 1)

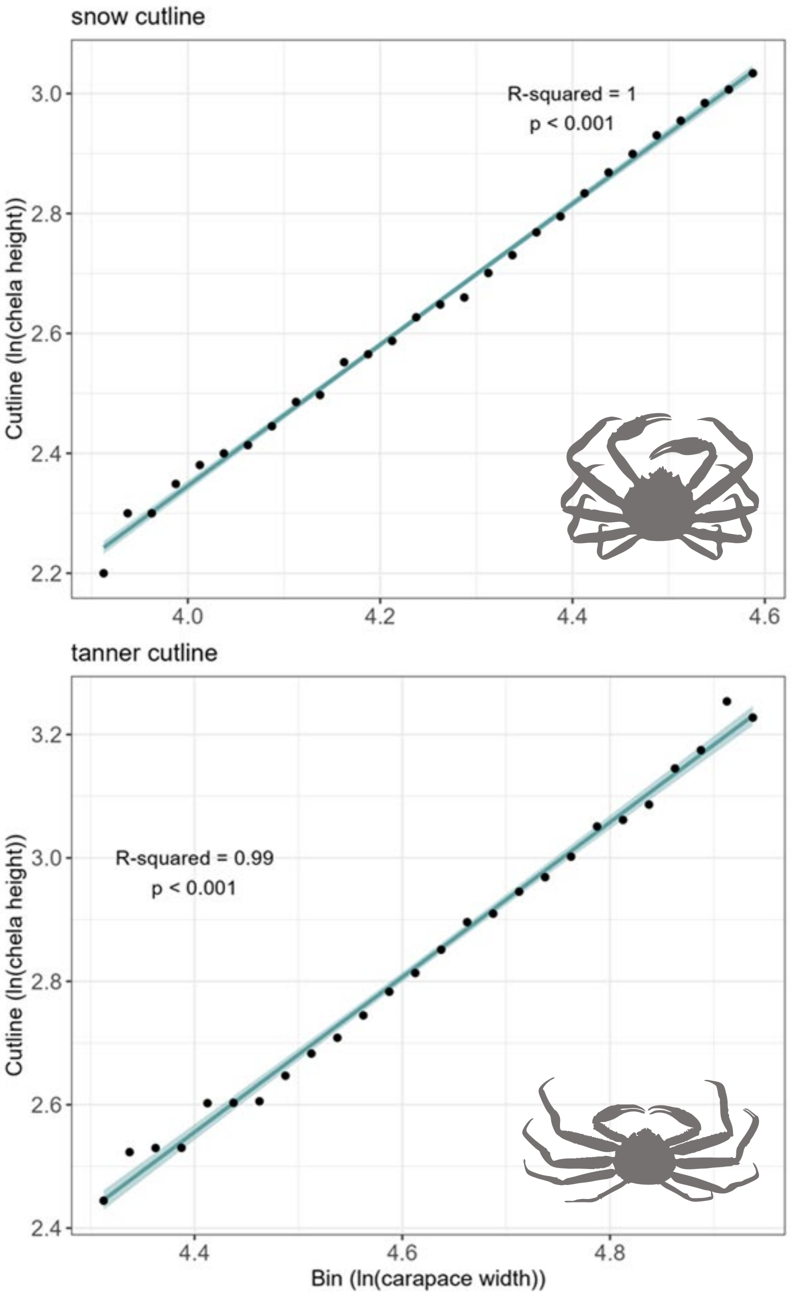
1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

3 Generate ln(CH) ~ ln(CW) maturity cutline

Only shell 2!

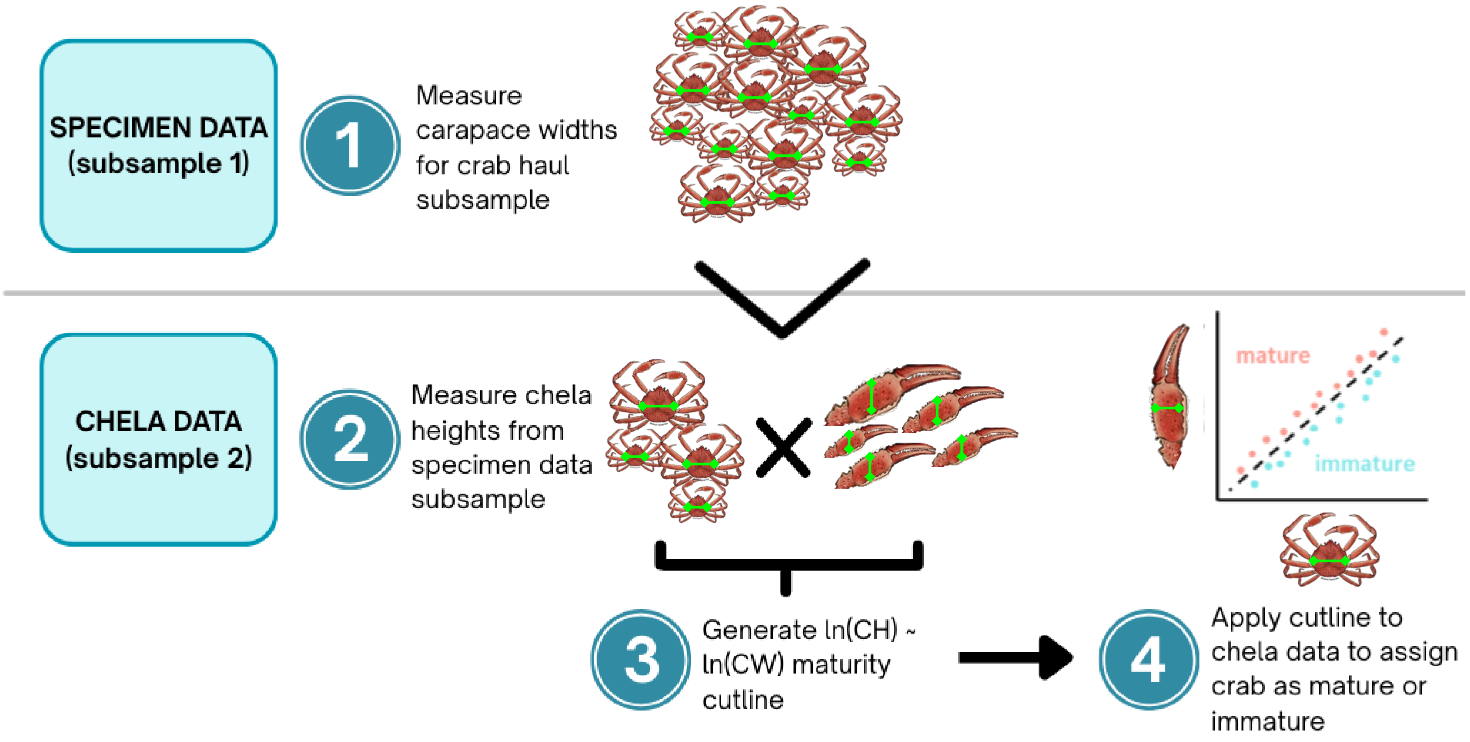

snow cutline

tanner cutline

Maturity cutlines established following Richar and Foy 2022

4

Applying maturity cutlines



SPECIMEN DATA (subsample 1)

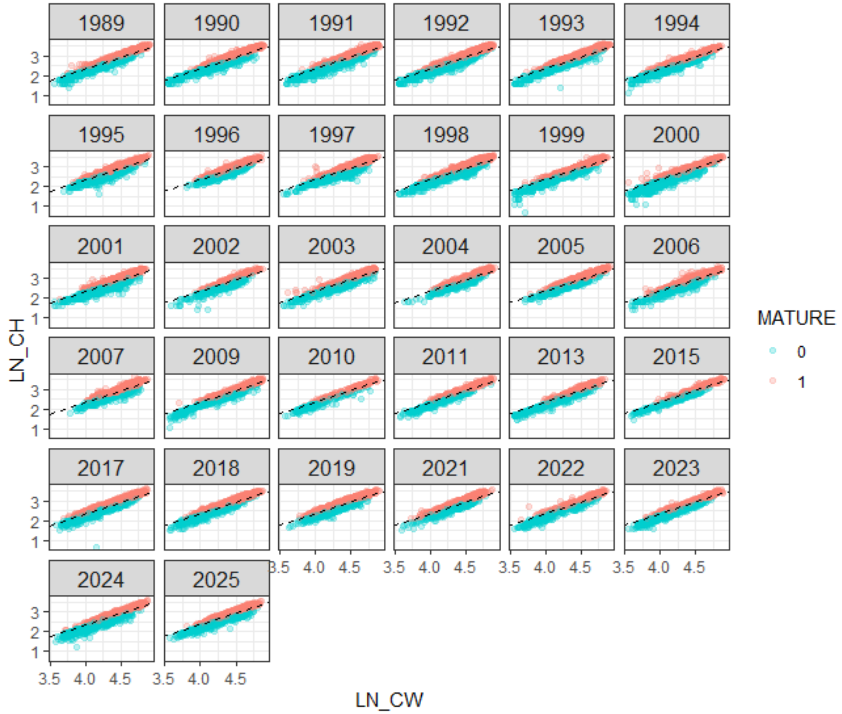
1 Measure carapace widths for crab haul subsample

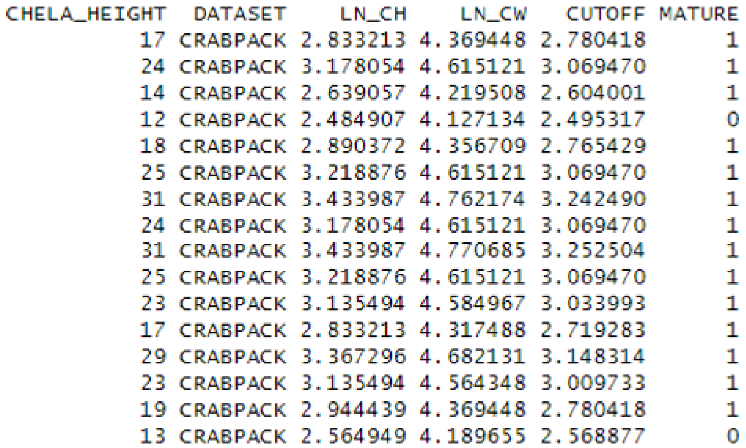
CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

