Progress on length based SAM

Anders Nielsen & Casper Berg
an@aqua.dtu.dk

DTU Aqua
National Institute of Aquatic Resources

Outline

- Background: SAM and time-varying selectivity
- Challenges in time-varying length based selectivity.
- Template Model Builder
- State-space Length structured ASsessment MOdel (SLASMO)

Focus here on fishing mortality $F_{a,y}$ (and later $F_{l,y}$)

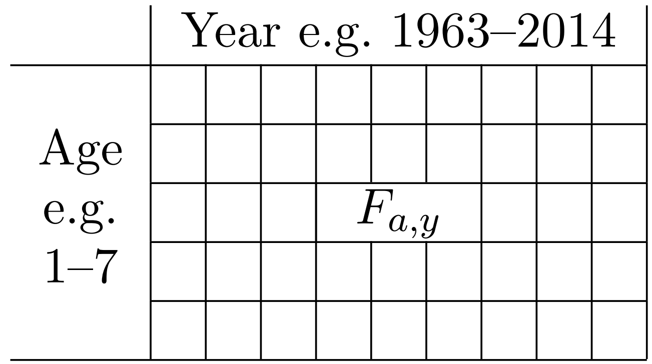
- $F_{a,y}$ is the instantaneous fishing mortality rate for age group a in year y
- Define the selectivity as:

$$S_{a,y} = \frac{F_{a,y}}{\sum_a F_{a,y}},$$

  Often the selectivity is scaled to have a maximum of one in each year.

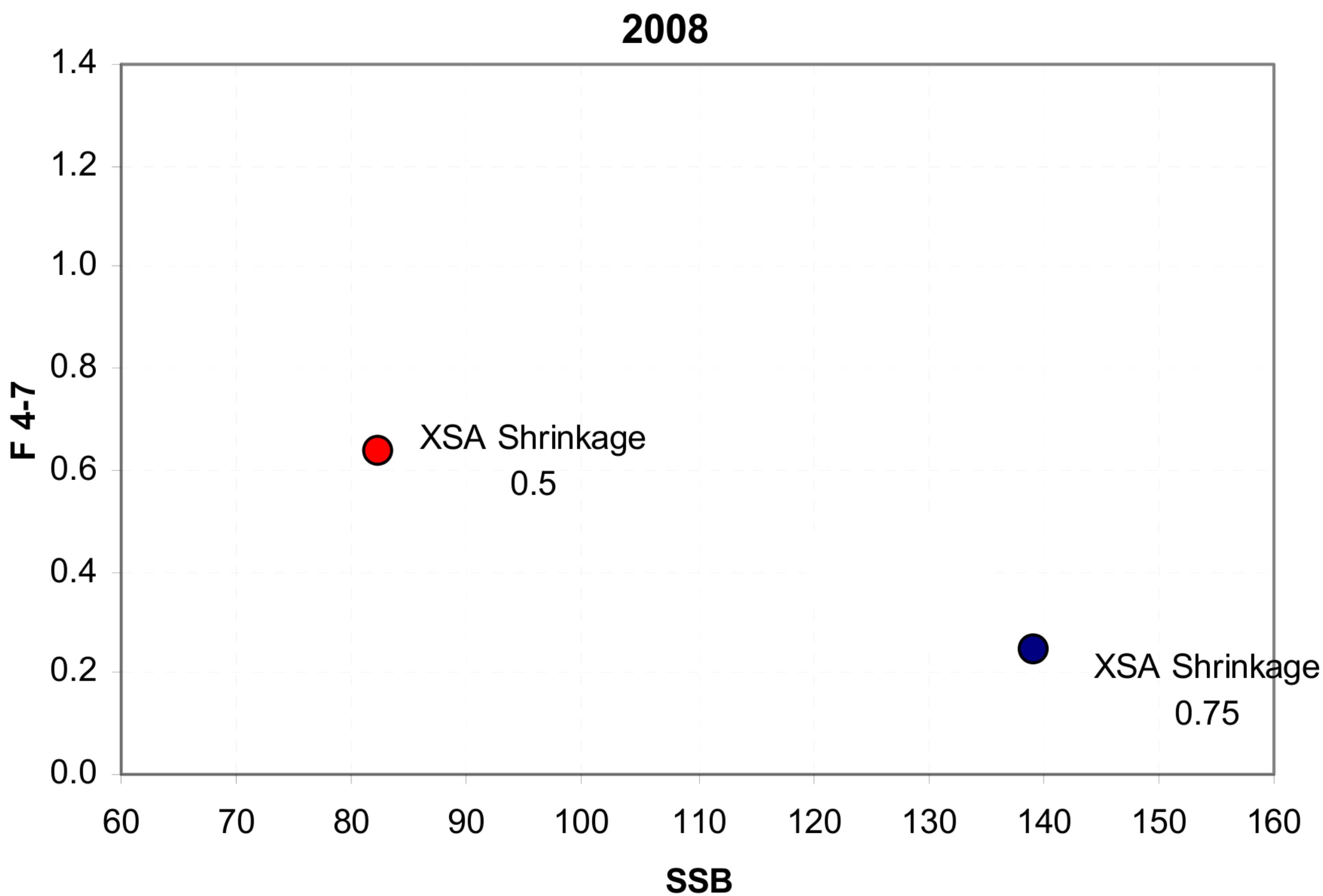
- There are many other interesting aspects in stock assessment models.
  - Natural mortality
  - Effort
  - Survey design
  - Stock recruitment model
  - Prediction ability
  - Growth
  - ...

