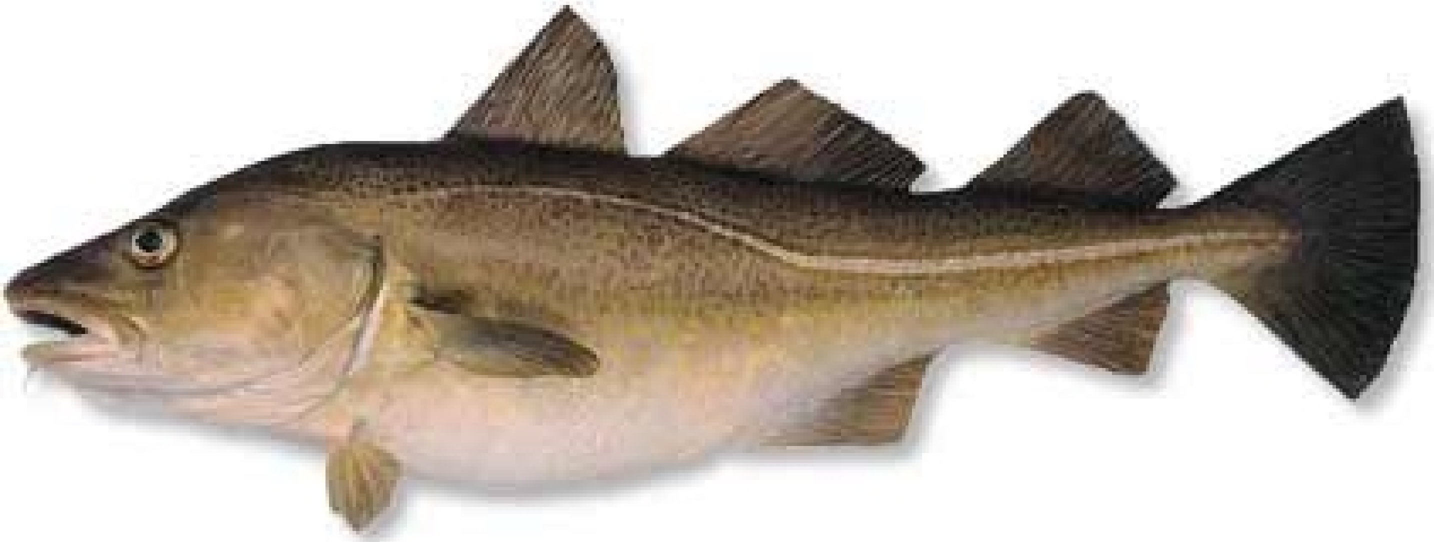
Norwegian Computing Center
APPLIED RESEARCH AND DEVELOPMENT

Estimating catch-at-age from market sampling data using a Bayesian hierarchical model

David Hirst[1], Sondre Aanes[2], Geir Storvik[1,3], Ragnar Bang Huseby[1] and Ingunn Fride Tvete[1]

1:Norwegian Computing Center, Oslo, Norway.
2:Institute for Marine Research, Bergen, Norway.
3:Department of Mathematics, University of Oslo

Norsk Regnesentral
ANVENDT DATAFORSKNING

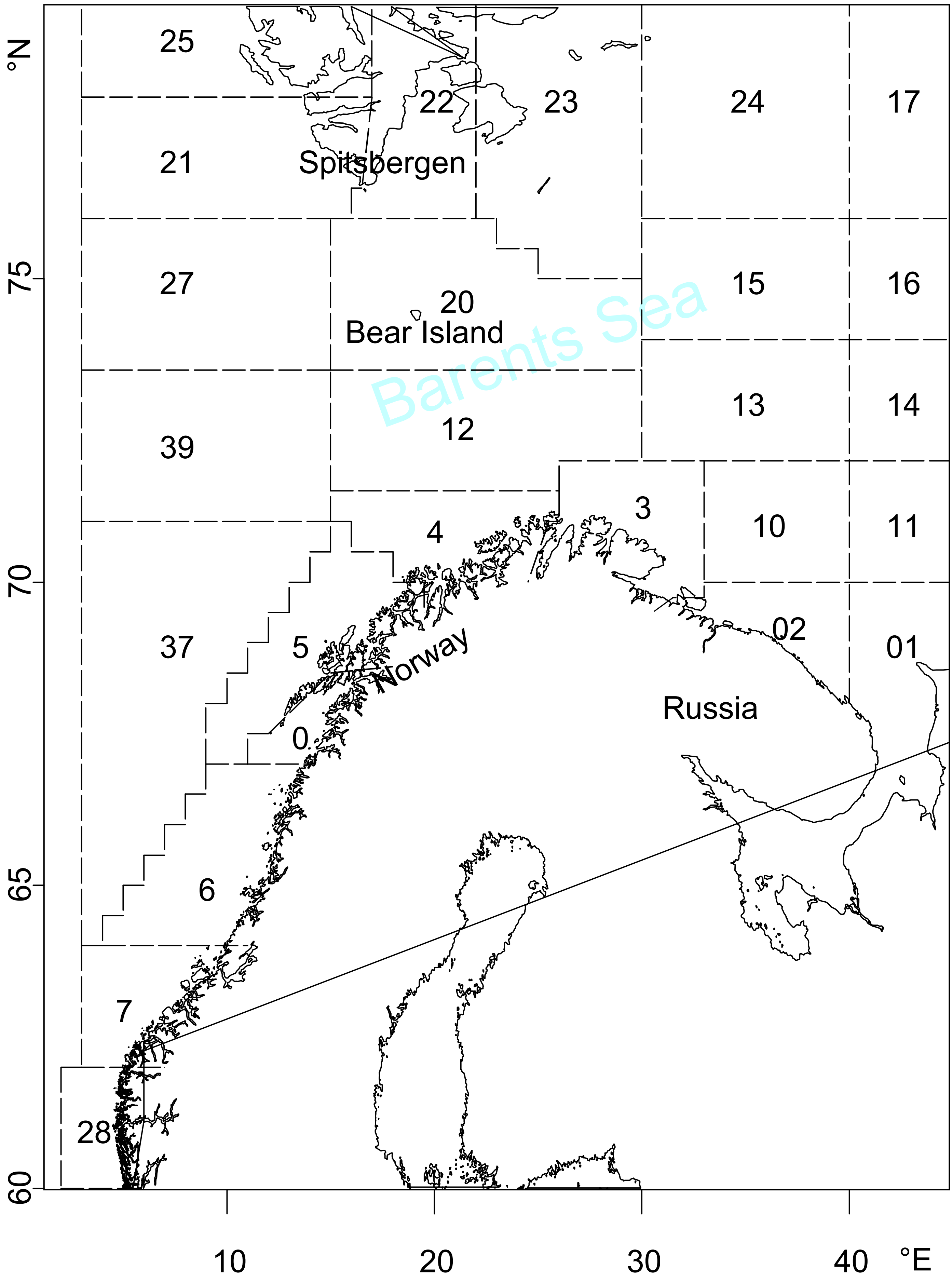


The Problem:

1) Estimate the catch at age of cod from very sparse sampling data

2) Provide a measure of the uncertainty

3) Do it quickly, reliably and repeatably



°N
75
70
65
60
10
20
30
40
°E
25
22
23
24
17
21
Spitsbergen
27
20
Bear Island
15
16
Barents Sea
39
12
13
14
3
4
10
11
37
5
Norway
02
01
0
Russia
6
7
28

The data:

Approximately 200 boats sampled every year,

80 fish sampled from each boat:
Age, weight and length measured.

16,000 fish per year!

But:

14 regions x 5 gears x 4 seasons = 280 ‘cells’

9 age classes

Large within boat correlation

Current Method:

1) Regard data as stratified sample, with many missing strata

2) Fill in missing strata manually

3) Find uncertainty by bootstrapping

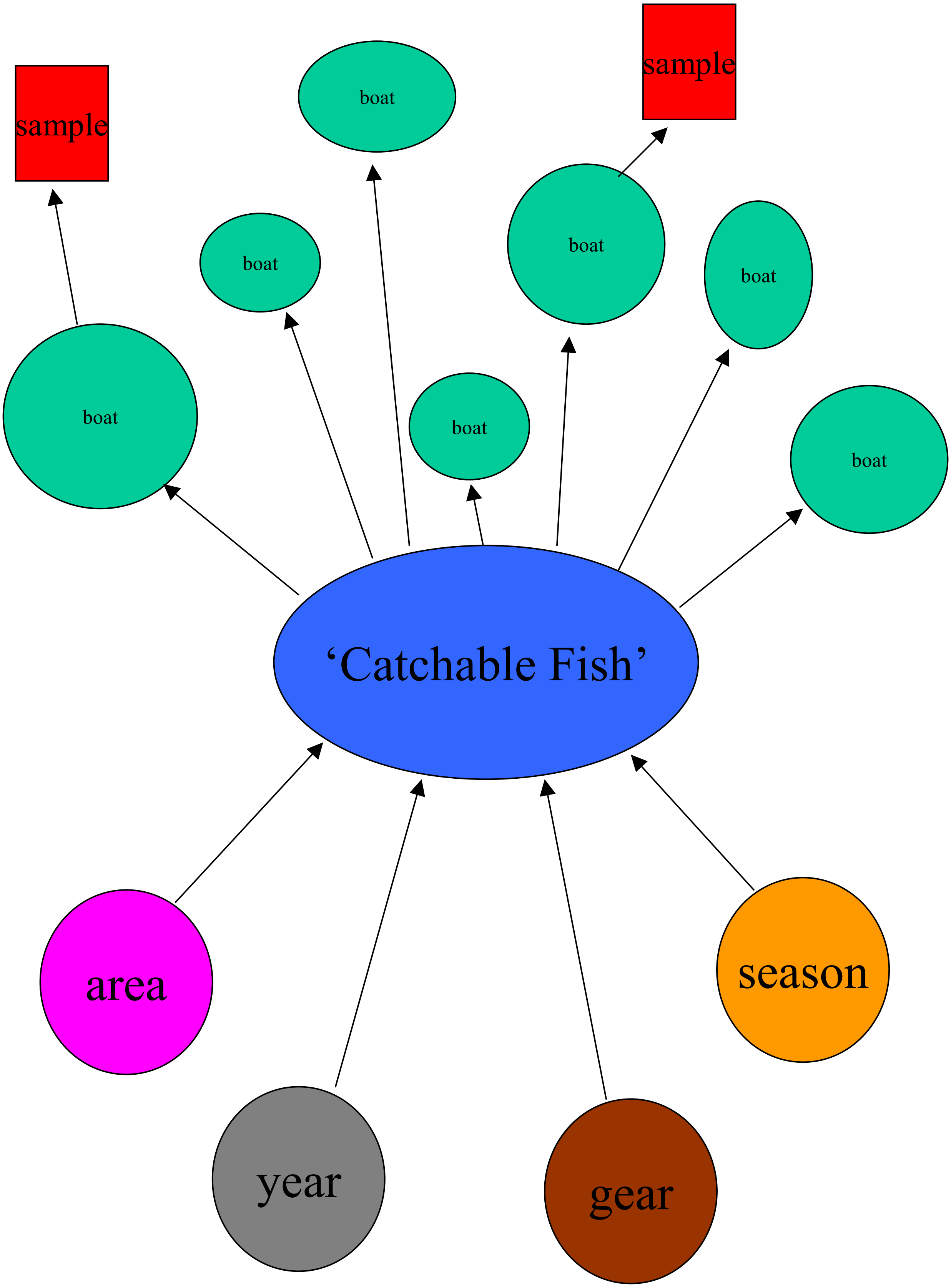
Bayesian Hierarchical Model

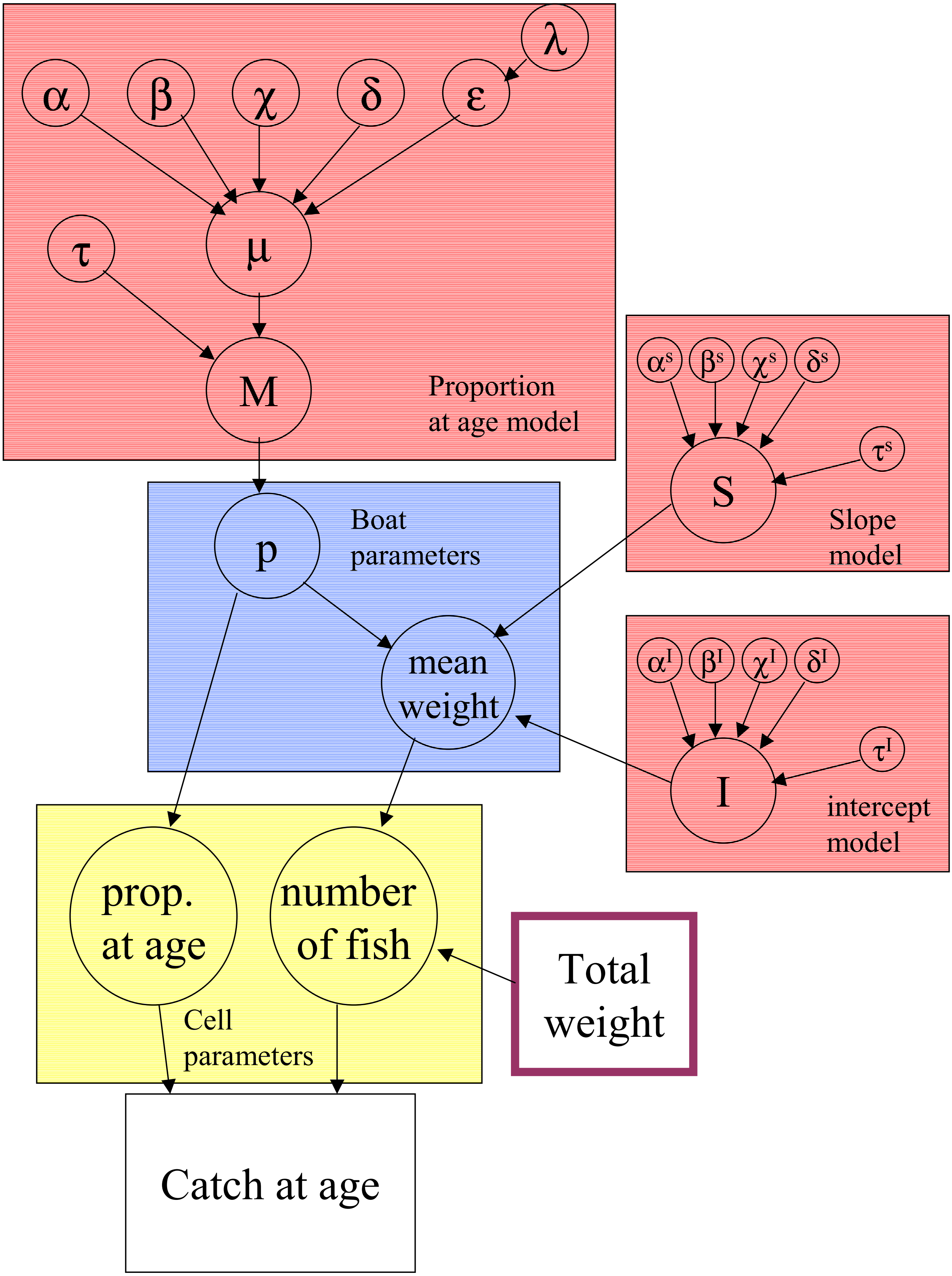
We model:

1) Proportion at age on individual boats

2) Weight given age on boats

3) Likelihood (including age-reading errors)


sample
boat
sample
boat
boat
boat
boat
boat
boat
boat
'Catchable Fish'
area
season
year
gear



λ
α
β
χ
δ
ε
τ
μ
M
Proportion
at age model
p
Boat
parameters
mean
weight
αs
βs
χs
δs
τs
S
Slope
model
αI
βI
χI
δI
τI
I
intercept
model
prop.
at age
number
of fish
Cell
parameters
Total
weight
Catch at age

Modelling proportion at age:

Let proportion at age *a* on boat *b* be $p_b(a)$.

reparameterise:

$$p_b(a) = \frac{\exp(\mathrm{M}_{ba})}{\sum_{a=1}^{A} \exp(\mathrm{M}_{ba})}$$

$$\mathrm{M}_{ba} \sim N\left(\mu_{ca}, \frac{1}{\tau}\right) \qquad a>1,\ \mathrm{M}_{b1}=0$$

Model the cell means in terms of season, gear, year and area effects:

$$\mu_{ca} = \alpha_a + \beta_{s(c)a} + \chi_{g(c)a} + \delta_{y(c)a} + \varepsilon_{l(c)a}$$

β, χ and *δ* are given vague exchangeable Gaussian priors

$$\beta_{sa} \sim N(0,1/0.0001), s \neq 1, a \neq 1$$
$$\beta_{s1} = \beta_{1a} = 0, \forall s, a$$

Spatial ε term has a Gaussian conditional autoregressive (CAR) prior distribution:

$$\varepsilon_{ia} \mid \varepsilon_{j \neq ia} \sim N\left(\bar{\varepsilon}_{ia}, \frac{1}{\lambda n_i} \right)$$

$$\bar{\varepsilon}_{ia} = n_i^{-1} \sum_{j \in \partial(i)} \varepsilon_{ja}$$

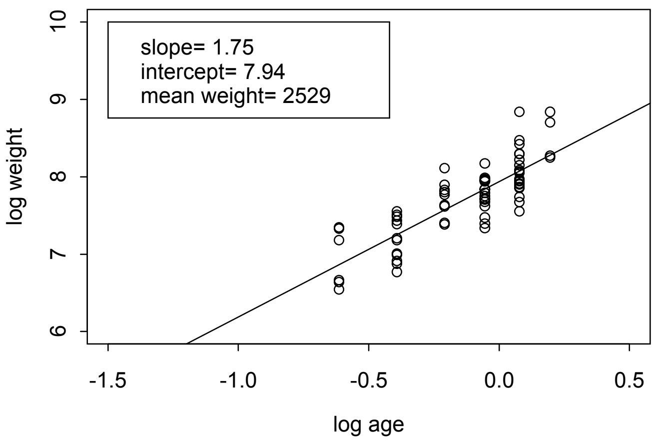
Modelling weight given age.

Assume:

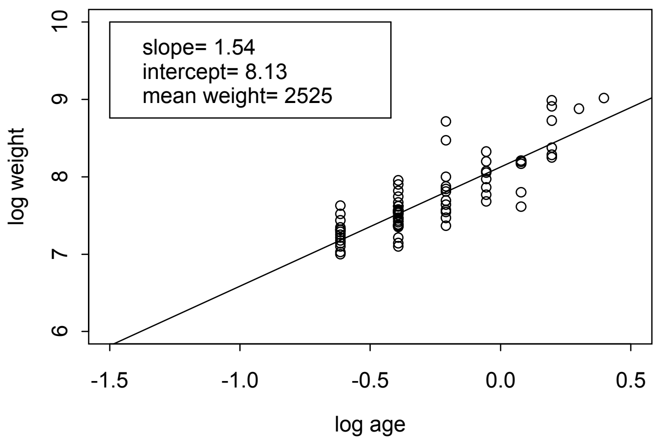
1) log(weight) is linearly related to log(age) on any given boat

Log(weight) vs log(age)

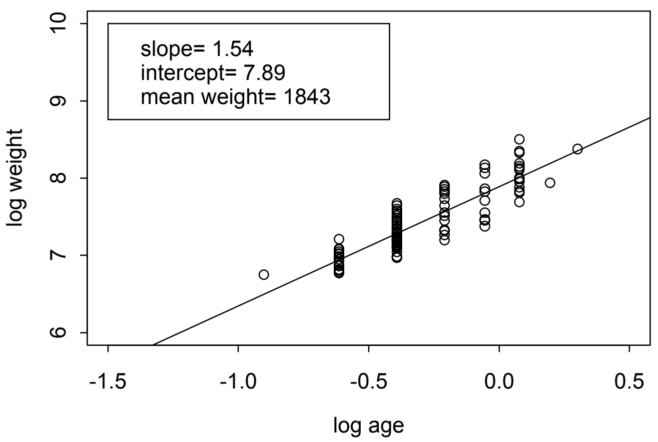
Boat 1



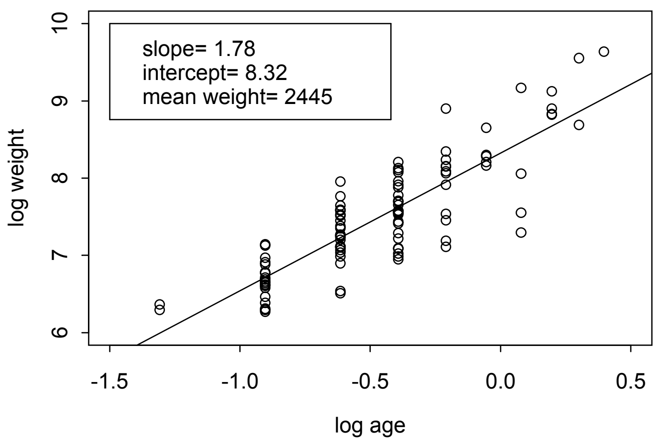
Boat 2



Boat 3



Boat 4



Modelling weight given age.

Assume:

1) log(weight) is linearly related to log(age) on any given boat

2) The slope and intercept parameters for the regressions are random variables, drawn from a cell population.