3 Generate ln(CH) ~ ln(CW) maturity cutline

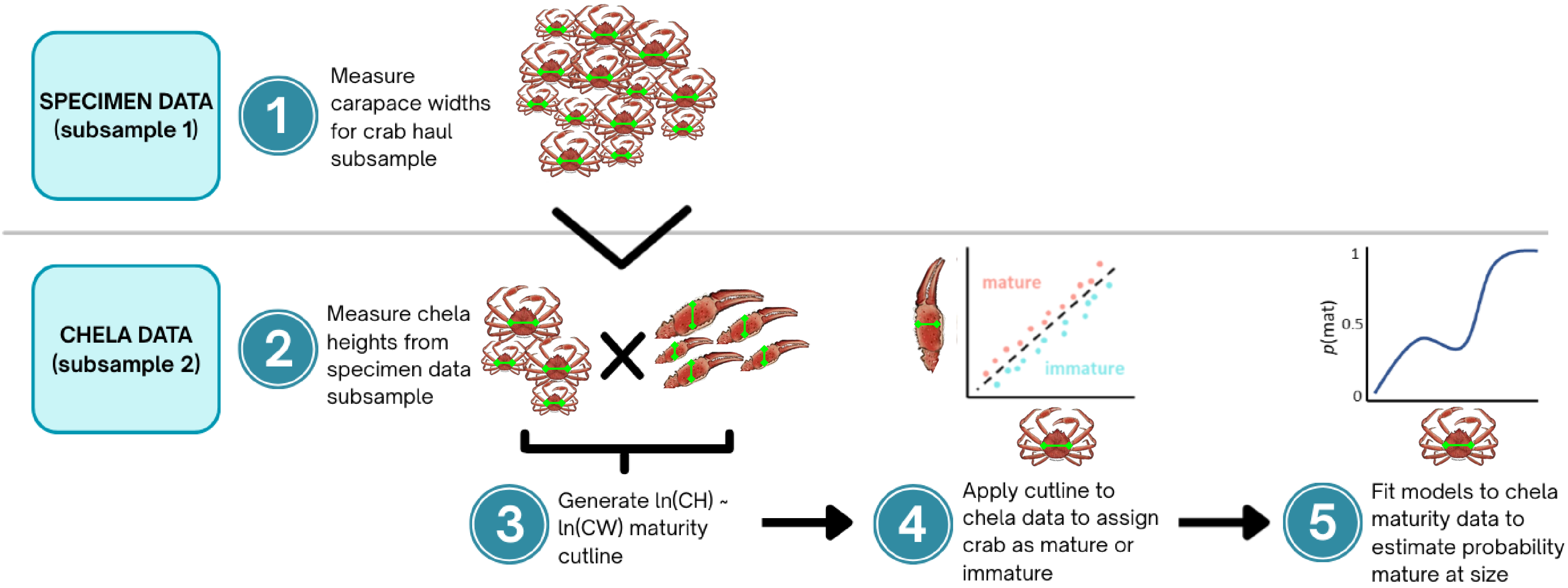
4 Apply cutline to chela data to assign crab as mature or immature





```
CHELA_HEIGHT  DATASET    LN_CH    LN_CW   CUTOFF MATURE
          17 CRABPACK 2.833213 4.369448 2.780418      1
          24 CRABPACK 3.178054 4.615121 3.069470      1
          14 CRABPACK 2.639057 4.219508 2.604001      1
          12 CRABPACK 2.484907 4.127134 2.495317      0
          18 CRABPACK 2.890372 4.356709 2.765429      1
          25 CRABPACK 3.218876 4.615121 3.069470      1
          31 CRABPACK 3.433987 4.762174 3.242490      1
          24 CRABPACK 3.178054 4.615121 3.069470      1
          31 CRABPACK 3.433987 4.770685 3.252504      1
          25 CRABPACK 3.218876 4.615121 3.069470      1
          23 CRABPACK 3.135494 4.584967 3.033993      1
          17 CRABPACK 2.833213 4.317488 2.719283      1
          29 CRABPACK 3.367296 4.682131 3.148314      1
          23 CRABPACK 3.135494 4.564348 3.009733      1
          19 CRABPACK 2.944439 4.369448 2.780418      1
          13 CRABPACK 2.564949 4.189655 2.568877      0
```

5 Fitting models to chela data to estimate probability mature at size



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

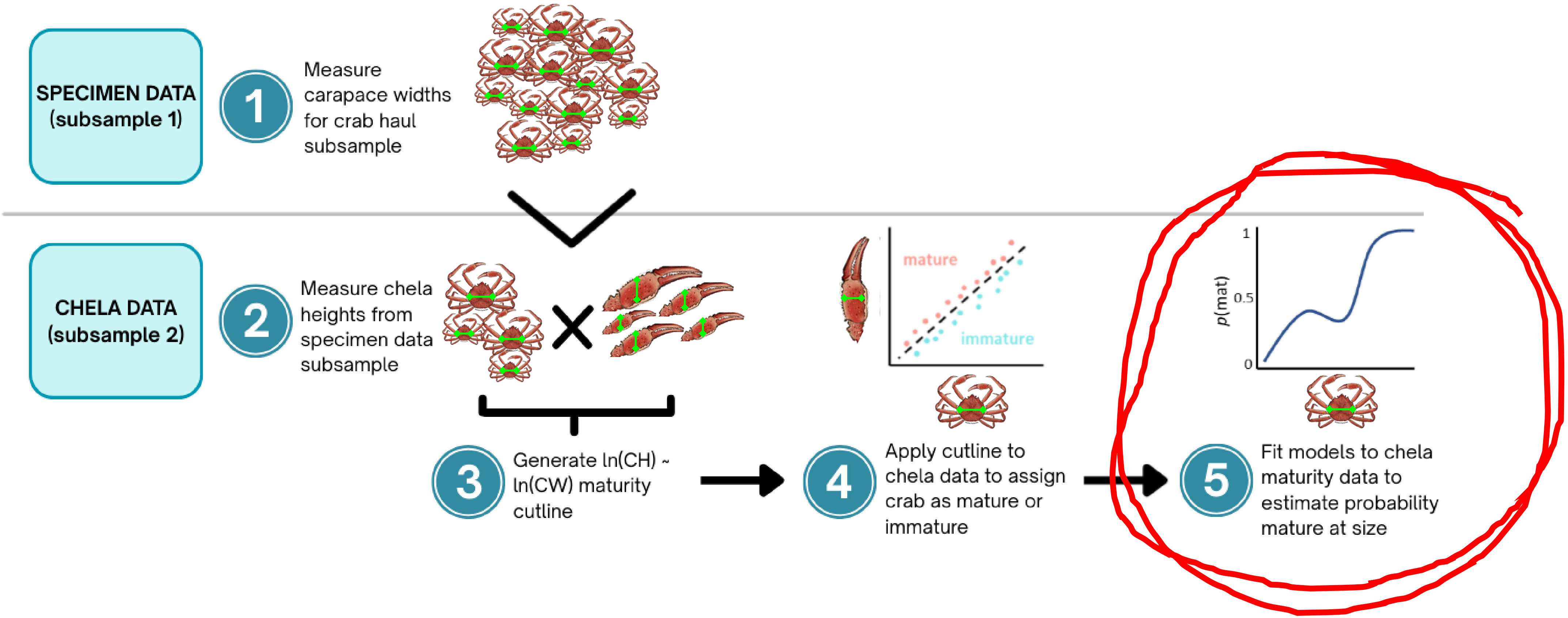
3 Generate ln(CH) ~ ln(CW) maturity cutline

4 Apply cutline to chela data to assign crab as mature or immature

5 Fit models to chela maturity data to estimate probability mature at size

5

Fitting models to chela data to estimate probability mature at size



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

3 Generate ln(CH) ~ ln(CW) maturity cutline

4 Apply cutline to chela data to assign crab as mature or immature

5 Fit models to chela maturity data to estimate probability mature at size

Legacy workflow

5

Fitting models to chela data to estimate probability mature at size

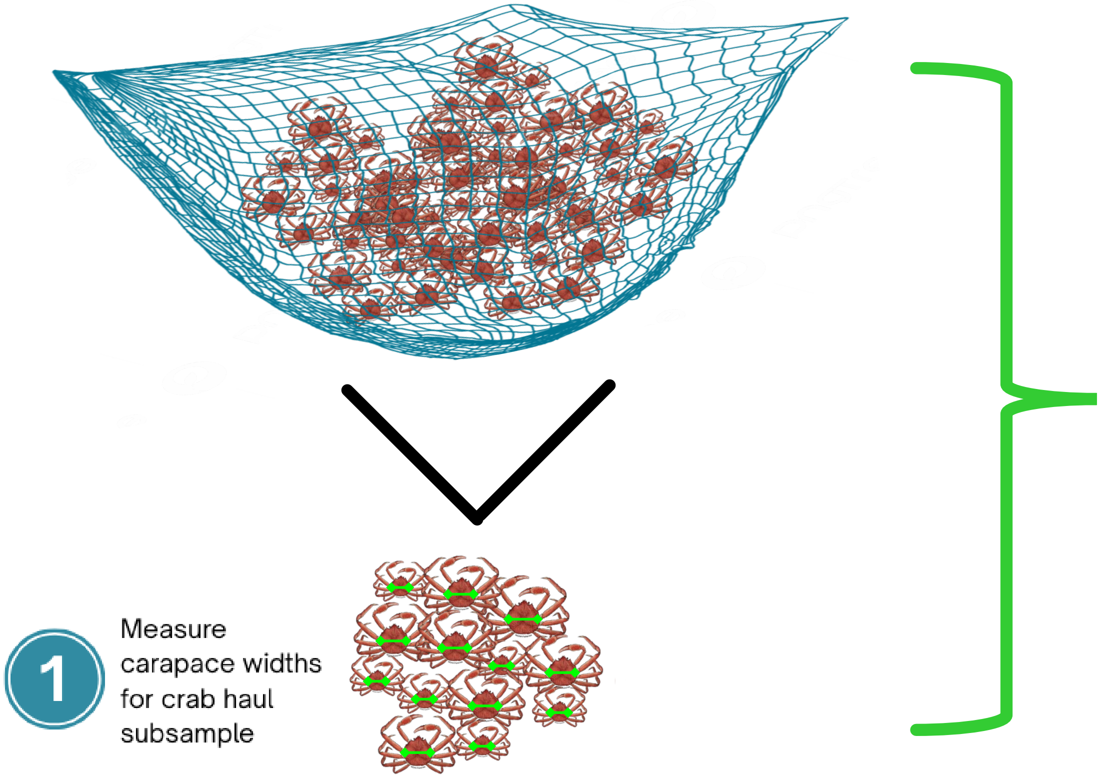
1. Bin chela data into 10mm bins
2. Calculate proportion mature within each bin using sampling factor