Approaches used to model F



- Deterministic (Catch-at-age assumed known without error)
  - Extremely flexible $F_{a,y}$ model.
  - Ad-hoc smoothing often added
  - No quantification of uncertainties
- Full parametric. E.g:
  - $F_{a,y} = S_a f_y$ (S may be a parametrized function)
  - $F_{a,y} = S_a f_y$ with separate S_a in time blocks
  - Splines with fixed degree of smoothness
  - Penalized deviances with fixed penalty

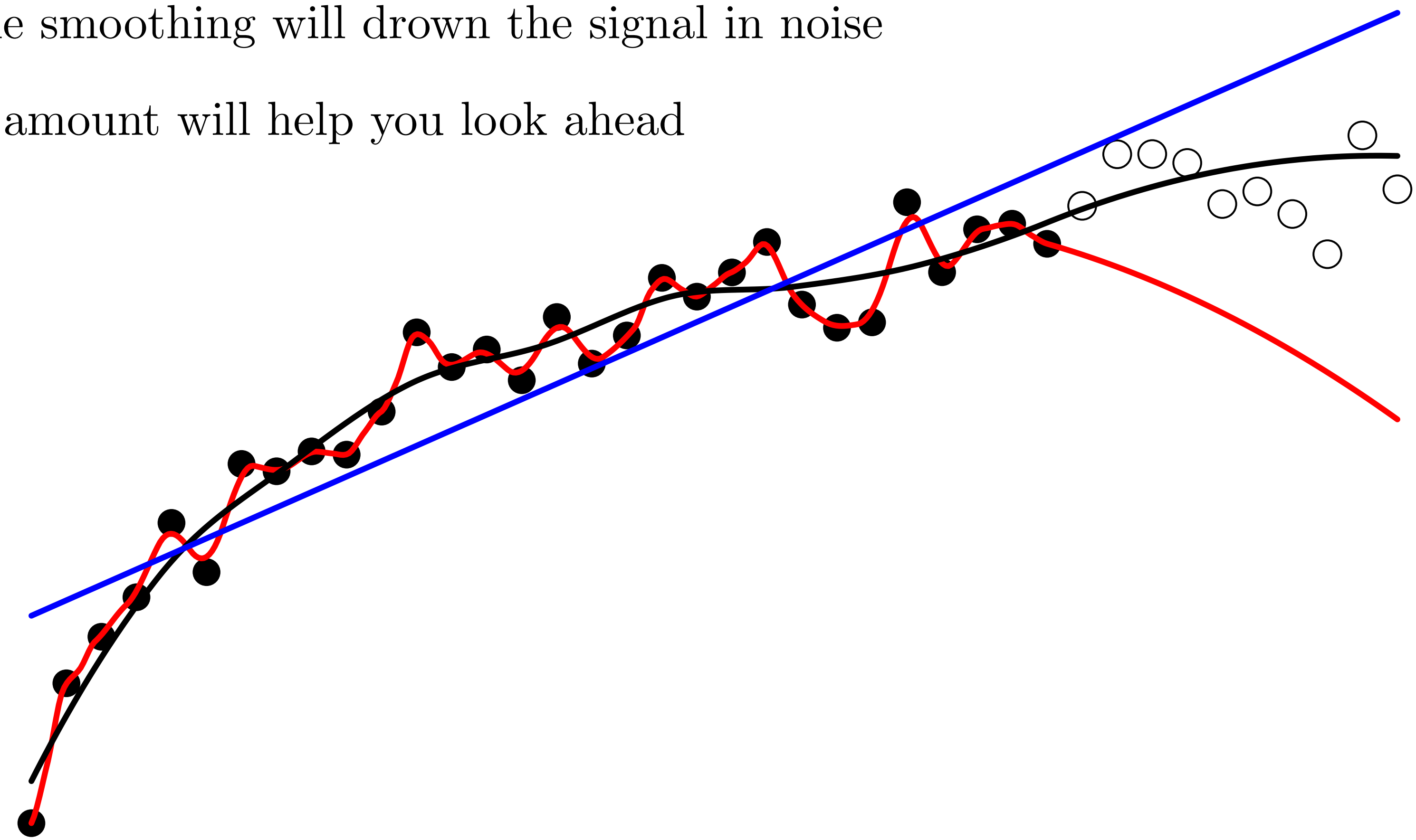
These choices are important



- These differences are not small and theoretical
- There are no objective way to choose between these two deterministic approaches
- There should really be an objective criteria. A statistical framework.

Further we need to predict

- Too much smoothing will bias the signal
- Too little smoothing will drown the signal in noise
- Correct amount will help you look ahead



- Correct amount should not be subjective.

State-space assessment models

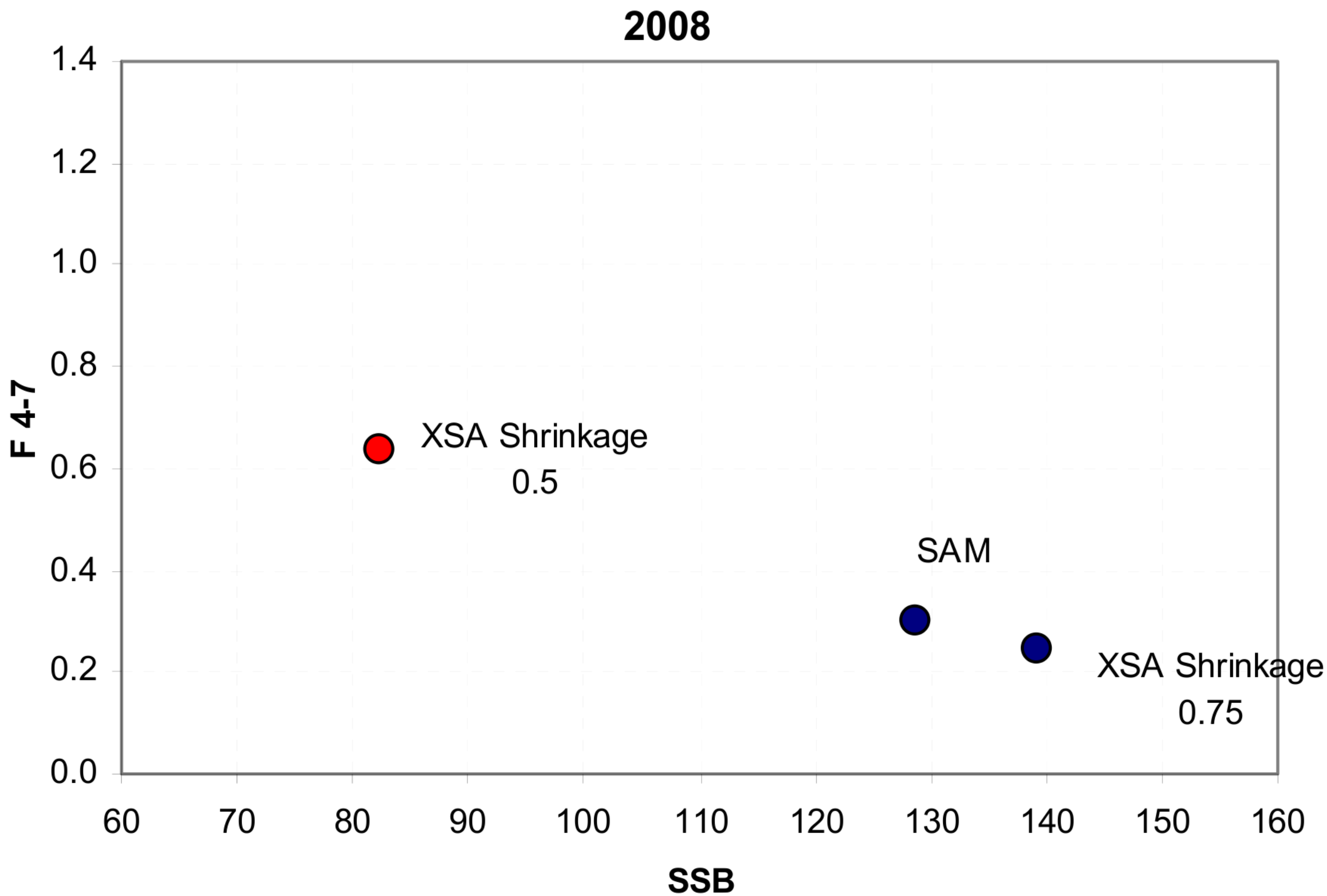
- Allow us to include things like:

$$F_{3,y} \text{ is a random walk with step variance } \sigma^2$$

- Importantly σ^2 is a model parameter estimated in the model.
- This model class[a] is used in most other quantitative fields
- It is a very useful extension to full parametric statistical models.
- Introduced for stock assessment by Gudmundsson (1987,1994) and Fryer (2001).
- The reason state-space models have not been more frequently used in stock assessment is that software to easily handle these models has not been available. It is now!
- Has recently received increased attention (e.g. Brinch et al., 2011; Gudmundsson and Gunnlaugsson, 2012; Berg et al., 2013; Nielsen and Berg 2014)
- Can give very flexible models with low number of model parameters

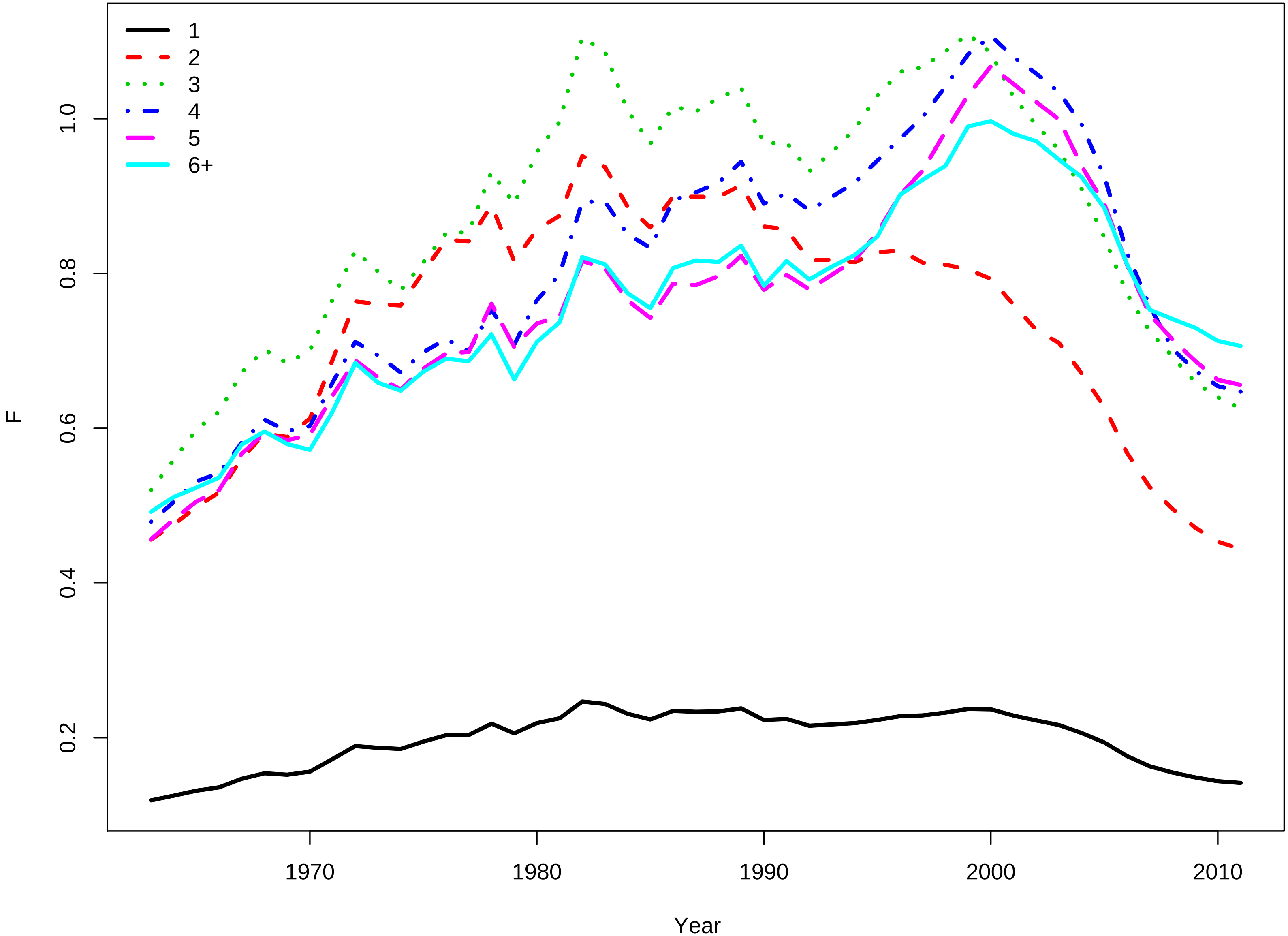
[a] a.k.a. random effects models, mixed models, latent variable models, hierarchical models, ...