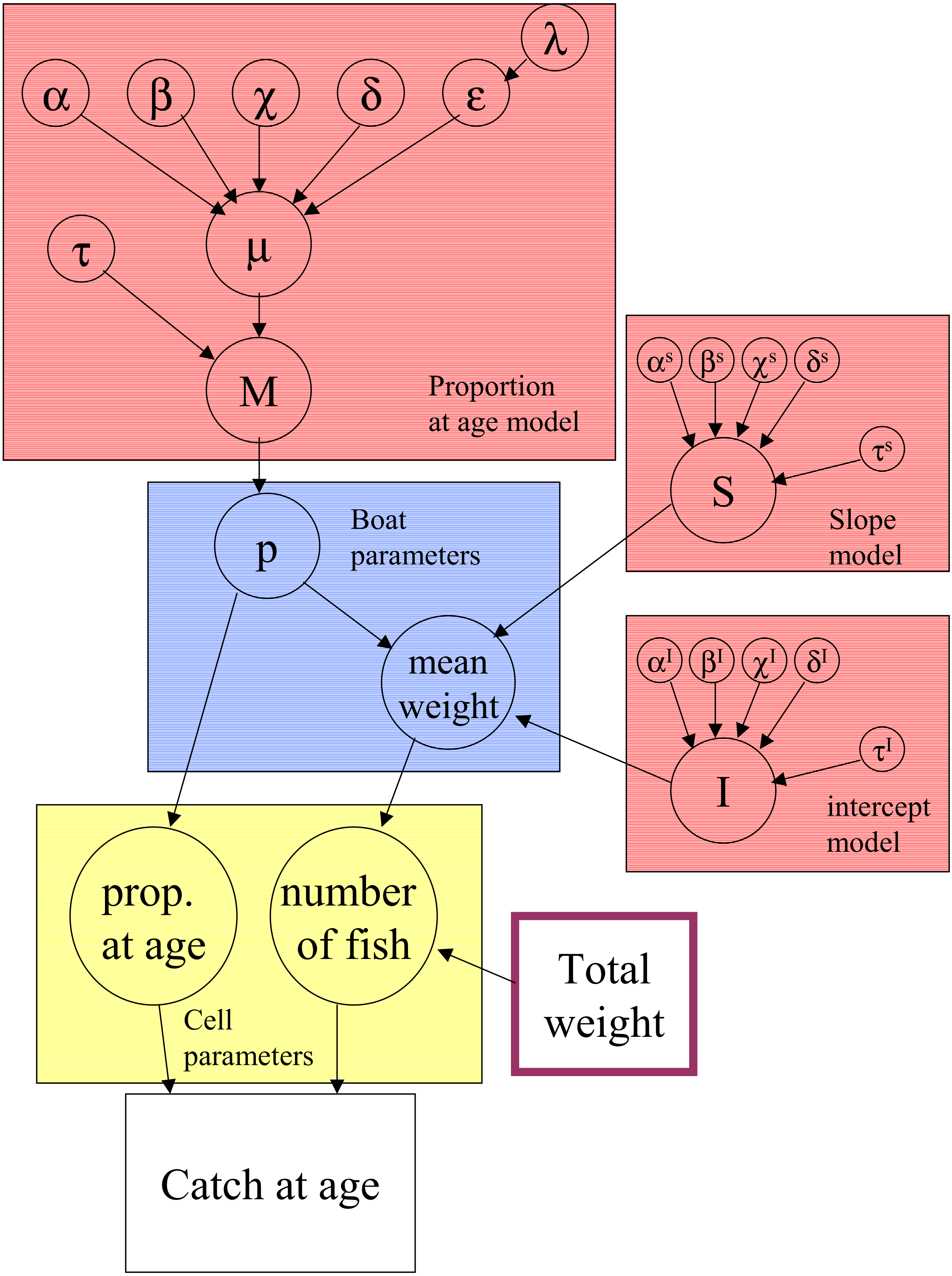
$$\log(weight_{bi}) = int_b + sl_b\left(\log(age_{bi}) - 2\right) + \xi_{bi}$$

$$sl_b \sim N\left(S_c, \frac{1}{\tau^s}\right)$$

$$S_c = \alpha^s + \beta^s_{s(c)} + \chi^s_{g(c)} + \delta^s_{y(c)}$$

$$int_b \sim N\left(I_c, \frac{1}{\tau^i}\right)$$

$$I_c = \alpha^i + \beta^i_{s(c)} + \chi^i_{g(c)} + \delta^i_{y(c)}$$


α
β
χ
δ
ε
λ
τ
μ
M
Proportion
at age model
p
Boat
parameters
mean
weight
α^s
β^s
χ^s
δ^s
τ^s
S
Slope
model
α^I
β^I
χ^I
δ^I
τ^I
I
intercept
model
prop.
at age
number
of fish
Cell
parameters
Total
weight
Catch at age

Estimating total catch at age

Assume:

1) There are a large number of fish on each individual boat, such that the mean weight given age on a boat is equal to its expected value:

$$E(w_b \mid a) = \left(\exp(int_b)\exp\left(sl_b(\log(a)-2)\right)\exp(1/2\tau^w)\right)$$

Mean weight on boat b = $\sum_a p_{ba} E(w_b \mid a)$

2) There are a large number of boats fishing in a cell, such that the mean weight of fish caught in the cell is equal to its expected value.

3) The proportion at age is independent of the weight given age.

The mean weight of fish in a cell is now:

$$w_{mean} = E_b\left(\sum_a p_{ba} E(w_b \mid a)\right)$$

$$= \sum_a E_b(p_{ba}) E_b\left(\exp(int_b)\exp\left(sl_b(\log(a)-2)\right)\exp(1/2\tau^w)\right)$$

$$= \sum_a E_b(p_{ba})\exp(I)\exp(1/2\tau^i)\exp\left(S(\ln(a)-2)\right) \text{ x}$$

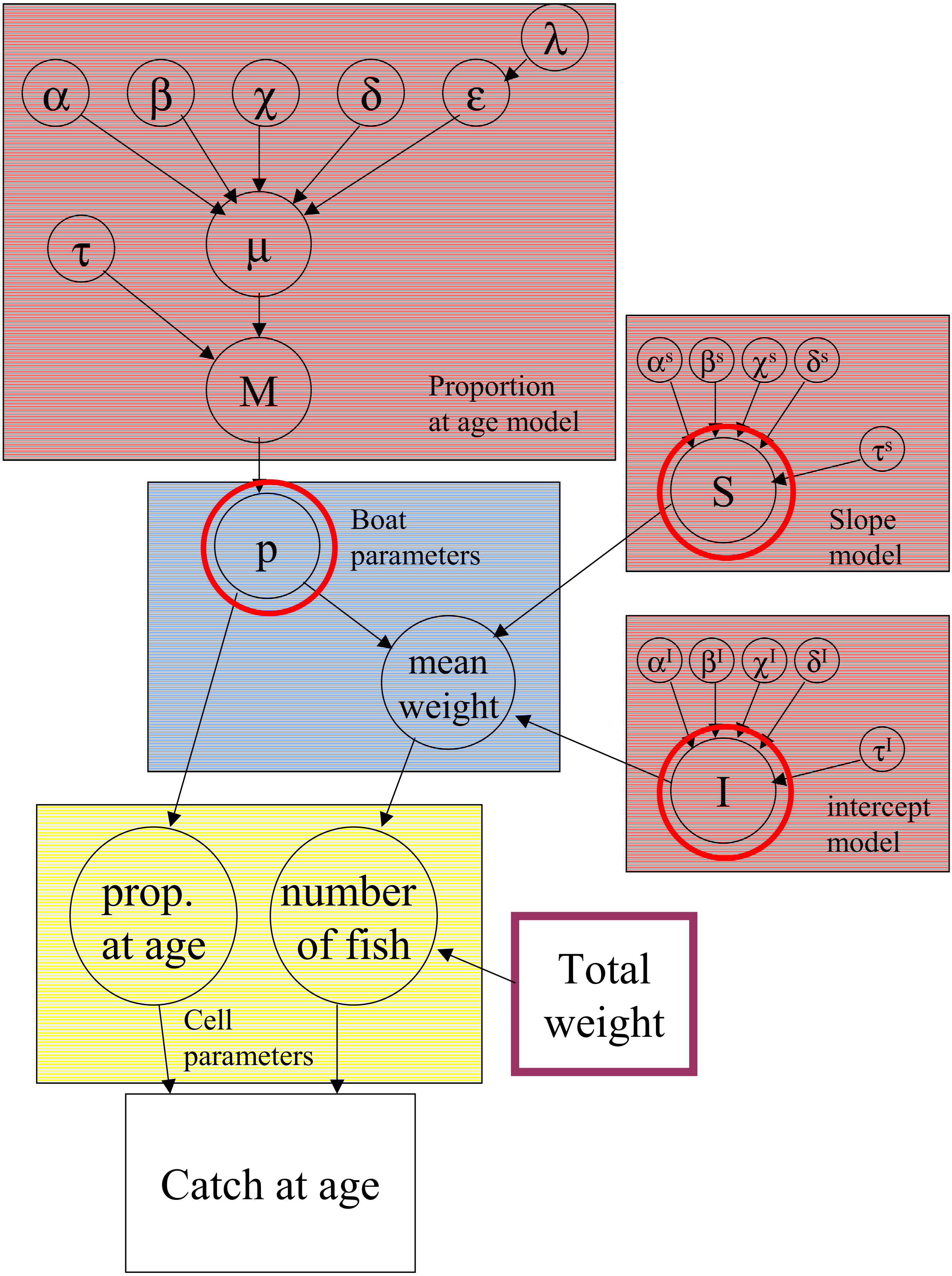
$$\exp\left((\ln(a)-2)^2/2\tau^s\right)\exp(1/2\tau^w)$$

4) The precision τ of the proportion at age parameters M_{ba} is sufficiently large compared to the range of the means μ_a that the expected value of the proportions at age is equal to the maximum likelihood estimator, ie

$$E(P_a) \approx \frac{\exp(\mu_a)}{\sum \exp(\mu_i)}$$

5) We are given without error the weight of the total catch W in each cell. Therefore the total number of fish caught in the cell is given by $T = W / w_{mean}$

6) The proportion at age in the cell is equal to its expected value, ie $T_a = TE(P_a)$


λ
α
β
χ
δ
ε
τ
μ
M
Proportion
at age model
α^s
β^s
χ^s
δ^s
τ^s
S
Slope
model
p
Boat
parameters
mean
weight
α^I
β^I
χ^I
δ^I
τ^I
I
intercept
model
prop.
at age
number
of fish
Cell
parameters
Total
weight
Catch at age

Likelihood for the Norwegian data:

Assume:

1) Boats are randomly selected from population within a cell

2) Fish are randomly selected from a boat

$$\boldsymbol{X_b} \sim \text{multinomial}(\boldsymbol{p_{b.}}, n_b)$$

Where $\boldsymbol{X_b}$ is the vector of numbers at age in the sample from boat b.

Estimating age from an otolith

Model for age errors:

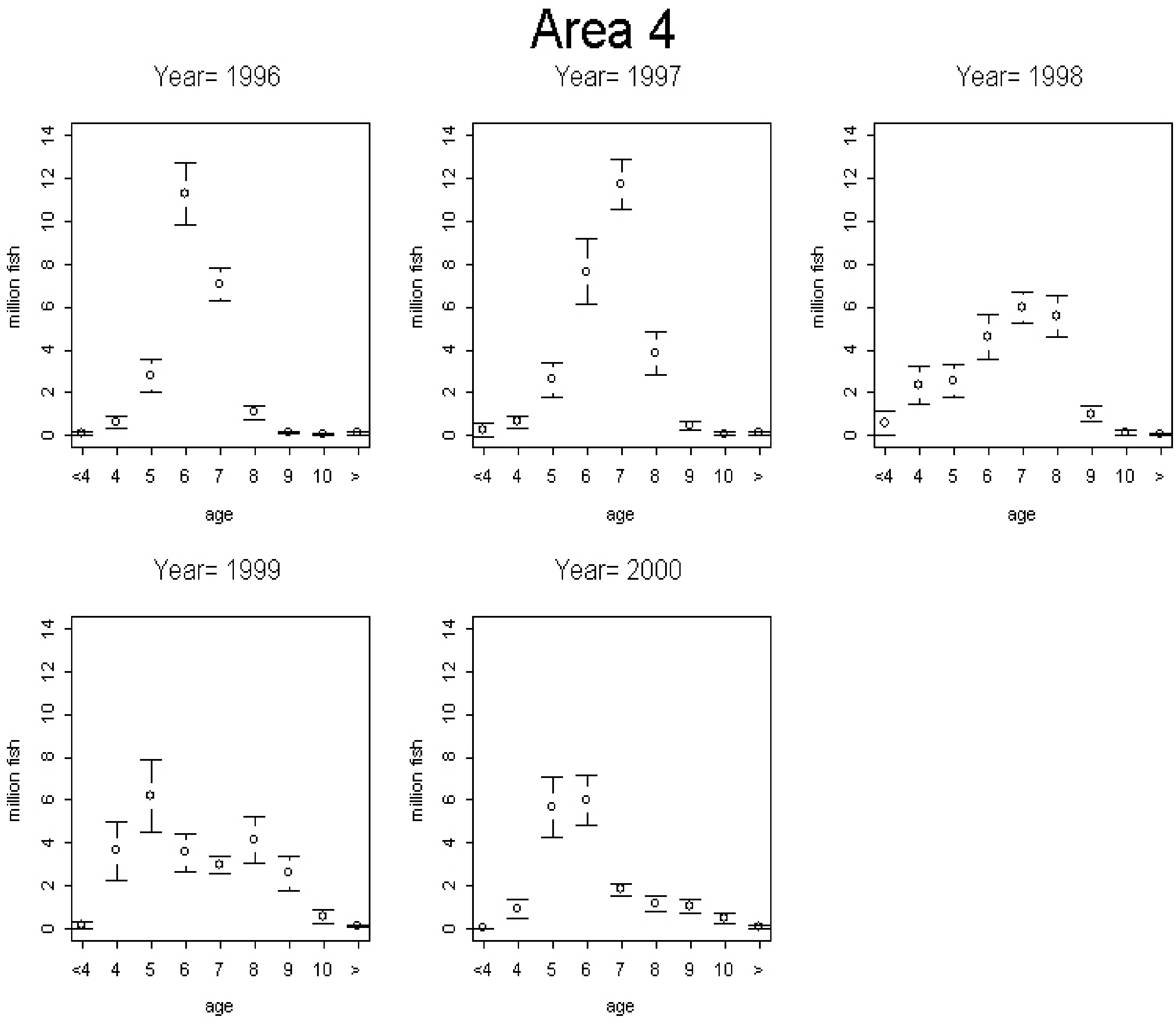
Suppose probability of mis-ageing is known,

$P(\text{error})_{ij}$= p(observed age=i|true age=j) known $\forall$i,j

eg “10% chance of age being wrong by 1 year”

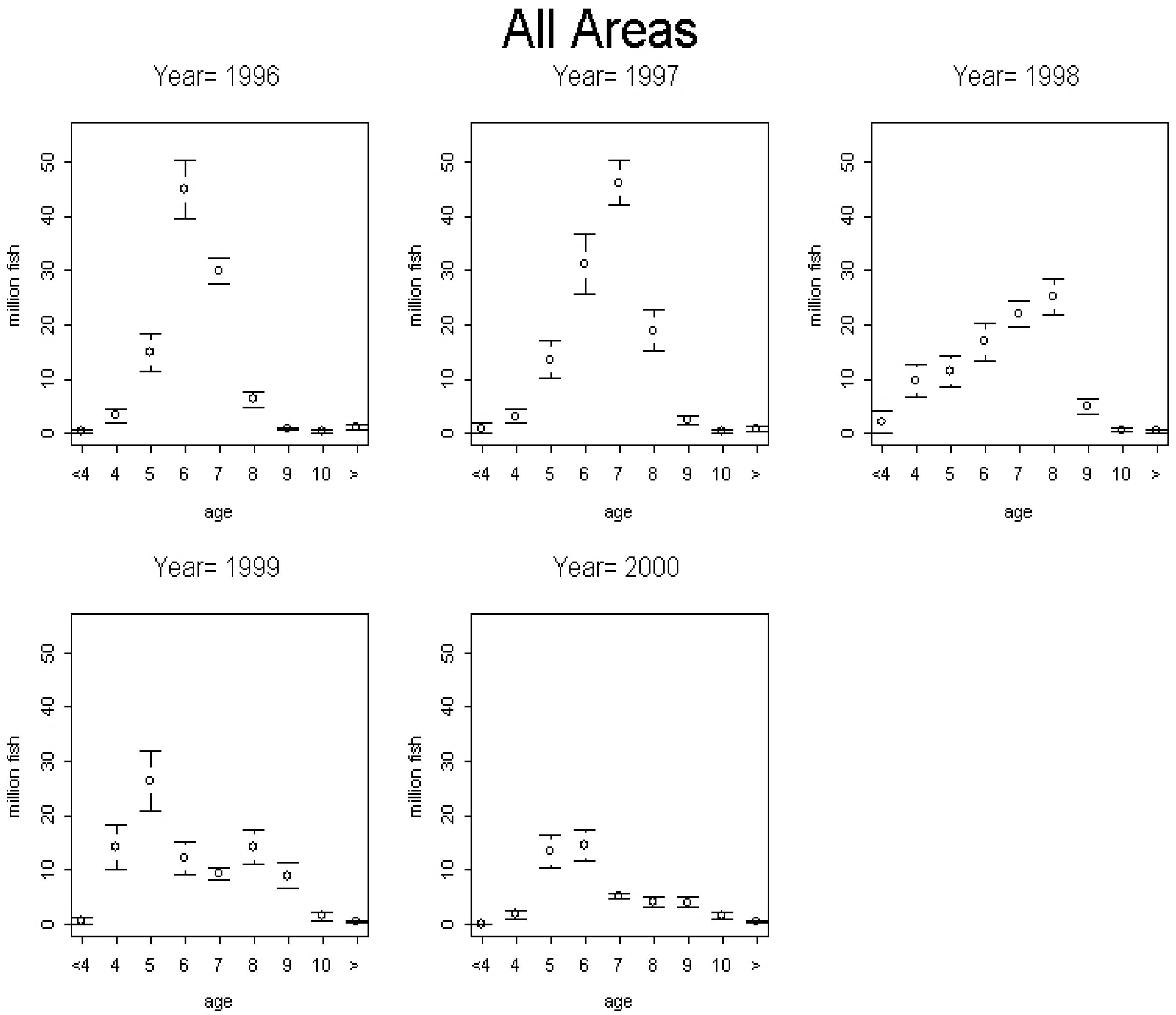
Observed ages are multinomial $P(\text{error})\boldsymbol{p}_b$

Total Catch

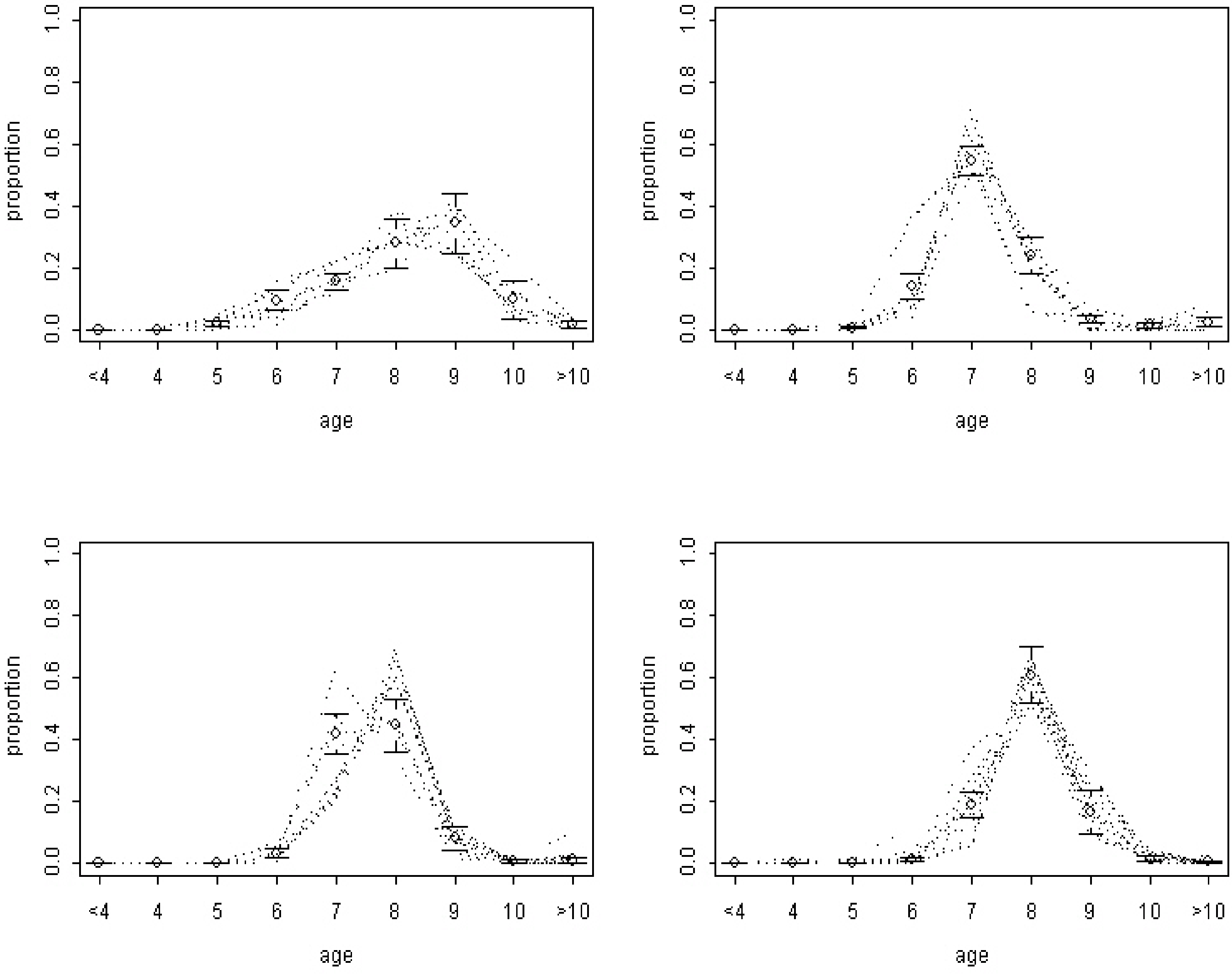


Area 4
Year= 1996
Year= 1997
Year= 1998
Year= 1999
Year= 2000
million fish
age
<4
4
5
6
7
8
9
10
>
0
2
4
6
8
10
12
14