5

Fitting models to chela data to estimate probability mature at size

1. Bin chela data into 10mm bins
2. Calculate proportion mature within each bin using sampling factor

5 Fitting models to chela data to estimate probability mature at size

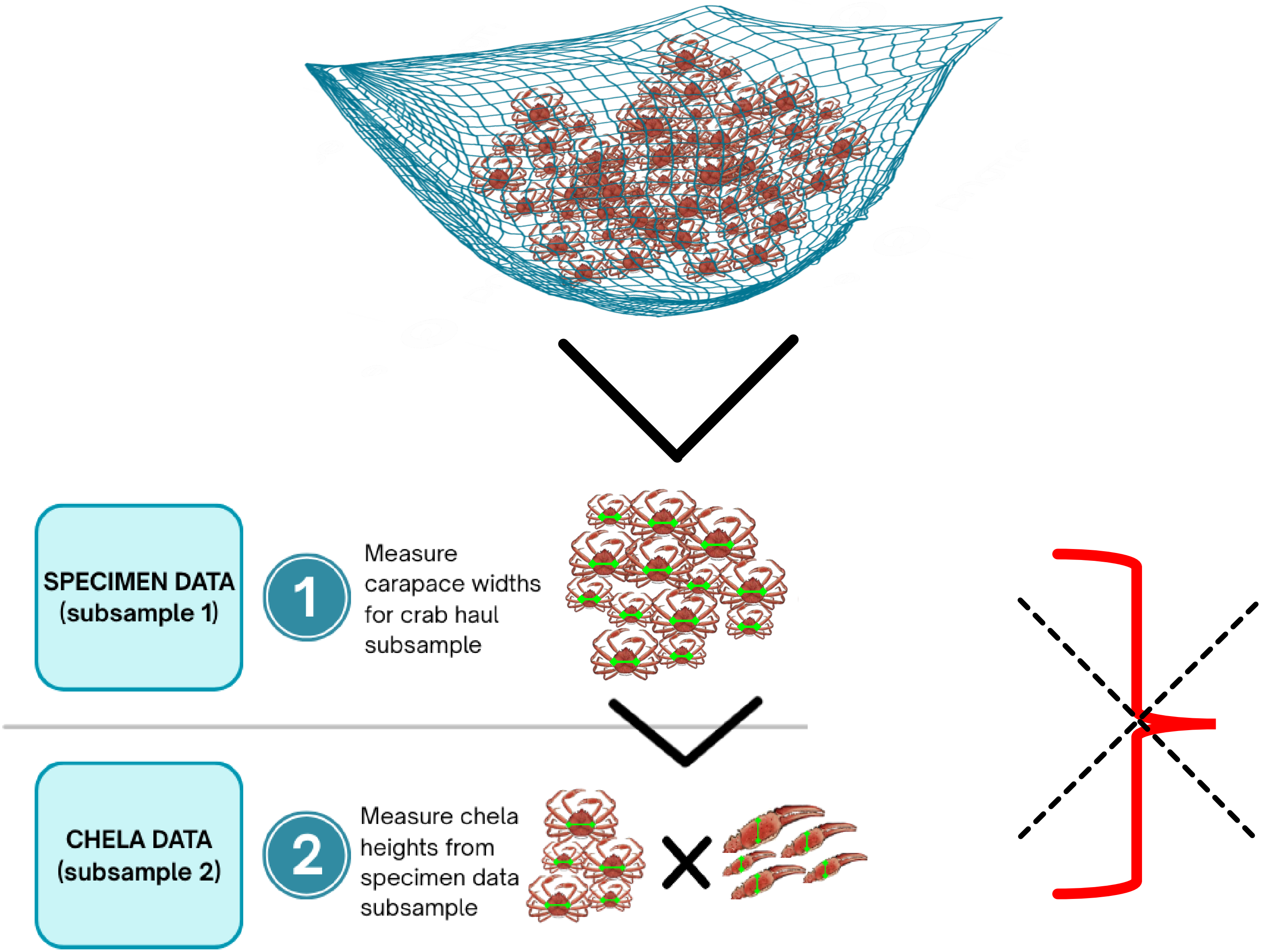


SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

Sampling factor = abundance at size in subsample 1 after accounting for subsampling intensity for subsample 1

5 Fitting models to chela data to estimate probability mature at size



SPECIMEN DATA (subsample 1)

① Measure carapace widths for crab haul subsample

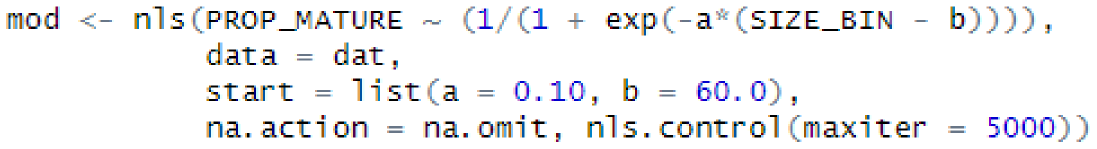
CHELA DATA (subsample 2)

② Measure chela heights from specimen data subsample

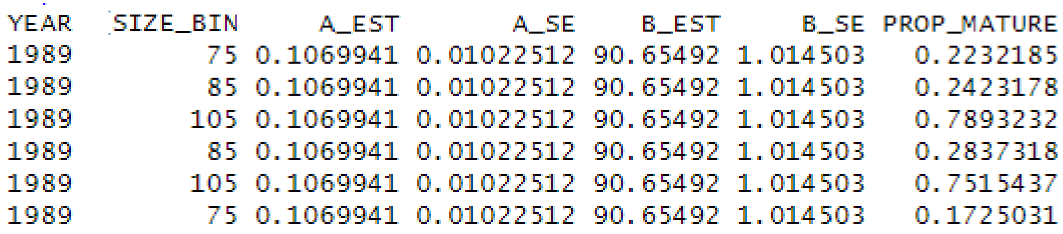
Sampling factor does not apply to subsampling for chela (subsample 2)

5 Fitting models to chela data to estimate probability mature at size

1. Bin chela data into 10mm bins
2. Calculate proportion mature within each bin using sampling factor
3. Fit logistic models to proportion mature by 10mm bin, separate by year



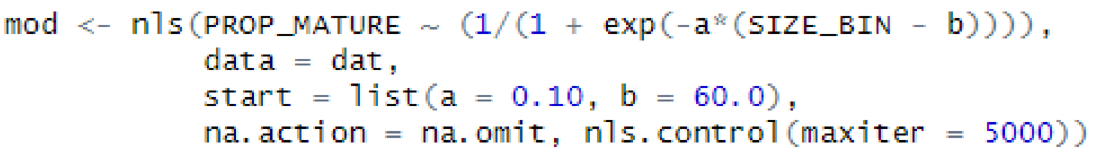
```
mod <- nls(PROP_MATURE ~ (1/(1 + exp(-a*(SIZE_BIN - b)))),
           data = dat,
           start = list(a = 0.10, b = 60.0),
           na.action = na.omit, nls.control(maxiter = 5000))
```



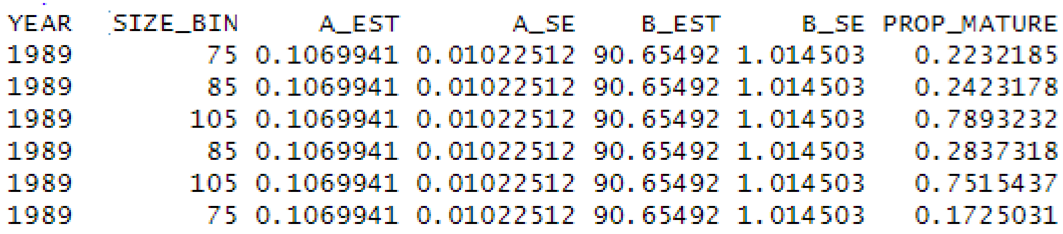
```
YEAR  SIZE_BIN     A_EST       A_SE    B_EST     B_SE PROP_MATURE
1989        75 0.1069941 0.01022512 90.65492 1.014503   0.2232185
1989        85 0.1069941 0.01022512 90.65492 1.014503   0.2423178
1989       105 0.1069941 0.01022512 90.65492 1.014503   0.7893232
1989        85 0.1069941 0.01022512 90.65492 1.014503   0.2837318
1989       105 0.1069941 0.01022512 90.65492 1.014503   0.7515437
1989        75 0.1069941 0.01022512 90.65492 1.014503   0.1725031
```

5 Fitting models to chela data to estimate probability mature at size

1. Bin chela data into 10mm bins
2. Calculate proportion mature within each bin using sampling factor
3. Fit logistic models to proportion mature by 10mm bin, separate by year
    - Does not account for uncertainty due to bin sample sizes
    - Does not account for spatiotemporal variability

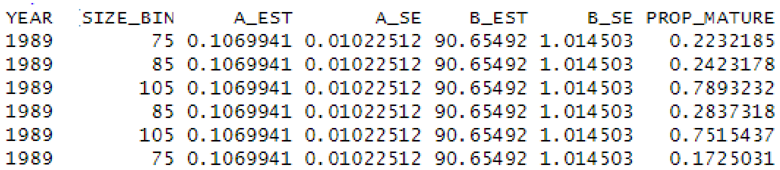


```
mod <- nls(PROP_MATURE ~ (1/(1 + exp(-a*(SIZE_BIN - b)))),
           data = dat,
           start = list(a = 0.10, b = 60.0),
           na.action = na.omit, nls.control(maxiter = 5000))
```

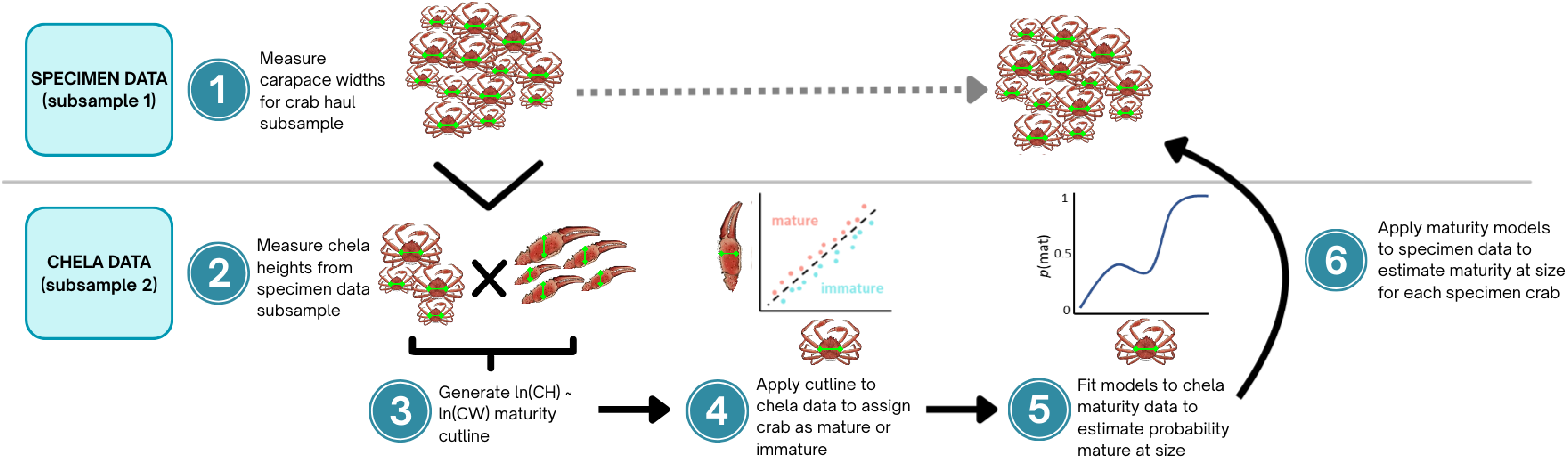


```
YEAR  SIZE_BIN     A_EST       A_SE    B_EST     B_SE PROP_MATURE
1989        75 0.1069941 0.01022512 90.65492 1.014503   0.2232185
1989        85 0.1069941 0.01022512 90.65492 1.014503   0.2423178
1989       105 0.1069941 0.01022512 90.65492 1.014503   0.7893232
1989        85 0.1069941 0.01022512 90.65492 1.014503   0.2837318
1989       105 0.1069941 0.01022512 90.65492 1.014503   0.7515437
1989        75 0.1069941 0.01022512 90.65492 1.014503   0.1725031
```

6 Use models to predict probability maturity at size for all specimen crab



```
YEAR  SIZE_BIN     A_EST       A_SE    B_EST     B_SE PROP_MATURE
1989        75 0.1069941 0.01022512 90.65492 1.014503   0.2232185
1989        85 0.1069941 0.01022512 90.65492 1.014503   0.2423178
1989       105 0.1069941 0.01022512 90.65492 1.014503   0.7893232
1989        85 0.1069941 0.01022512 90.65492 1.014503   0.2837318
1989       105 0.1069941 0.01022512 90.65492 1.014503   0.7515437
1989        75 0.1069941 0.01022512 90.65492 1.014503   0.1725031
```