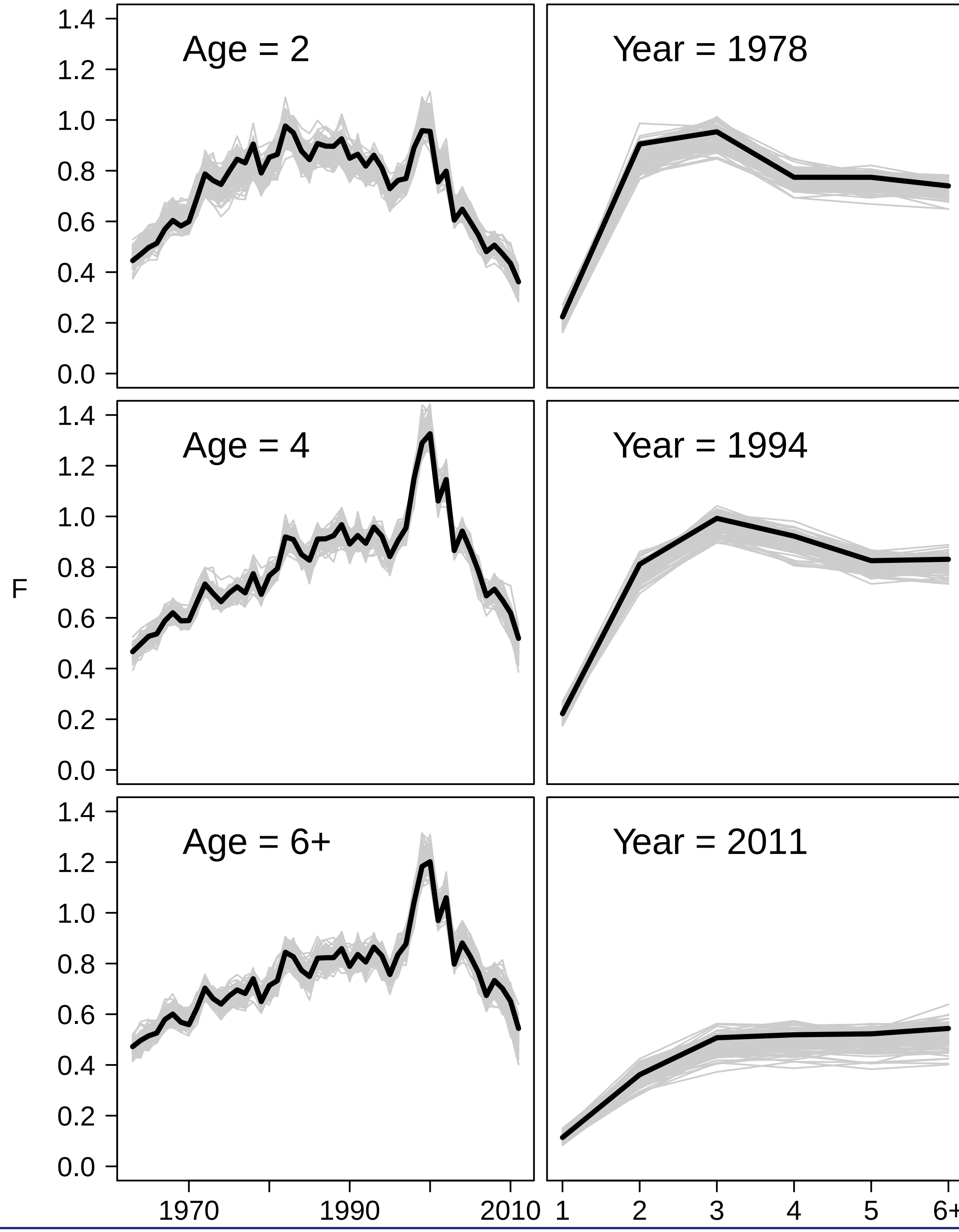
Avoiding ad-hoc choices — Eastern Baltic Cod



- Using the State-space Assessment Model (SAM) gives us an objective criteria

Evolving selectivity — North Sea Cod





Age = 2
Year = 1978
Age = 4
Year = 1994
Age = 6+
Year = 2011
F
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1970
1990
2010
1
2
3
4
5
6+

Model selection ex: Correlated Random Walks

- Instead of independent random walks for F at different ages, we can allow those random walks to be correlated. Define $\Delta \log F_y = \log F_y - \log F_{y-1}$, then:

$$\Delta \log F_y \sim \mathcal{N}(0, \Sigma)$$

- For all combination of ages $(a \neq \tilde{a})$:

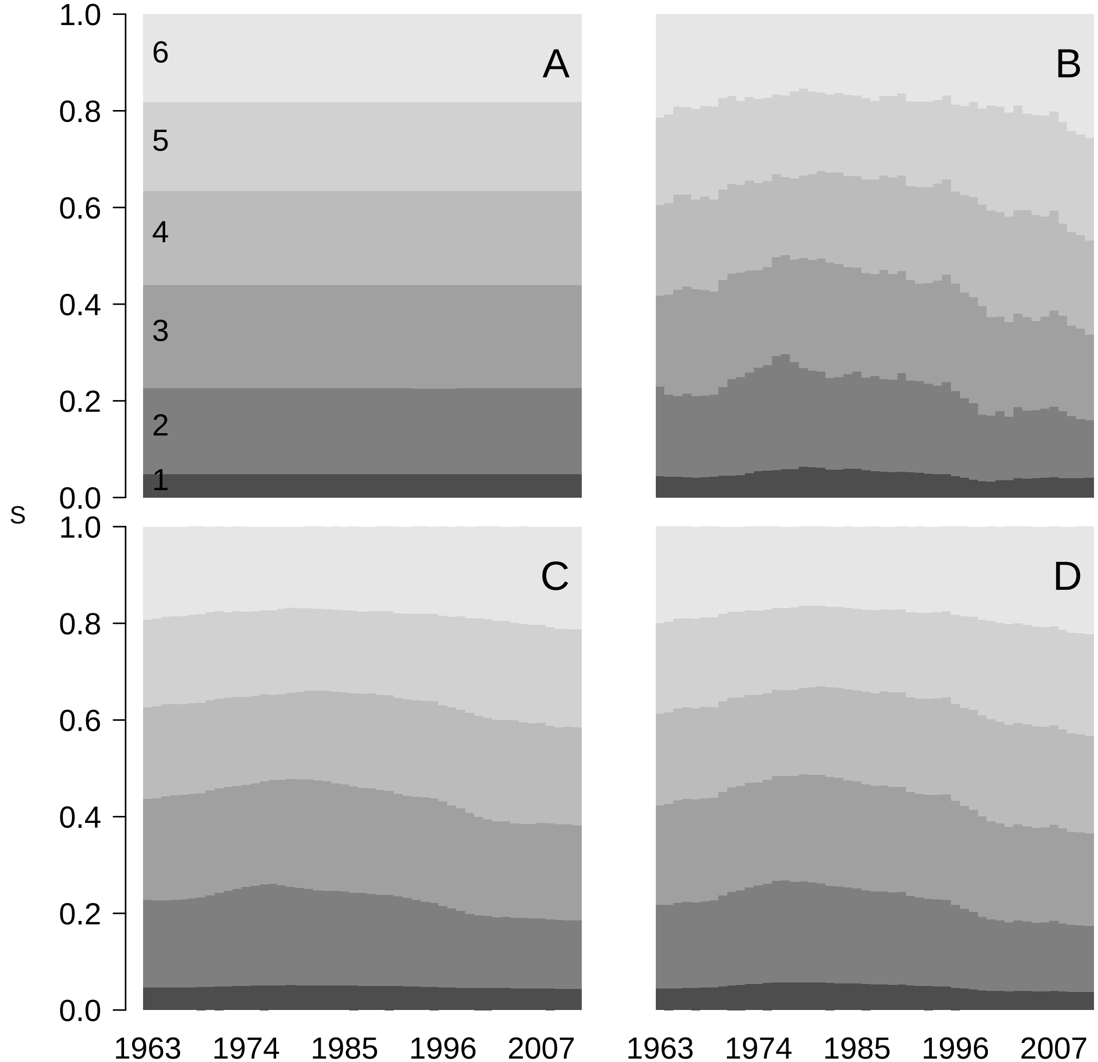
  A) Parallel: $\Sigma_{a,\tilde{a}} = \sqrt{\Sigma_{a,a}\Sigma_{\tilde{a},\tilde{a}}}$