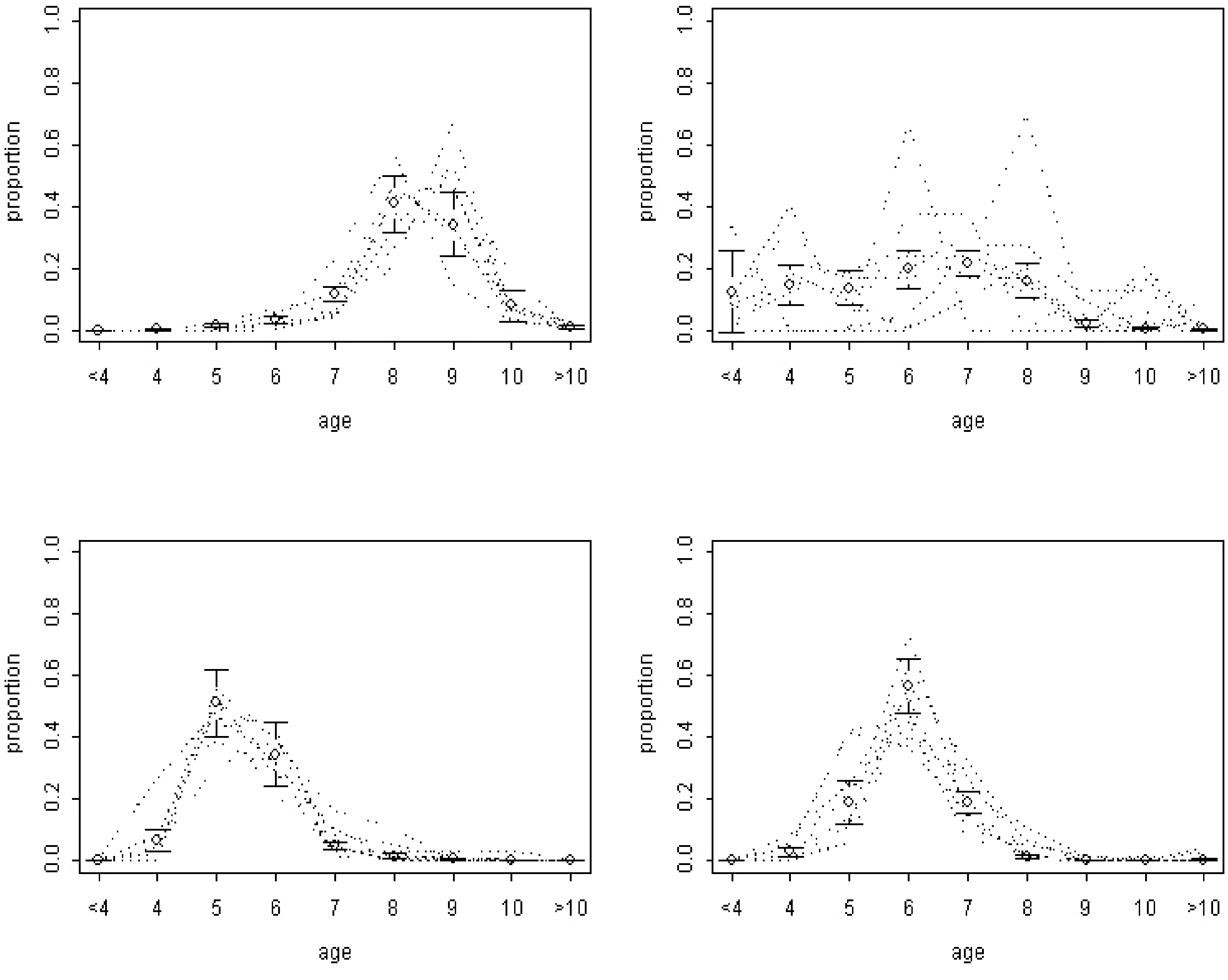
Total Catch


All Areas
Year= 1996
Year= 1997
Year= 1998
Year= 1999
Year= 2000
million fish
age
0
10
20
30
40
50
<4
4
5
6
7
8
9
10
>