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

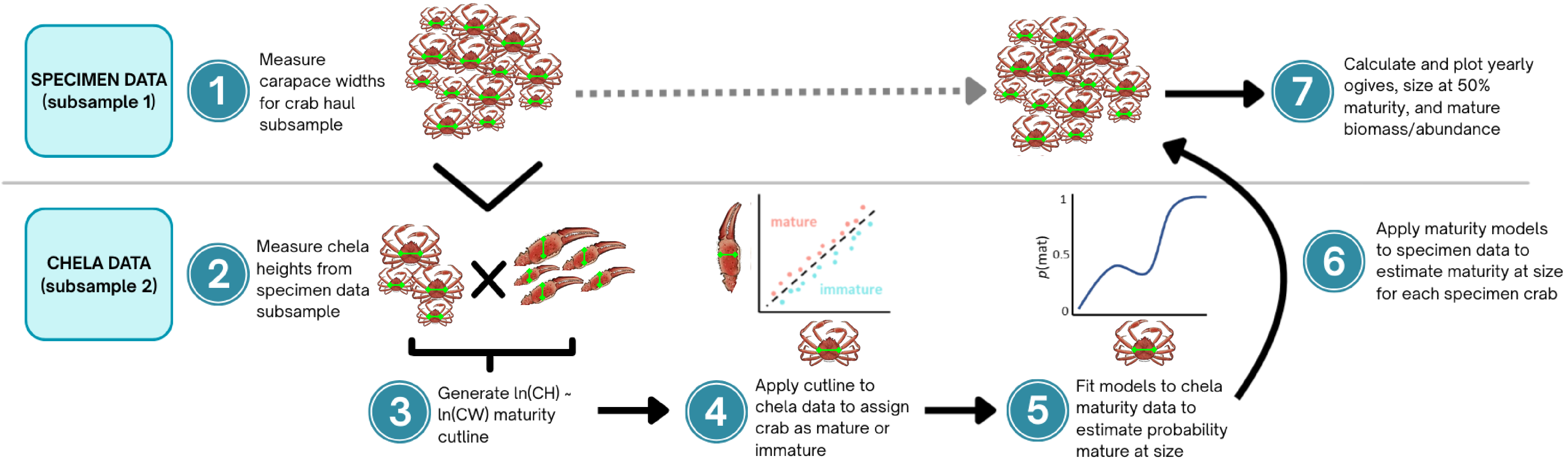
3 Generate ln(CH) ~ ln(CW) maturity cutline

4 Apply cutline to chela data to assign crab as mature or immature

5 Fit models to chela maturity data to estimate probability mature at size

6 Apply maturity models to specimen data to estimate maturity at size for each specimen crab

7 Calculate ogives, SAM, and mature biomass/abundance



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

7 Calculate and plot yearly ogives, size at 50% maturity, and mature biomass/abundance

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

3 Generate ln(CH) ~ ln(CW) maturity cutline

4 Apply cutline to chela data to assign crab as mature or immature

5 Fit models to chela maturity data to estimate probability mature at size

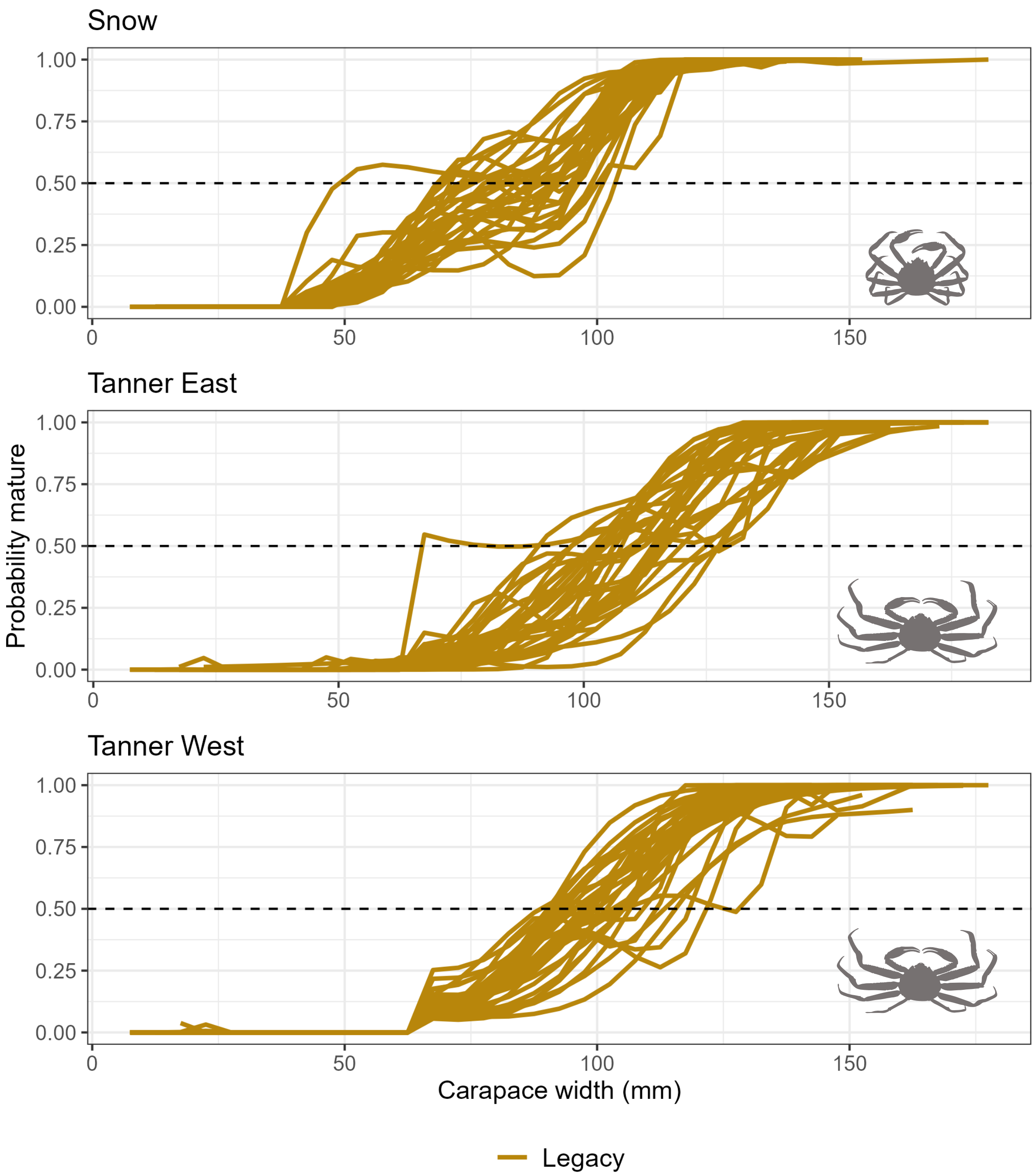
6 Apply maturity models to specimen data to estimate maturity at size for each specimen crab

7

Calculate ogives, SAM, and mature biomass/abundance

Ogives calculated by:

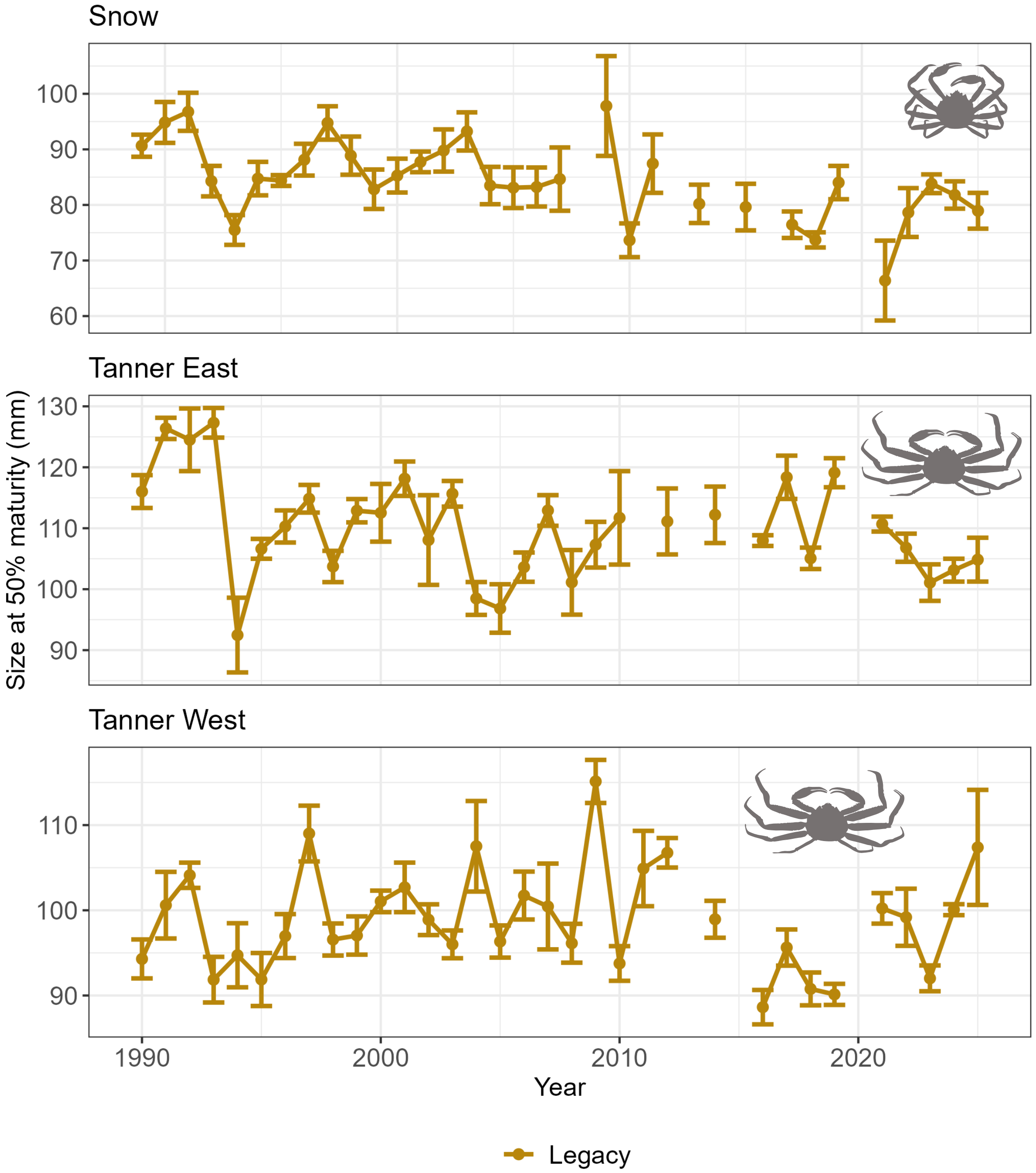
1. Computing the mean probability mature by 10mm size bin each year, interpolated to 5mm bins for comparison with updated workflow



7 Calculate ogives, SAM, and mature biomass/abundance

Size-at-50% maturity (SAM) calculated by:

1. Using yearly SAM parameter estimates from the models
2. Using yearly SAM standard error estimates from the models for uncertainty