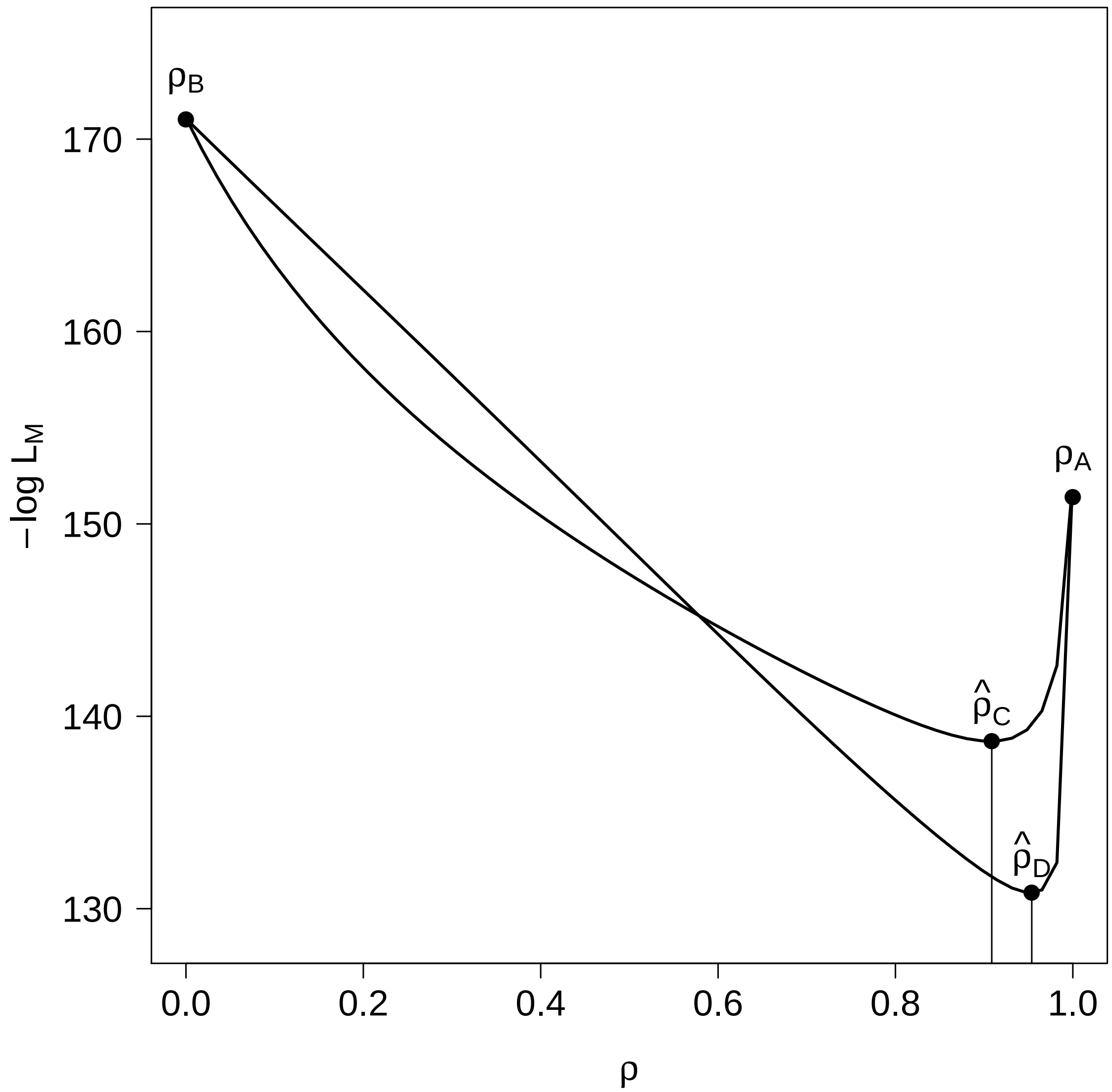
  B) Independent: $\Sigma_{a,\tilde{a}} = 0$

  C) Compound symmetry: $\Sigma_{a,\tilde{a}} = \rho\sqrt{\Sigma_{a,a}\Sigma_{\tilde{a},\tilde{a}}}$

  D) AR(1): $\Sigma_{a,\tilde{a}} = \rho^{|a-\tilde{a}|}\sqrt{\Sigma_{a,a}\Sigma_{\tilde{a},\tilde{a}}}$