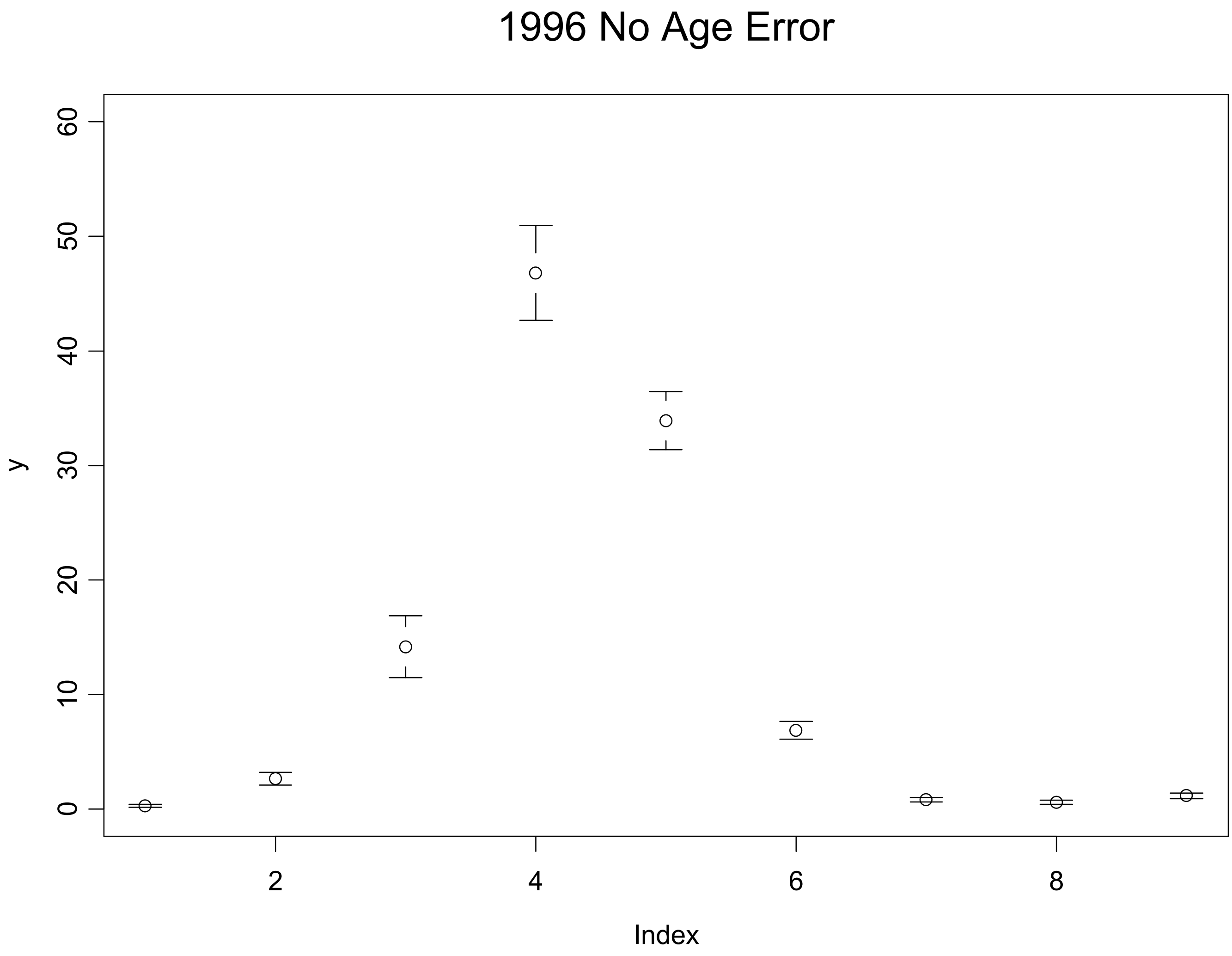
The fit within a cell



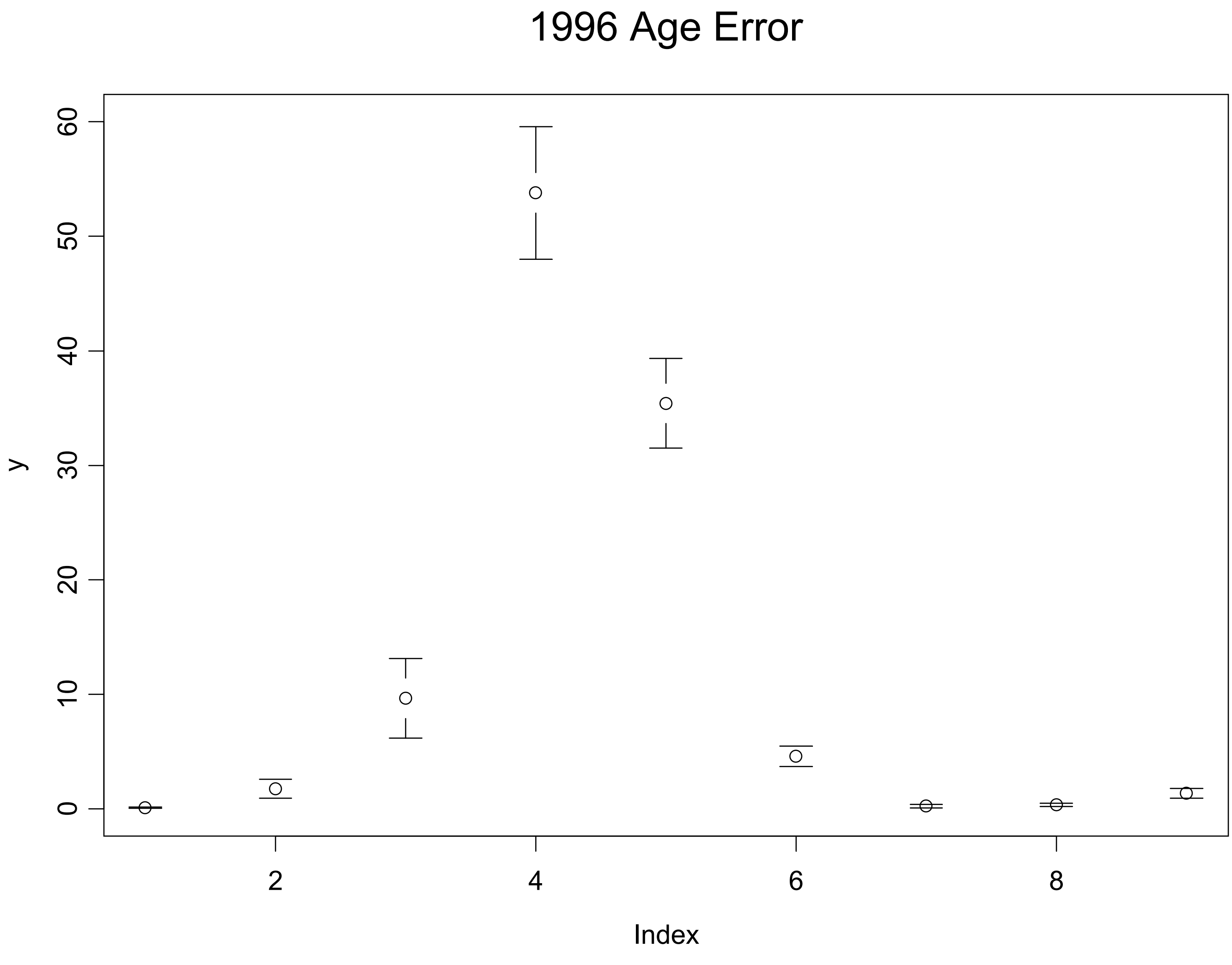
The fit within a cell



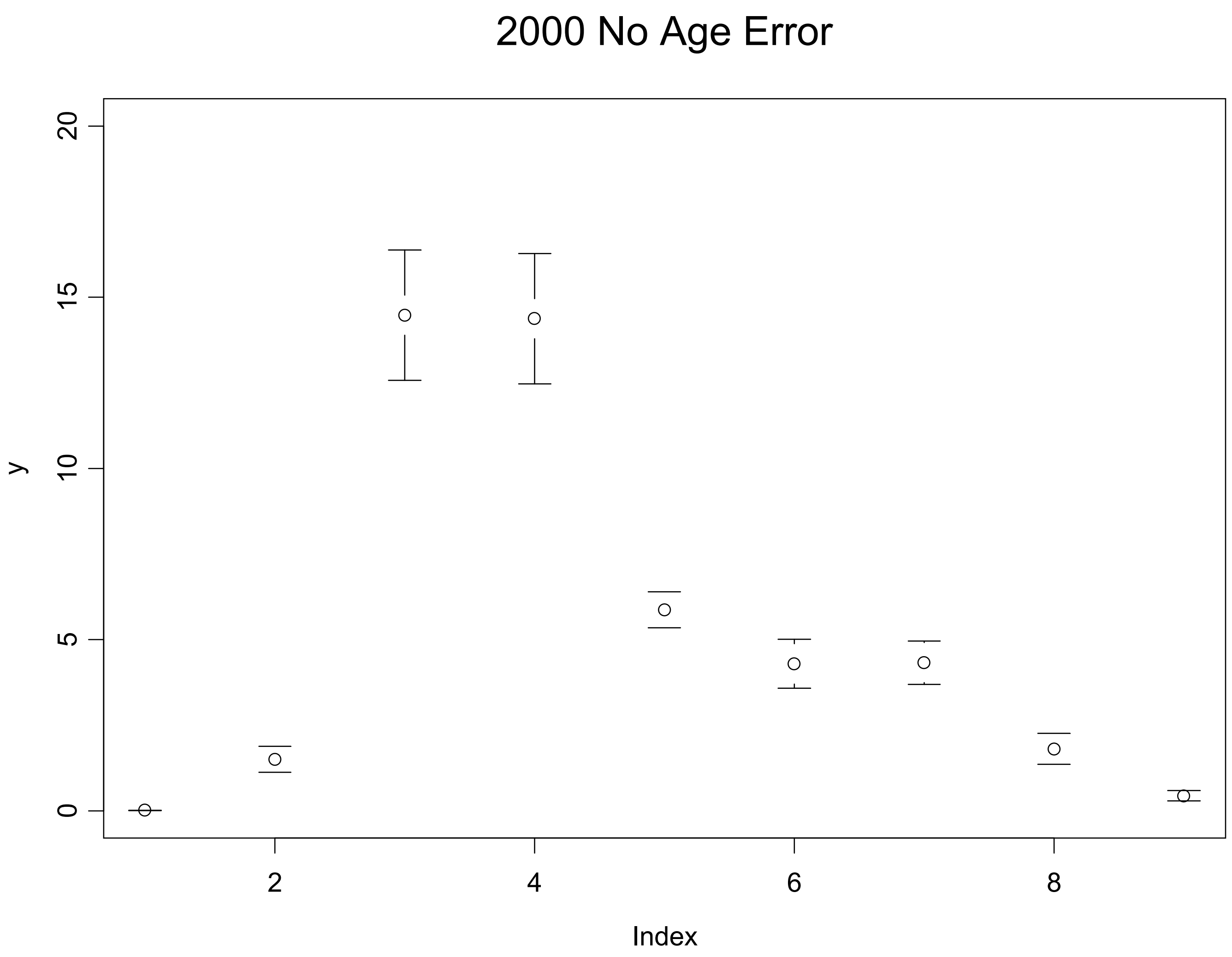
Effect of Age Errors



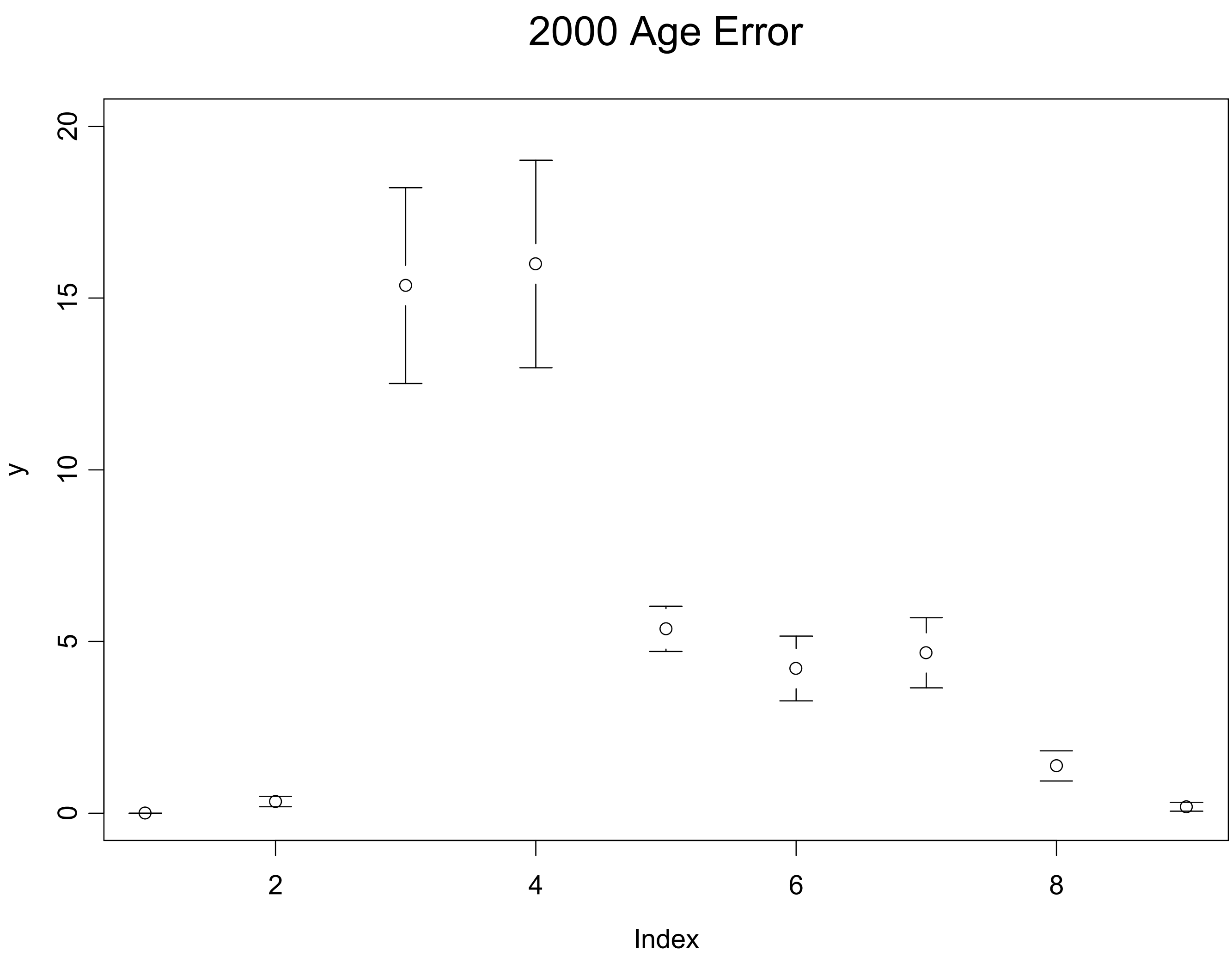
Effect of Age Errors



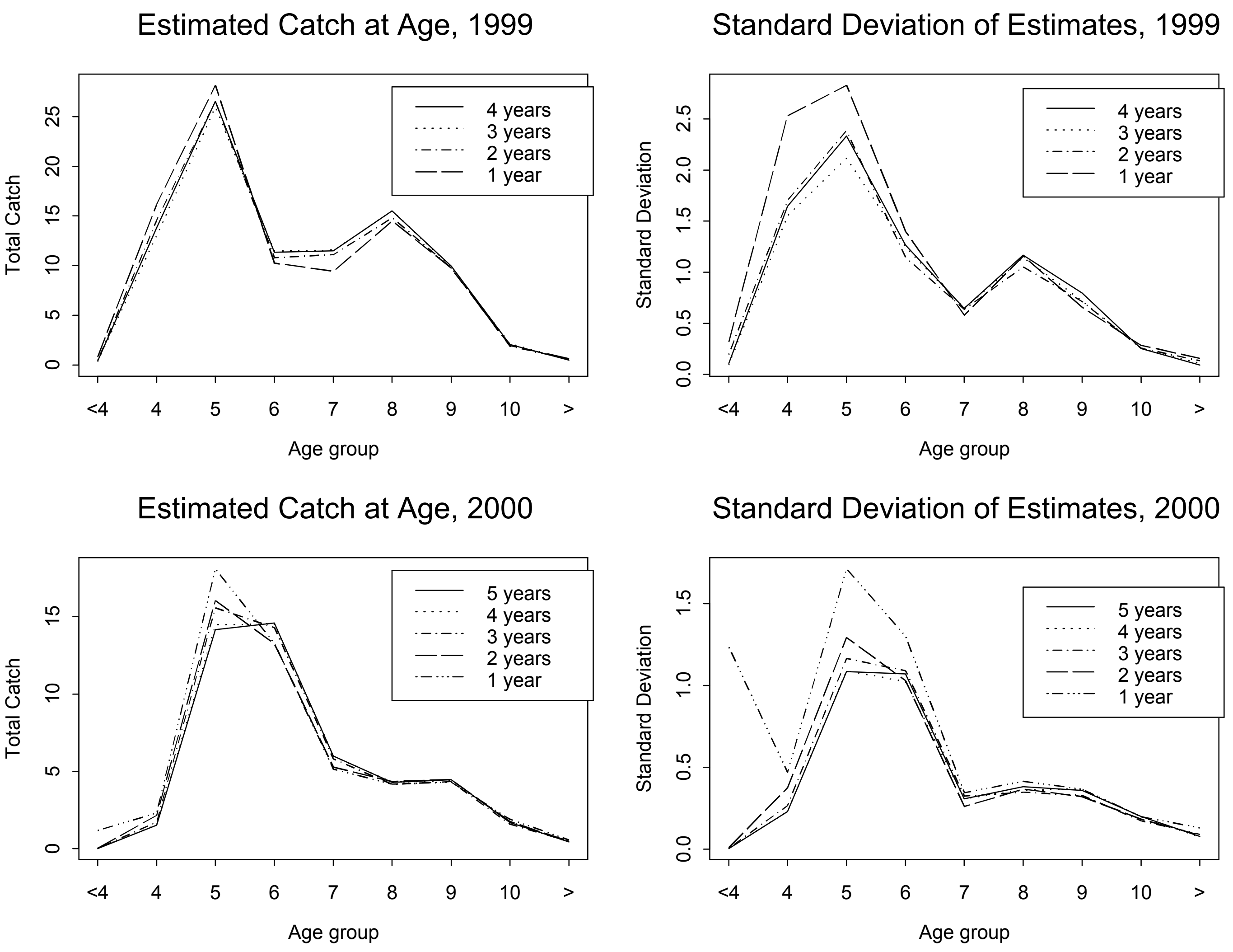
Effect of Age Errors



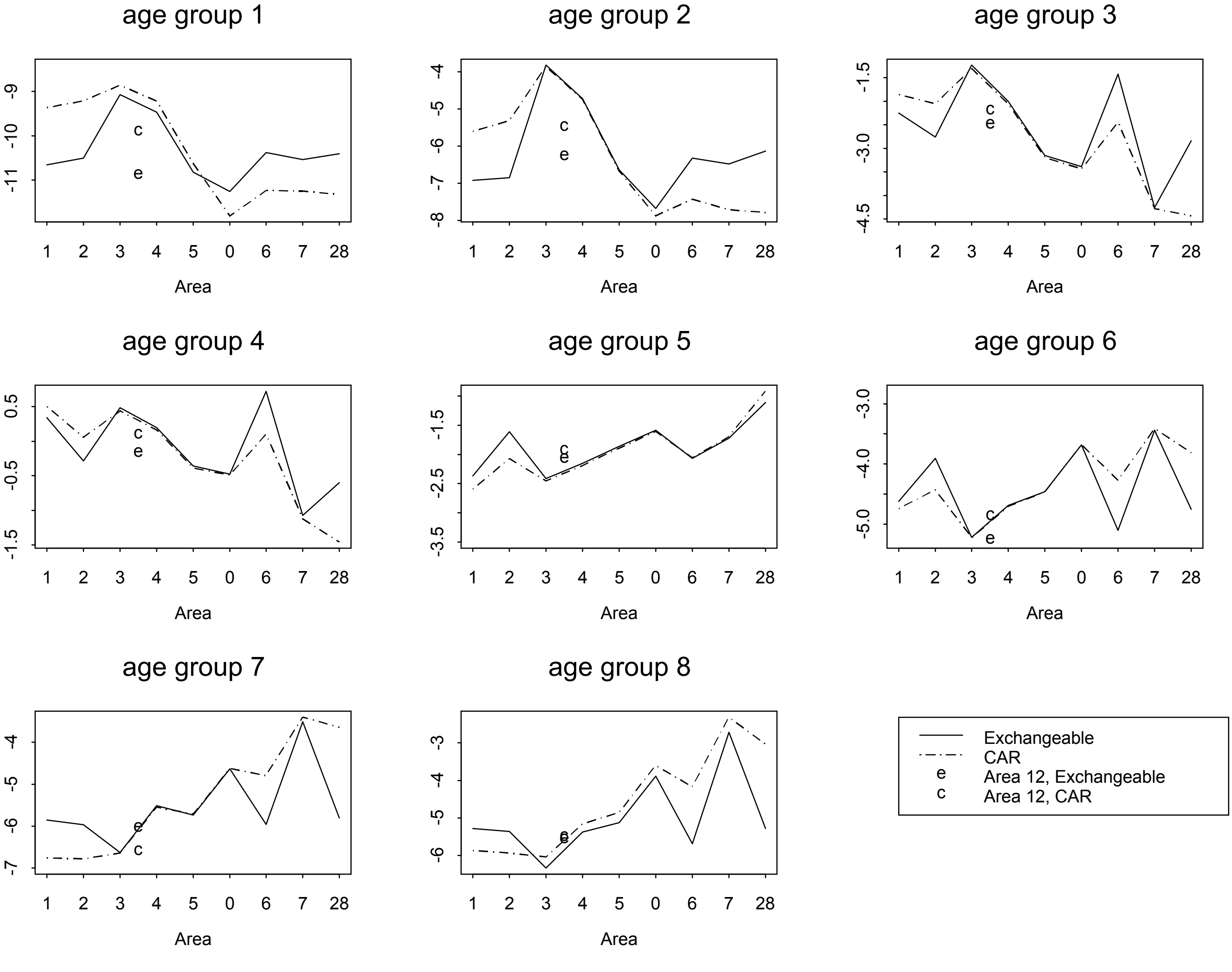
Effect of Age Errors



The effect of using multiple years



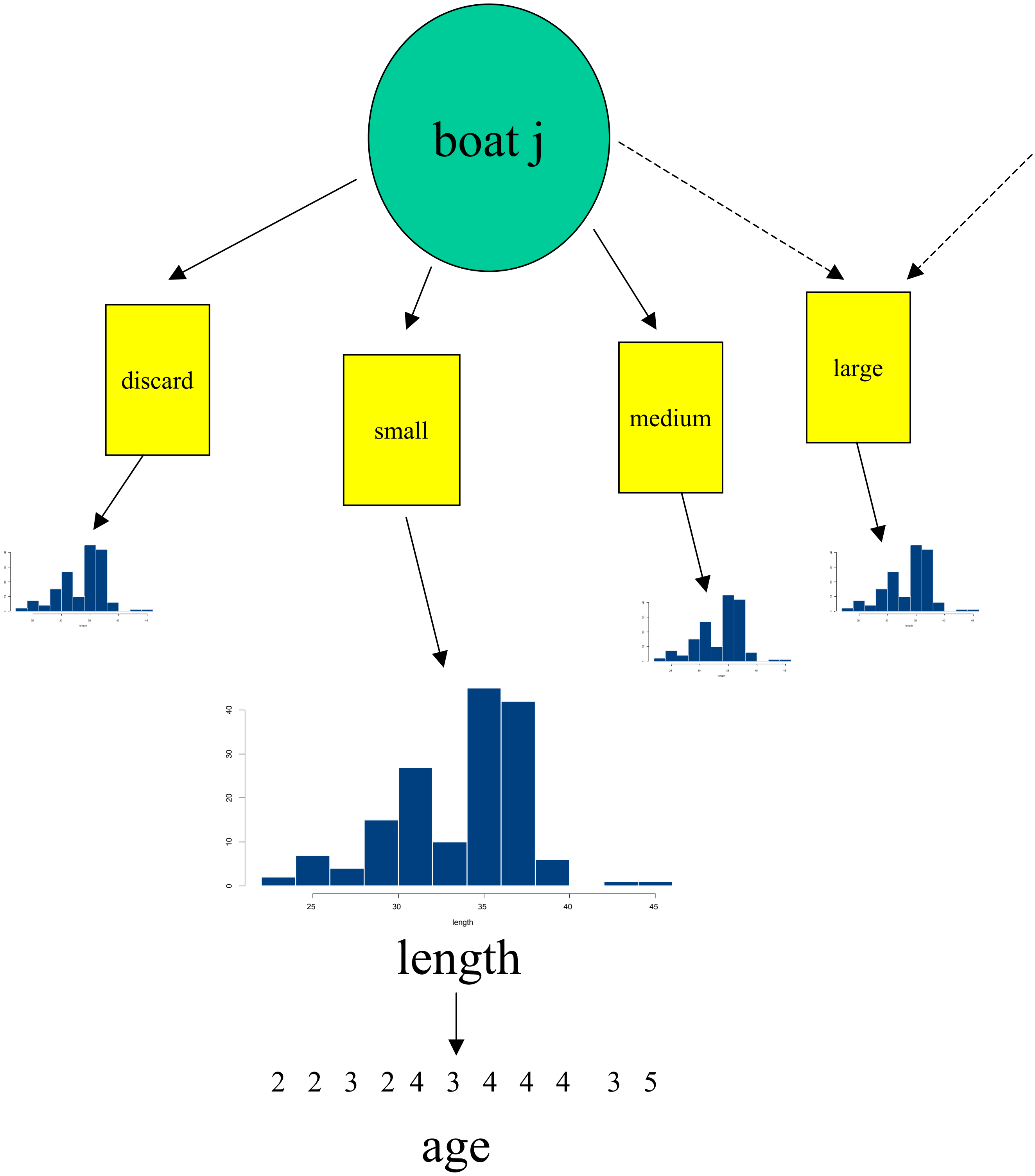
Spatial effects



The Likelihood for the Baltic data:

Samples are stratified first by weight class, then by length. One fish in each length class is aged, and the total number in each class counted.