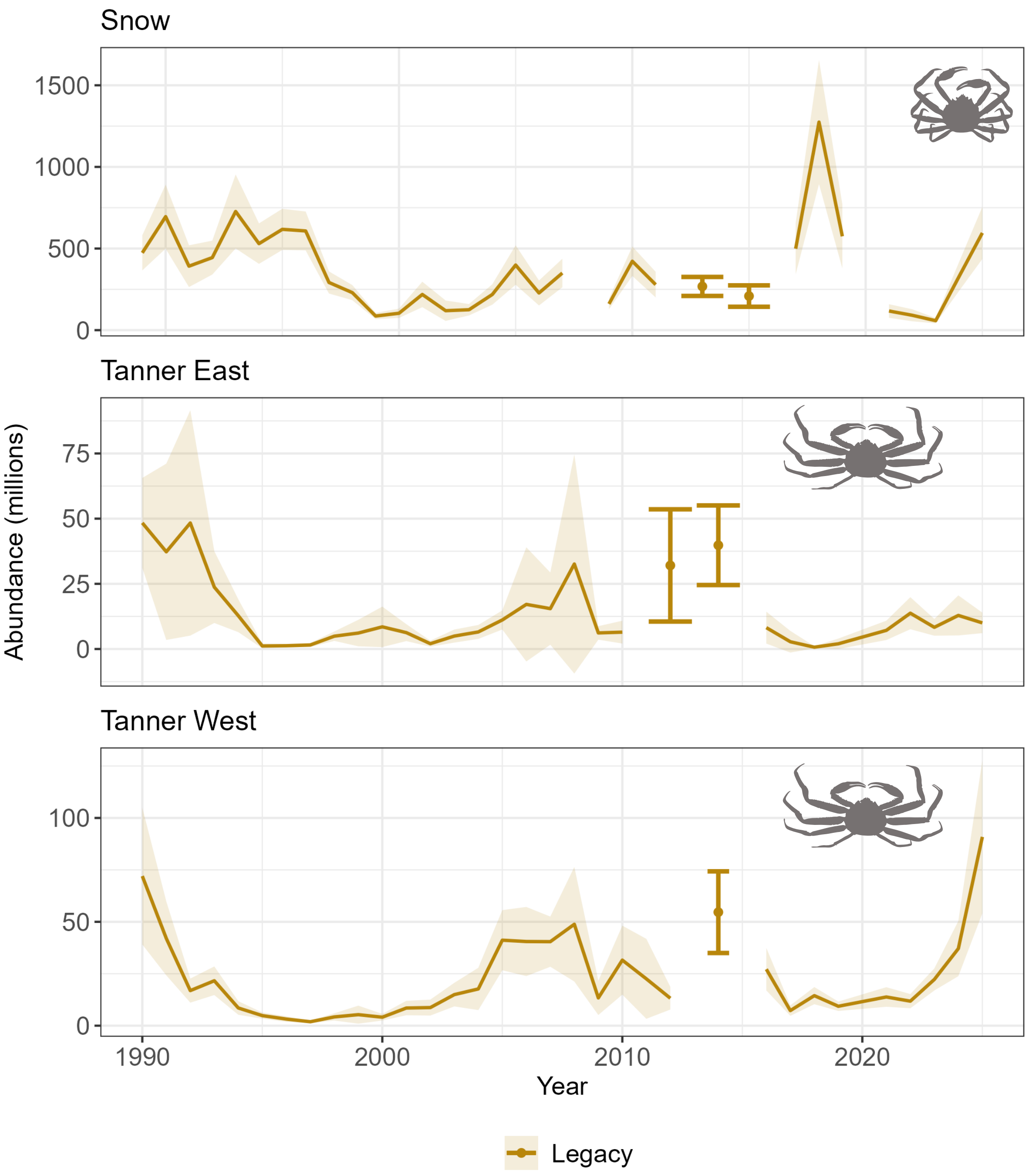


LEGACY

7 Calculate ogives, SAM, and mature biomass/abundance

Mature abundance/biomass calculated by:

1. Multiplying probability mature at size and abundance at size (sampling factor) to get mature abundance/biomass at size
2. Mature abundance/biomass at size is then expanded via standard design-based methods



LEGACY

Legacy workflow	**Proposed workflow**
Uses 10mm bins	Uses 5mm bins

Legacy workflow	**Proposed workflow**
Uses 10mm bins	Uses 5mm bins
Fits nls logistic models to chela data	Fits sdmTMB models to chela data

Legacy workflow	**Proposed workflow**
Uses 10mm bins	Uses 5mm bins
Fits nls logistic models to chela data	Fits sdmTMB models to chela data
Uses proportion mature as the response, calculated using sampling factor	Uses binomial (mature/immature) as response

Legacy workflow	**Proposed workflow**
Uses 10mm bins	Uses 5mm bins
Fits nls logistic models to chela data	Fits sdmTMB models to chela data
Uses proportion mature as the response, calculated using sampling factor	Uses binomial (mature/immature) as response
Assumes logistic maturity-at-size ogive shape	Flexibly models maturity-at-size via gam-style smooths

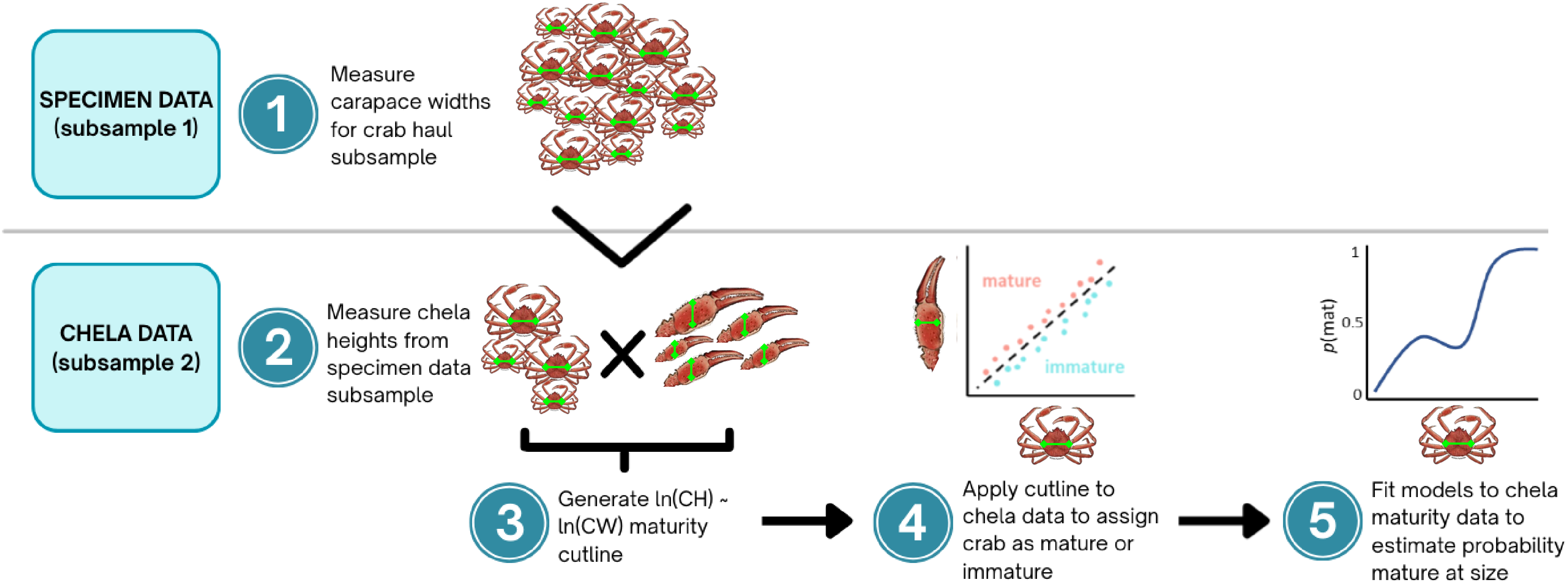
Legacy workflow	**Proposed workflow**
Uses 10mm bins	Uses 5mm bins
Fits nls logistic models to chela data	Fits sdmTMB models to chela data
Uses proportion mature as the response, calculated using sampling factor	Uses binomial (mature/immature) as response
Assumes logistic maturity-at-size ogive shape	Flexibly models maturity-at-size via gam-style smooths
Does not account for uncertainty in maturity due to uneven sampling across space, time, and size bins	Robust to spatiotemporal gaps and variability in maturity through random fields; uncertainty is also tied to sample sizes via binomial structure and propagated across all workflow steps

Legacy workflow	**Proposed workflow**
Uses 10mm bins	Uses 5mm bins
Fits nls logistic models to chela data	Fits sdmTMB models to chela data
Uses proportion mature as the response, calculated using sampling factor	Uses binomial (mature/immature) as response
Assumes logistic maturity-at-size ogive shape	Flexibly models maturity-at-size via gam-style smooths
Does not account for uncertainty in maturity due to uneven sampling across space, time, and size bins	Robust to spatiotemporal gaps and variability in maturity through random fields; uncertainty is also tied to sample sizes via binomial structure and propagated across all workflow steps
Models not vetted through diagnostics(?)	Models vetted through diagnostics

Legacy workflow	**Proposed workflow**
Uses 10mm bins	Uses 5mm bins
Fits nls logistic models to chela data	Fits sdmTMB models to chela data
Uses proportion mature as the response, calculated using sampling factor	Uses binomial (mature/immature) as response
Assumes logistic maturity-at-size ogive shape	Flexibly models maturity-at-size via gam-style smooths
Does not account for uncertainty in maturity due to uneven sampling across space, time, and size bins	Robust to spatiotemporal gaps and variability in maturity through random fields; uncertainty is also tied to sample sizes via binomial structure and propagated across all workflow steps
Models not vetted through diagnostics(?)	Models vetted through diagnostics
SAM and ogives do not account for population size	SAM and ogives are weighted by abundance to account for population size

Proposed workflow

Steps 1-4 the same as the legacy workflow



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

3 Generate ln(CH) ~ ln(CW) maturity cutline

4 Apply cutline to chela data to assign crab as mature or immature

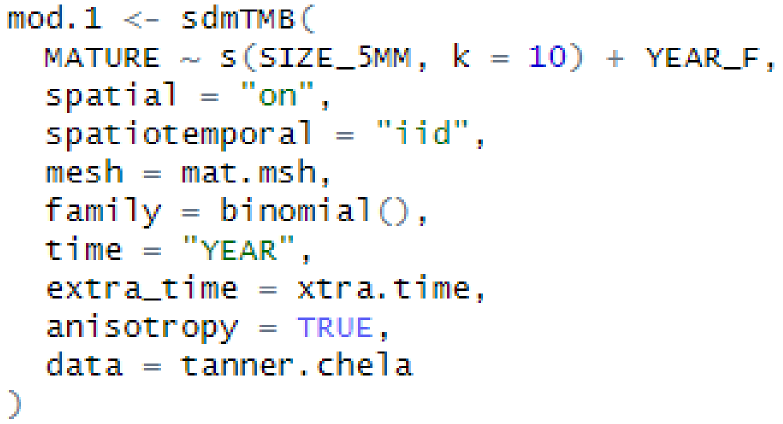
5 Fit models to chela maturity data to estimate probability mature at size

PROPOSED

5

Fitting models to chela data to estimate probability mature at size

1. Bin chela data into 5mm bins
2. Fit sdmTMB models to binomial (mature/immature) individual crab data by 5mm size bin
    - Binomial: bins with more samples carry more weight in the likelihood
    - Spatial ="on": captures persistent space-only deviations
    - Spatiotemporal = "iid": captures year-specific spatial deviations, though no correlation across years ("ar1")
    - Uneven sampling accounted for in random fields
    - Anisotropy in spatial deviations
    - Also spatial_varying() and time_varying() arguments
3. Evaluate various model parameterizations using sanity(), k-fold cv, and DHARMa residuals

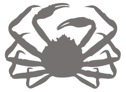


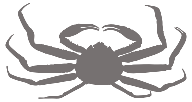
```
mod.1 <- sdmTMB(
  MATURE ~ s(SIZE_5MM, k = 10) + YEAR_F,
  spatial = "on",
  spatiotemporal = "iid",
  mesh = mat.msh,
  family = binomial(),
  time = "YEAR",
  extra_time = xtra.time,
  anisotropy = TRUE,
  data = tanner.chela
)
```