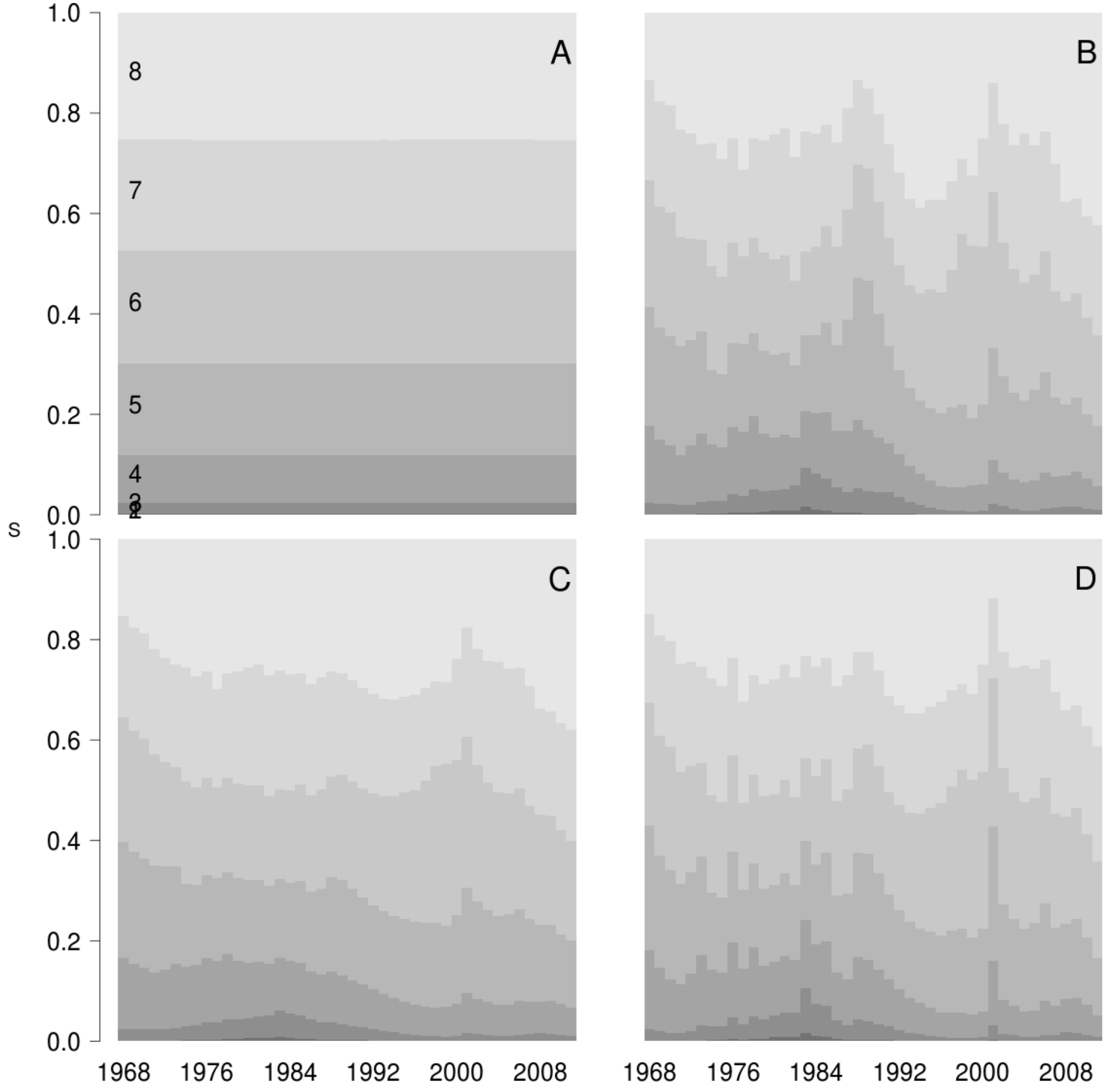
Selectivities: North Sea cod



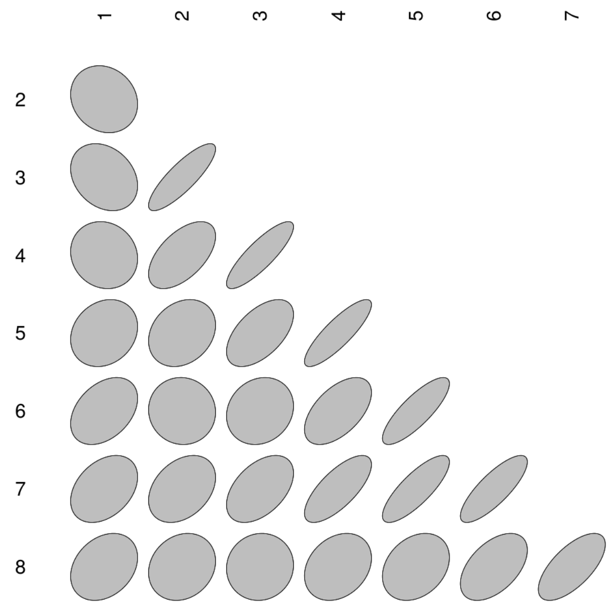


North Sea cod: Profile likelihood for the ρ-parameter for models C and D, $\rho = 1$ corresponds to model A, and $\rho = 0$ corresponds to model B.

Preliminary run: Namibian hake

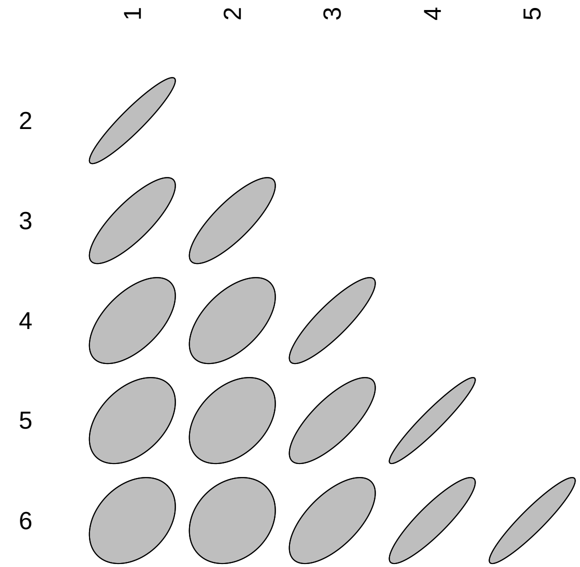


Namibian Hake



	Model			
Name	A	B	C	D
σ_F	0.30	0.41	0.38	0.50
σ_R	1.16	0.69	0.81	0.77
σ_S	0.67	0.44	0.53	0.53
σ_C	0.40	0.33	0.32	0.15
$\sigma_{I,1}$	0.95	0.99	0.98	0.98
$\sigma_{I,2}$	0.51	0.62	0.56	0.54
$\sigma_{I,3}$	0.27	0.38	0.32	0.31
$\sigma_{I,4}$	0.55	0.61	0.56	0.59
ρ			0.62	0.83
AIC	1898.23	1901.69	1866.18	**1816.04**