Age, length and weight should be modelled simultaneously.


boat j
discard
small
medium
large
length
2 2 3 2 4 3 4 4 4 3 5
age

We just consider one weight class, and assume

1) The length samples from boat *b* are multinomial, with the probability of a fish being in length category *l* given by

$$p_b(l) = \sum_a p_b(l \mid a) p_b(a)$$

2) Age given length is also multinomial

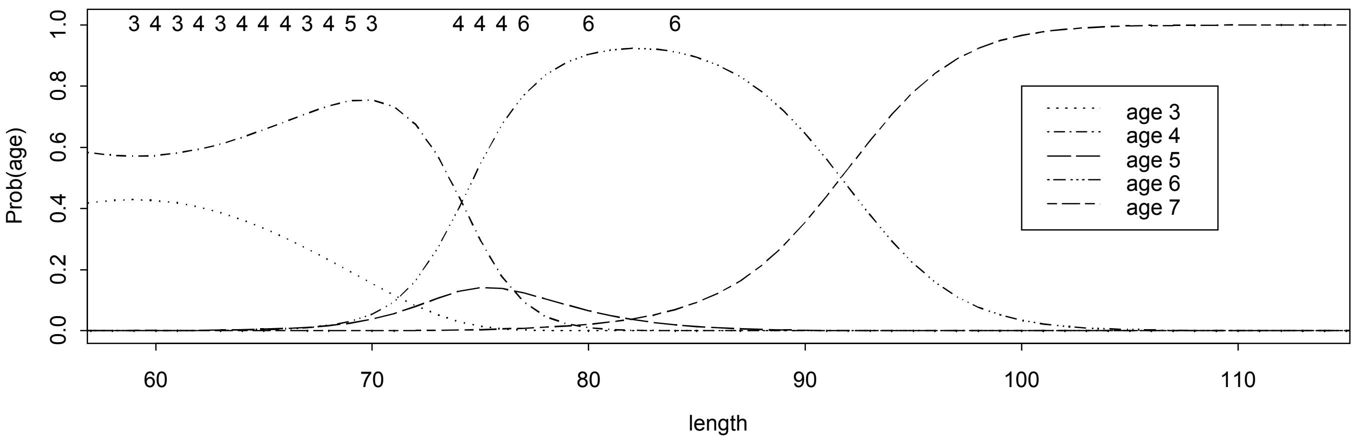
$$p_b(a \mid l) = \frac{p_b(l \mid a)}{p_b(l)} p_b(a)$$

3) Length given age is Gaussian:

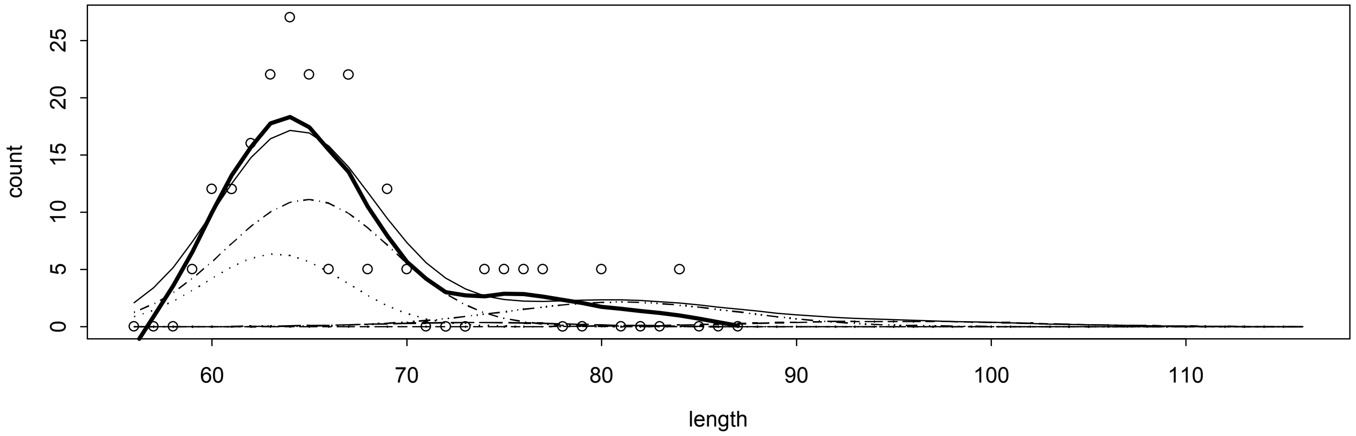
$$(l \mid a)_b \sim N\left(\lambda_{ba}, \frac{1}{\varsigma_{ba}}\right)$$