PROPOSED

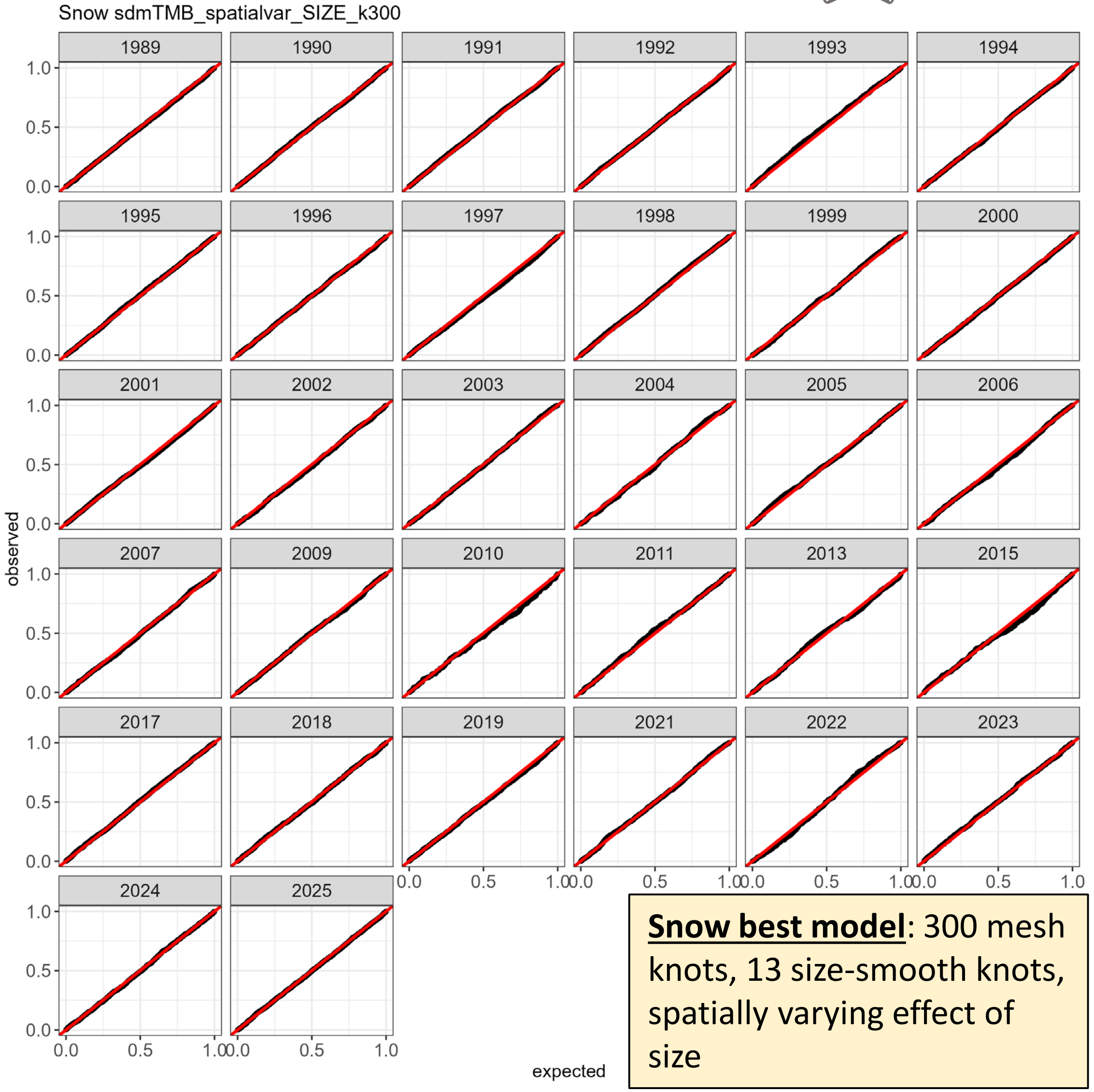
5

Fitting models to chela data to estimate probability mature at size

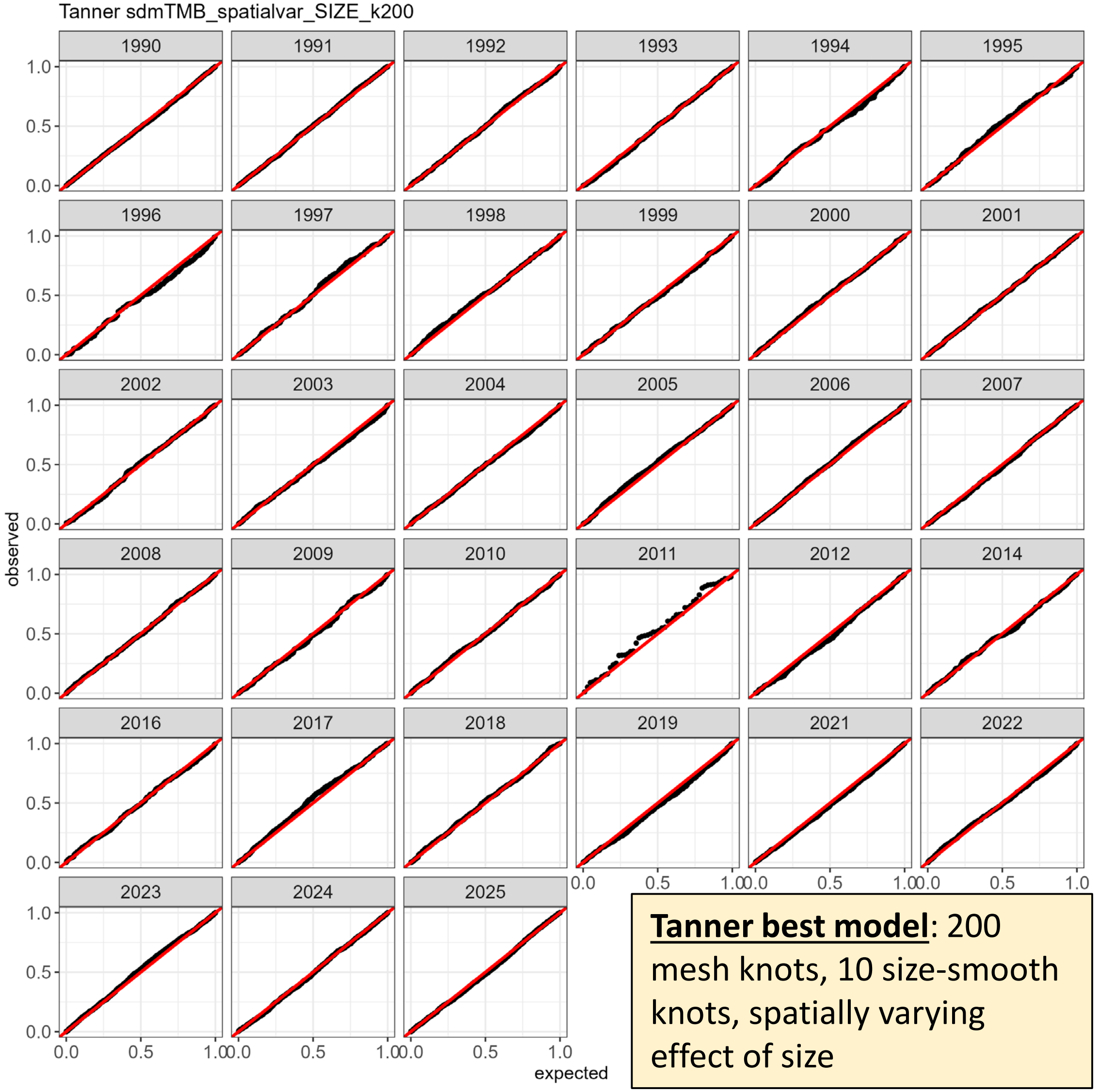




*see document for more diagnostics



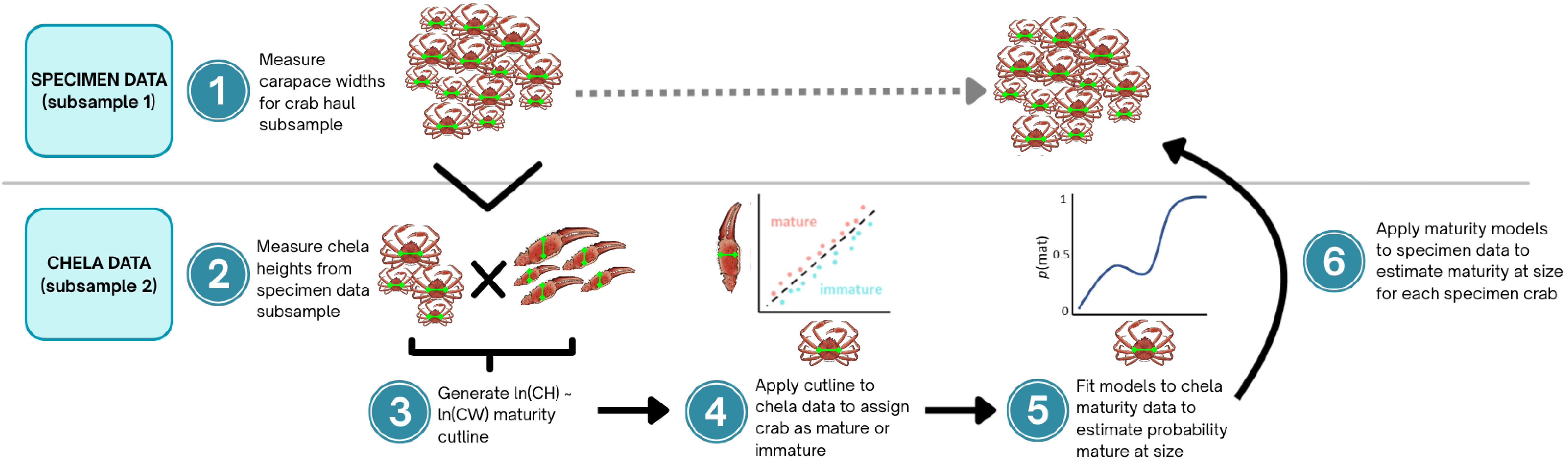
Snow best model: 300 mesh knots, 13 size-smooth knots, spatially varying effect of size



Tanner best model: 200 mesh knots, 10 size-smooth knots, spatially varying effect of size

PROPOSED

6 Use models to predict probability maturity at size for all specimen crab



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

3 Generate ln(CH) ~ ln(CW) maturity cutline

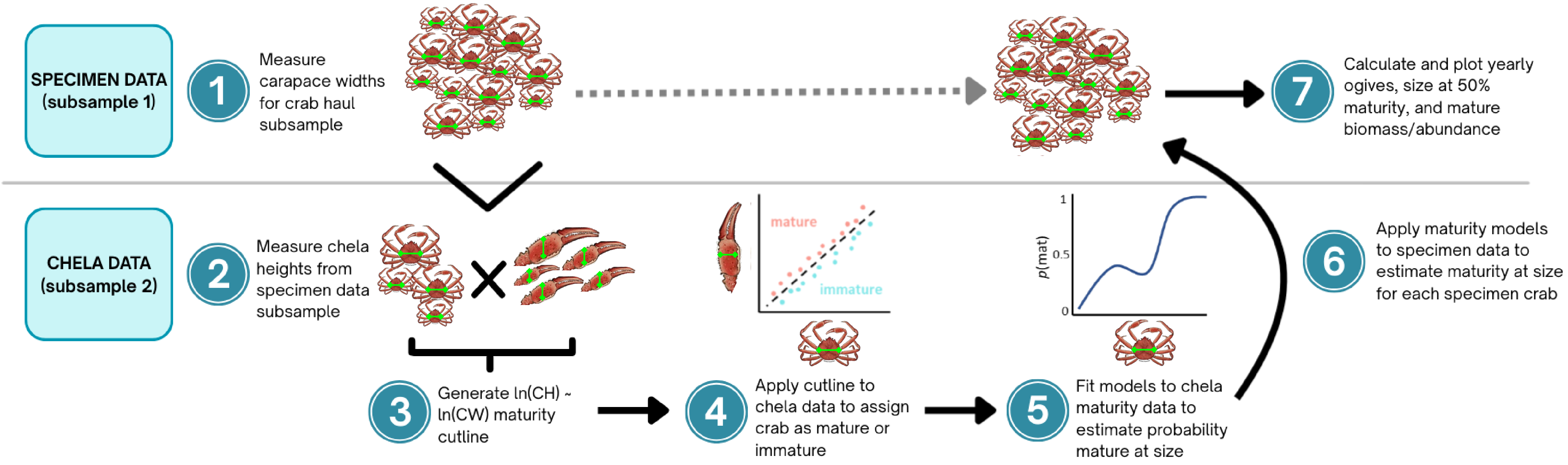
4 Apply cutline to chela data to assign crab as mature or immature