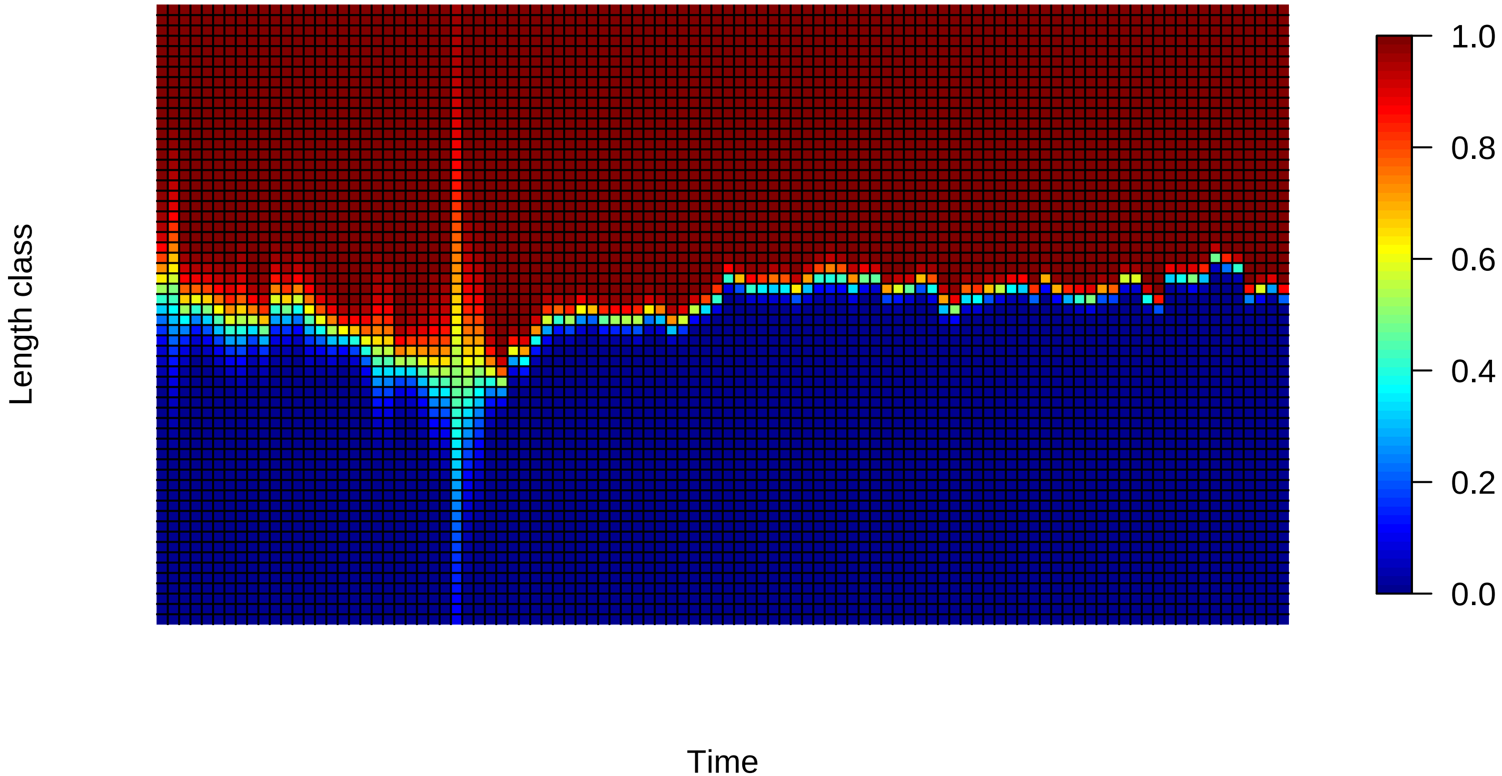
North Sea Cod



	Model			
Name	A	B	C	D
σ_F	0.11	0.10	0.11	0.11
σ_R	0.51	0.52	0.50	0.51
σ_S	0.14	0.09	0.10	0.10
$\sigma_{C_{a=1}}$	0.80	0.69	0.76	0.71
$\sigma_{C_{a=2}}$	0.28	0.16	0.21	0.20
$\sigma_{C_{a\geq 3}}$	0.08	0.11	0.09	0.09
$\sigma_{I_{a=1}}$	0.61	0.64	0.63	0.64
$\sigma_{I_{a\geq 2}}$	0.27	0.28	0.28	0.28
ρ			0.91	0.95
AIC	333.44	372.04	309.40	**293.63**

Challenges in time-varying length based selectivity.

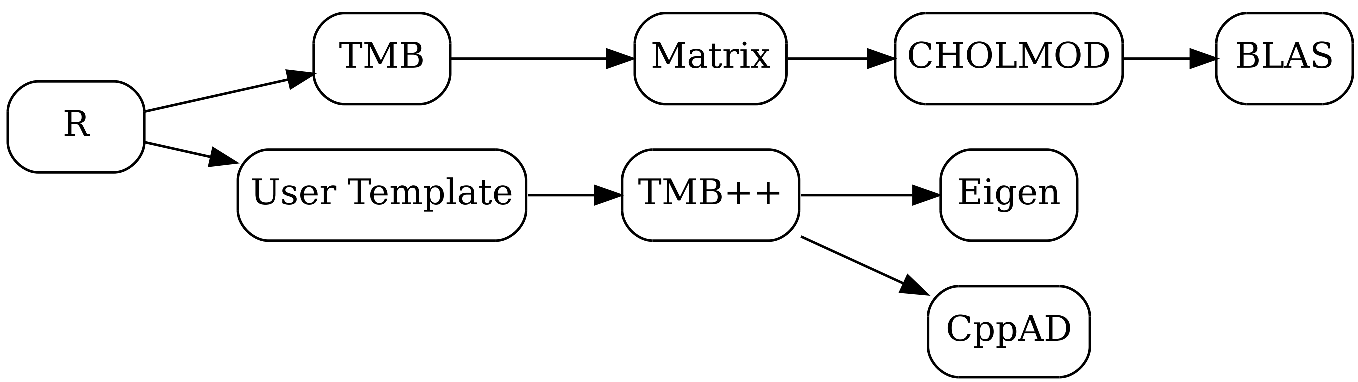
- The dimension is typically one order of magnitude higher
- More time steps are often needed
- Even more important to get correlation structure right



All of the above means longer run times — what can be done about that?