Baltic Boat 1

Age data

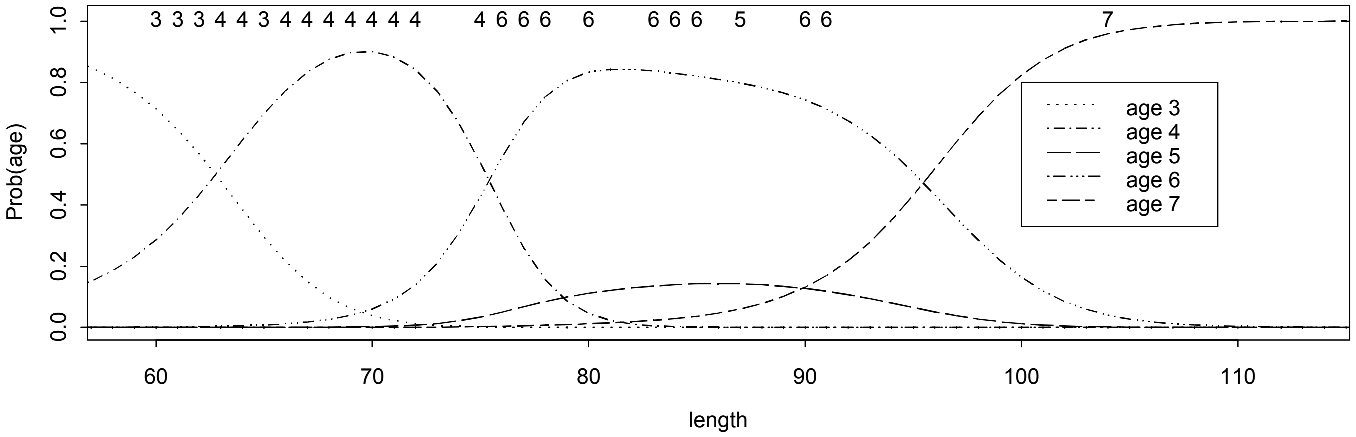


Length data

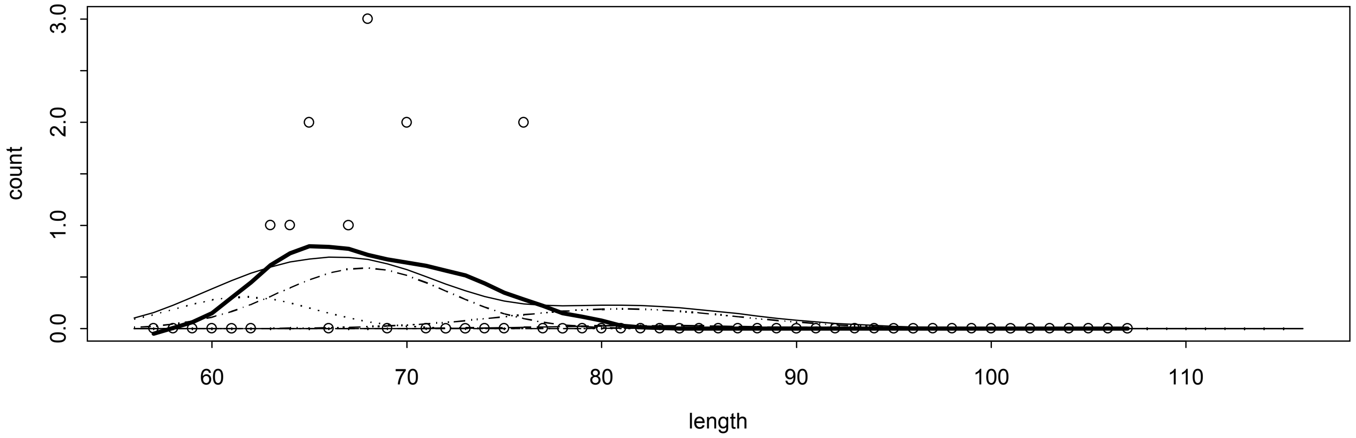


Baltic Boat 2

Age data

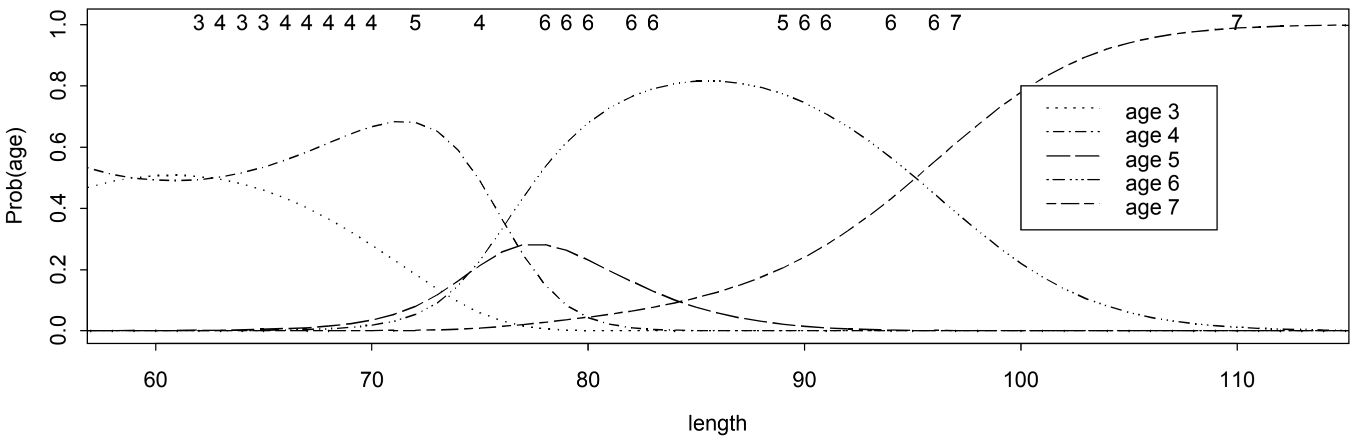


Length data

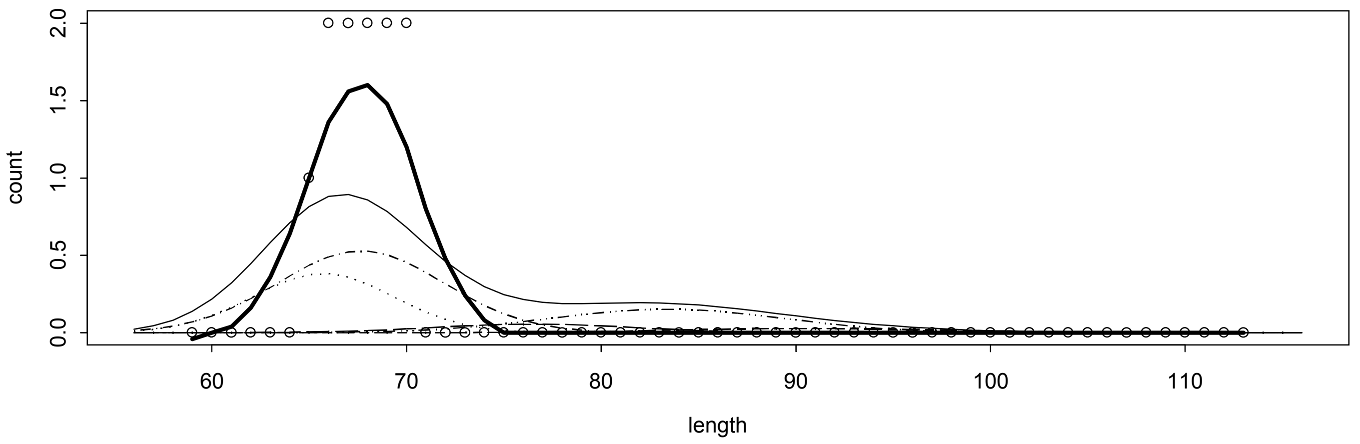


Baltic Boat 3

Age data

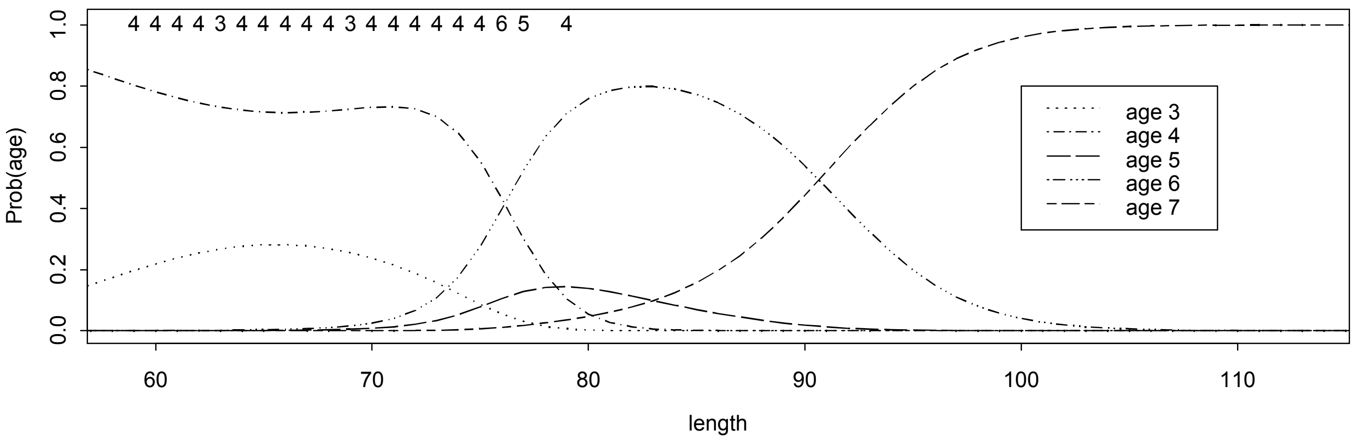


Length data

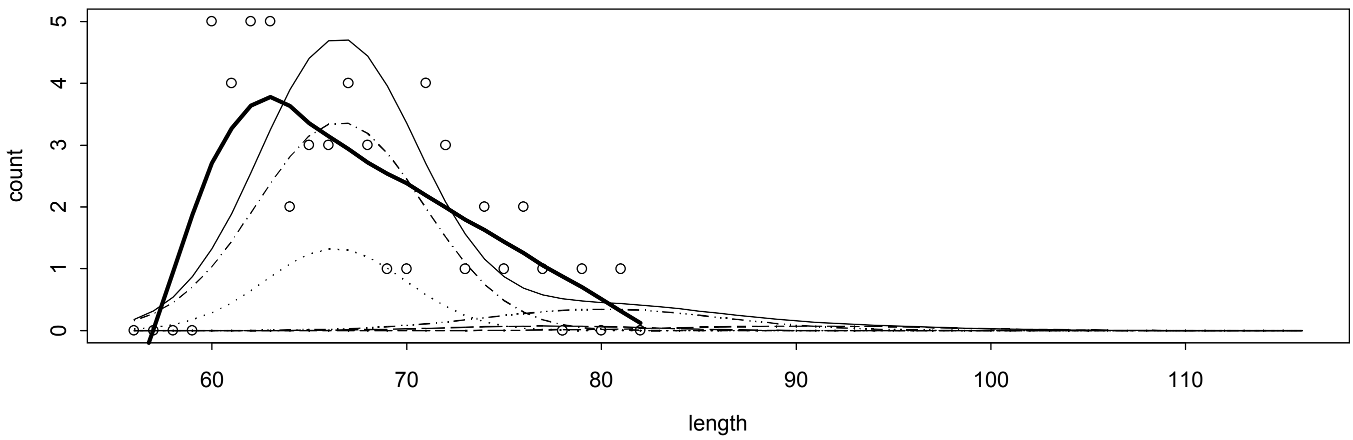


Baltic Boat 4

Age data



Length data



Conclusions

1) It is possible to model catch-at-age in a relatively simple way

2) Uncertainty is easily found

3) The model is fast, repeatable and easy to use

4) The potential saving in effort is huge