5 Fit models to chela maturity data to estimate probability mature at size

6 Apply maturity models to specimen data to estimate maturity at size for each specimen crab

PROPOSED

7 Calculate ogives, SAM, and mature biomass/abundance



SPECIMEN DATA (subsample 1)

1 Measure carapace widths for crab haul subsample

CHELA DATA (subsample 2)

2 Measure chela heights from specimen data subsample

3 Generate ln(CH) ~ ln(CW) maturity cutline

4 Apply cutline to chela data to assign crab as mature or immature

5 Fit models to chela maturity data to estimate probability mature at size

6 Apply maturity models to specimen data to estimate maturity at size for each specimen crab

7 Calculate and plot yearly ogives, size at 50% maturity, and mature biomass/abundance

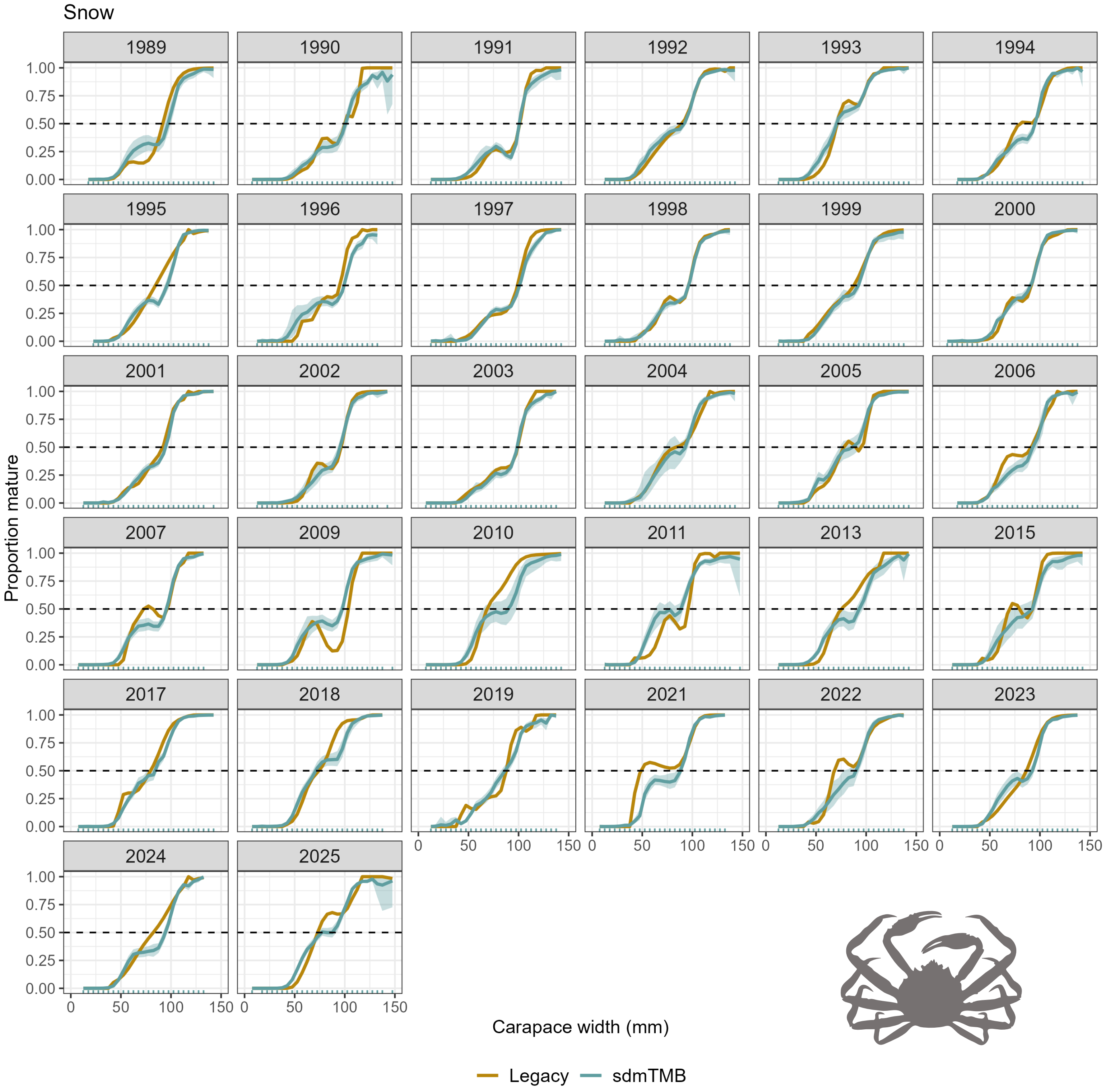
PROPOSED

7

Calculate ogives, SAM, and mature biomass/abundance

Ogives calculated by:

1. Simulating 500 draws from each model’s joint precision matrix
2. Computing the mean probability mature at size across draws each year, weighted by abundance at size
3. Quantifying uncertainty by summarizing ogive variation across draws



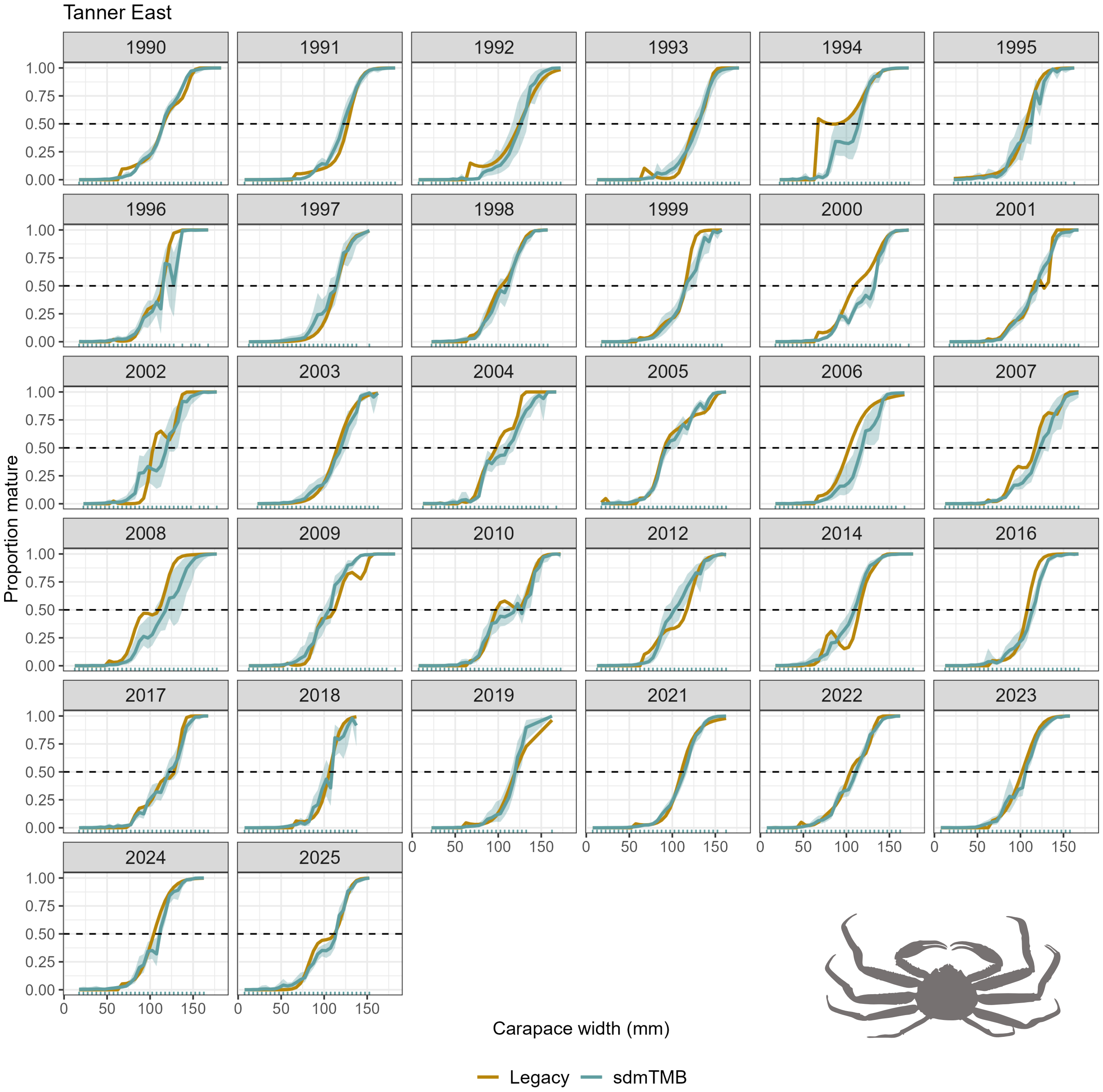
PROPOSED

7

Calculate ogives, SAM, and mature biomass/abundance

Ogives calculated by:

1. Simulating 500 draws from each model's joint precision matrix
2. Computing the mean probability mature at size across draws each year, weighted by abundance at size
3. Quantifying uncertainty by summarizing ogive variation across draws



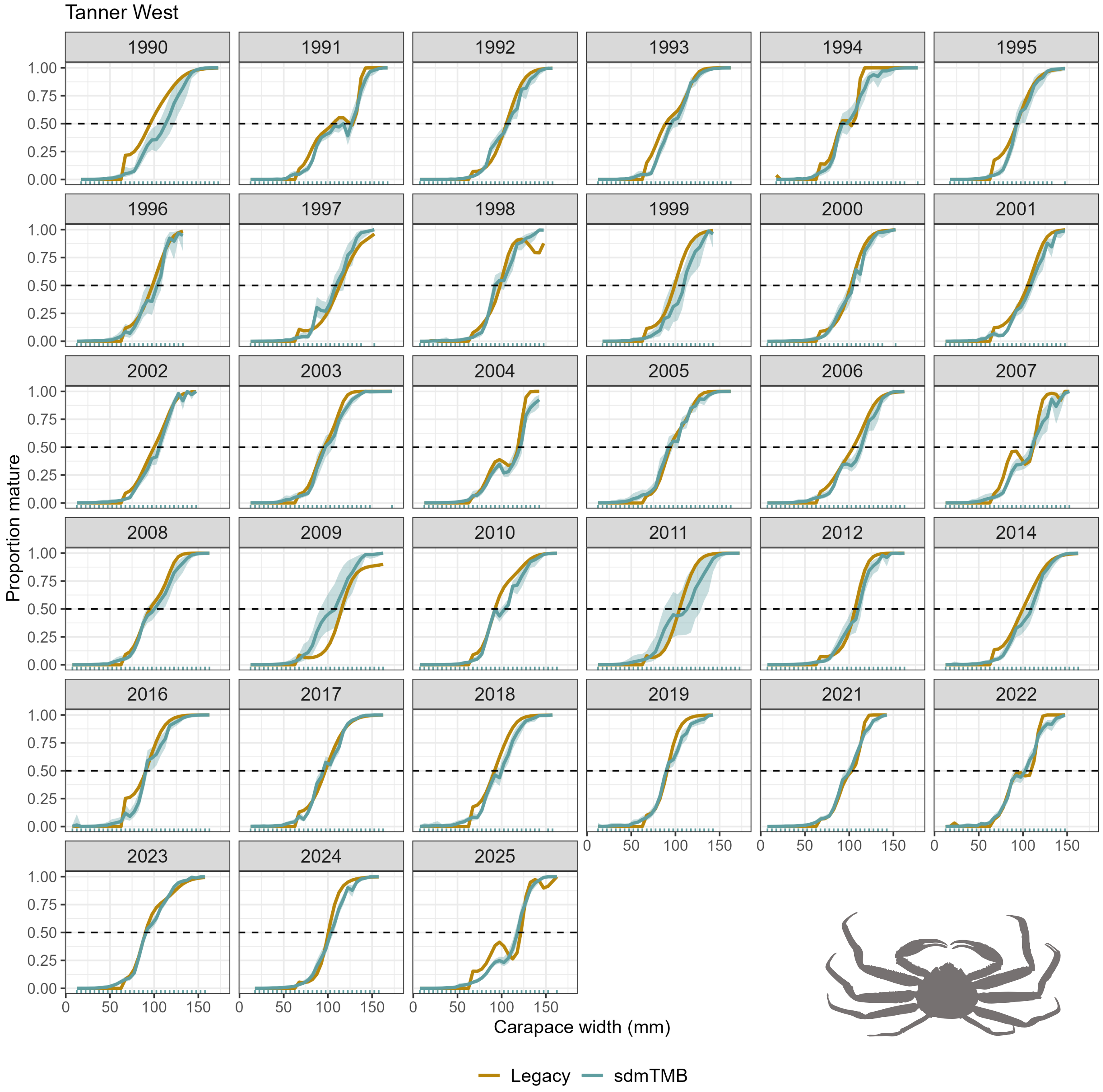
PROPOSED

7

Calculate ogives, SAM, and mature biomass/abundance

Ogives calculated by:

1. Simulating 500 draws from each model's joint precision matrix
2. Computing the mean probability mature at size across draws each year, weighted by abundance at size
3. Quantifying uncertainty by summarizing ogive variation across draws



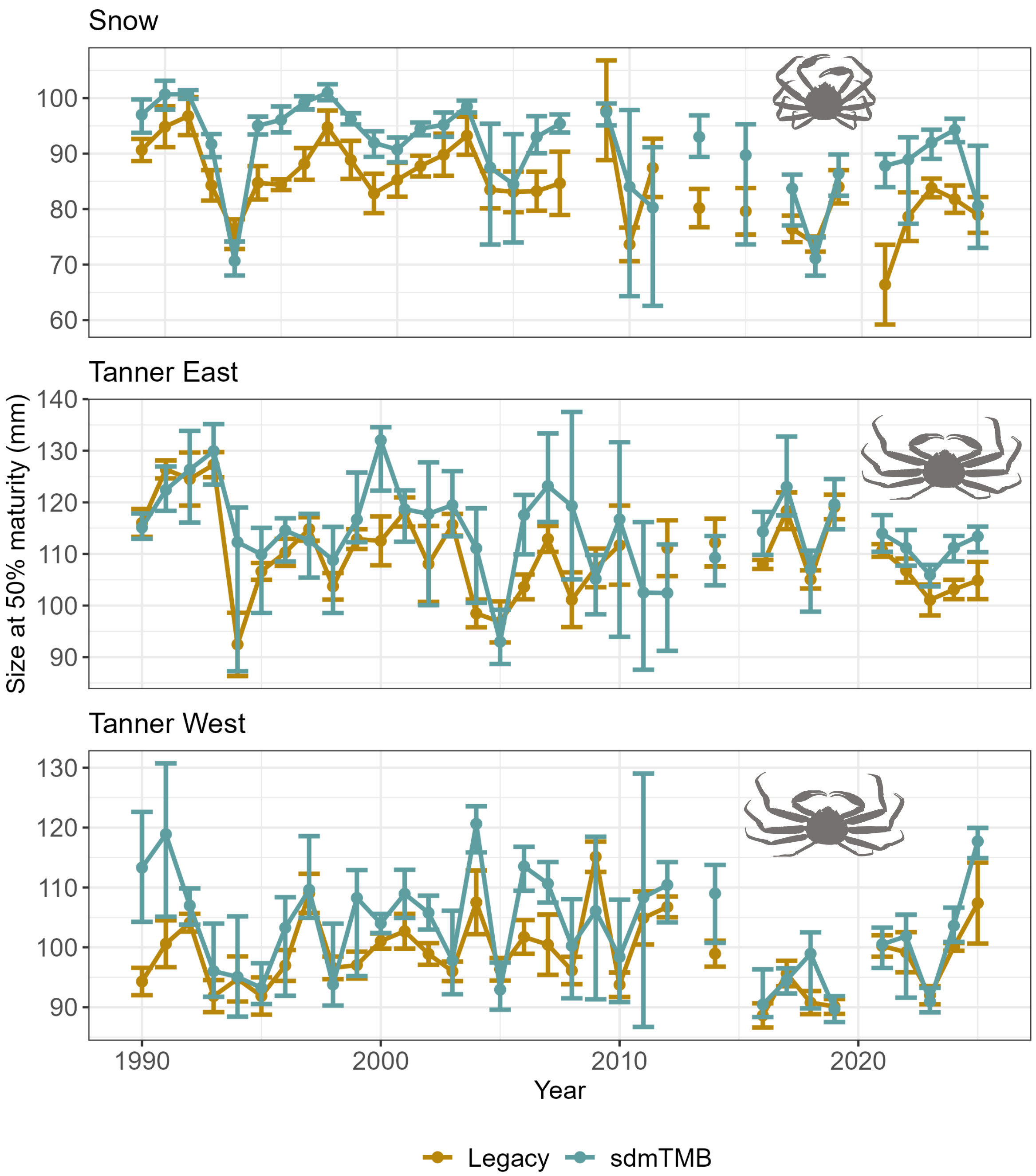
PROPOSED

7

Calculate ogives, SAM, and mature biomass/abundance

SAM calculated by:

1. Using weighted probability mature to calculate the size at which each yearly ogive first crosses 50%, averaged across the 500 draws
2. Quantifying uncertainty by summarizing SAM variation across draws

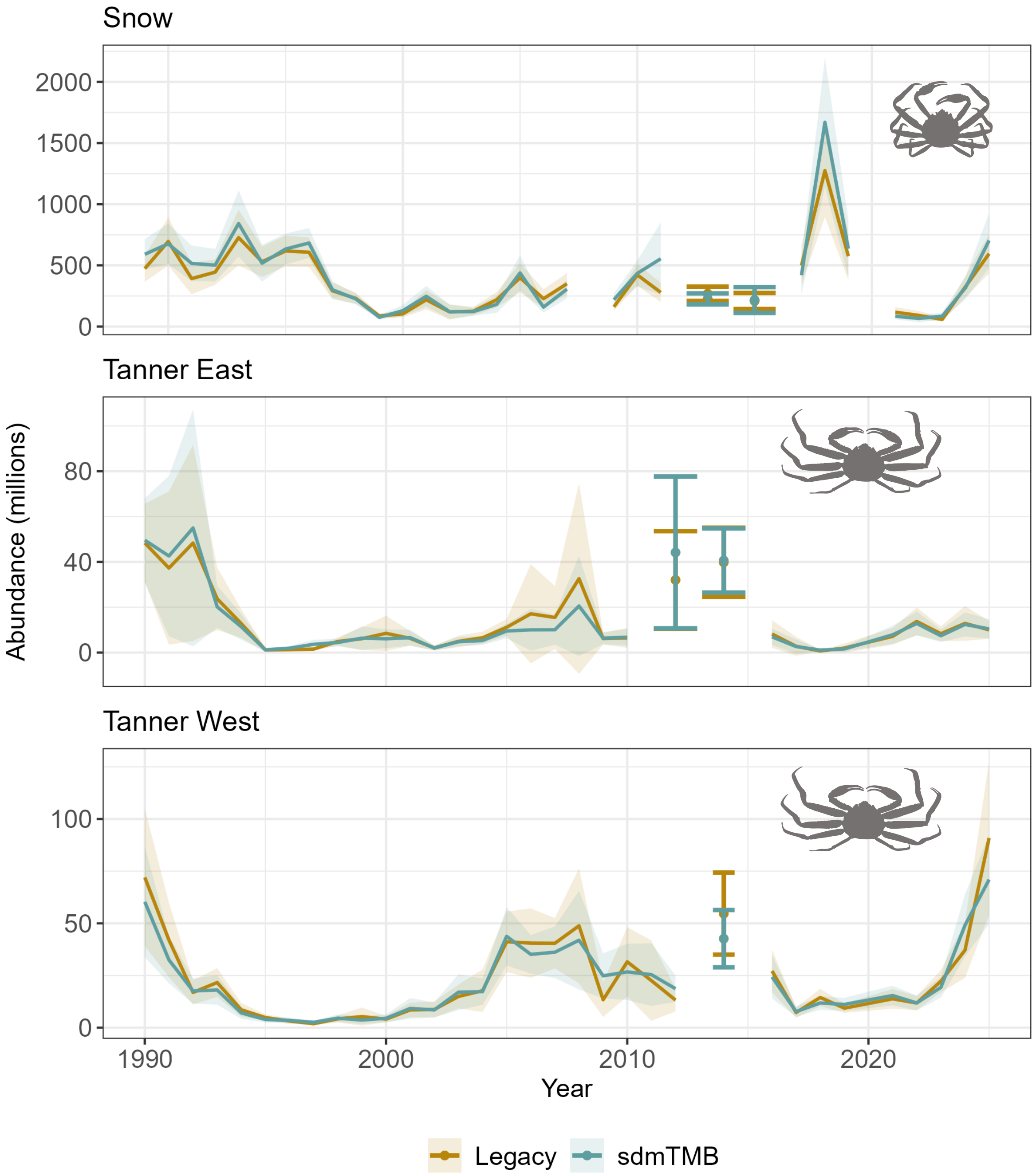


PROPOSED

7 Calculate ogives, SAM, and mature biomass/abundance

Mature abundance/biomass calculated by:

1. Expanding mature abundance/biomass at size via standard design-based methods for each draw, then averaged across the 500 draws
2. Quantifying uncertainty as model variation across draws combined with design-based expansion uncertainty



Conclusions

Accounting for spatiotemporal variation and multiple uncertainty sources in *Chionoecetes* maturity is essential for stock biology and reliable assessment inputs.

The updated workflow:

1. Removes the use of the sampling factor to model maturity in subsample two, eliminating inappropriate commingling of subsamples.
2. Uses sdmTMB to account for spatiotemporal variation, handles data gaps from uneven sampling, and enables future spatiotemporal maturity research.
3. Propagates uncertainty from model-based maturity estimates into downstream outputs.
4. Flexibly models maturity-at-size, strengthening confidence that non-logistic ogives represent biological signal via strong diagnostics