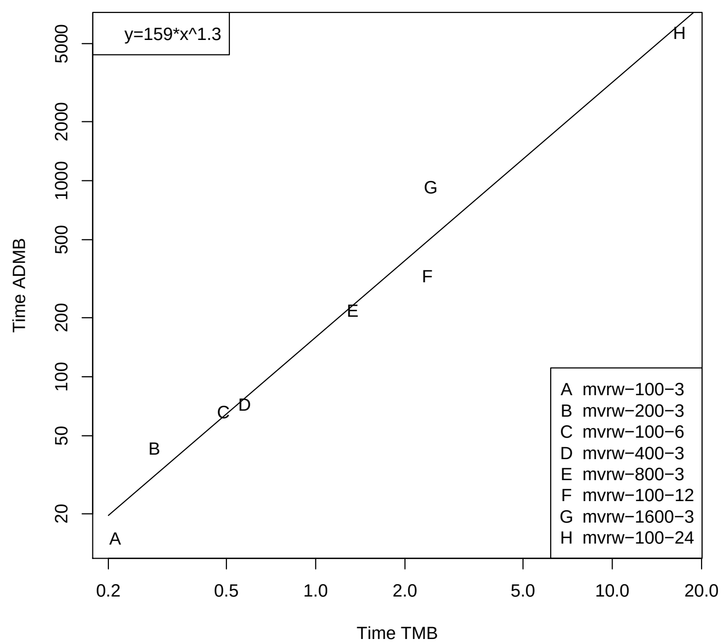
Template Model Builder (TMB):

- Developed by Kasper Kristensen, DTU
- ADMB inspired R-package
- Combines external libraries: CppAD, Eigen, CHOLMOD
- Continuously developed since 2009, $<$ 10000 lines of code
- Implements Laplace approximation for random effects
- C++ Template based
- Automatic sparseness detection
- Parallel computing supported on three levels: BLAS,parallel templates, and R.



Timings!

Example	Time (TMB)	Speedup (TMB vs ADMB)
longlinreg	11.3	0.9
mvrw	0.3	97.9
nmix	1.2	26.2
orange_big	5.3	51.3
sam	3.1	60.8
sdv_multi	11.8	37.8
socatt	1.6	6.9
spatial	8.3	1.5
thetalog	0.3	22.8



TMB: summary

- TMB combines automatic differentiation and the Laplace approximation for estimation in complex non-linear latent variable models (state-space, hierarchical models, Gaussian Markov random fields)
- Can handle very high dimensional problems ($\sim 10^6$ random effects) through automatic detection and exploitation of sparseness structure wrt. random effects
- Orders of magnitude faster than other well-known methods (ADMB, BUGS)
- Good support for parallel computations
- R package and examples available at `http://tmb-project.org`

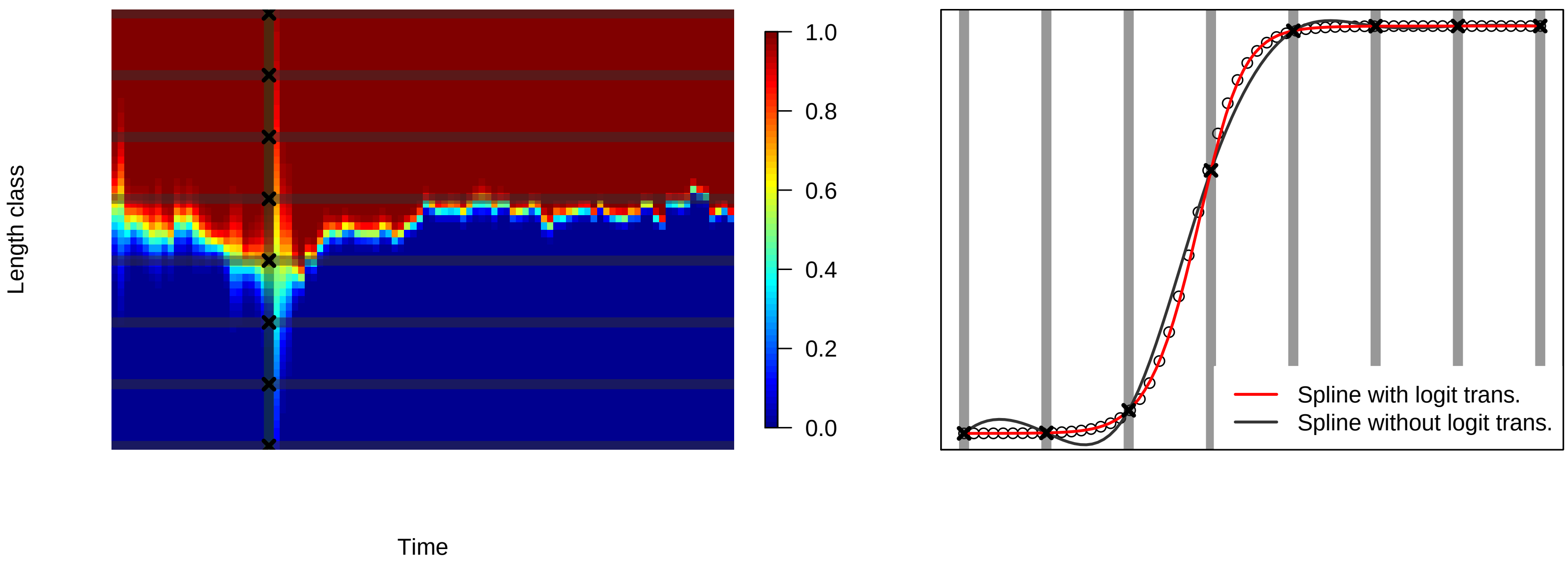
Reduce number of processes

- Too slow to have a process for each length class
- Proposed solution:
  - Select a number of lengths $\tilde{l}_i$ (e.g. $\tilde{l}_1, \ldots, \tilde{l}_8$ equidistant)
  - Define a multivariate β_t process as:

$$\beta_{t+1} = \beta_t + \eta_t \text{ , with } \eta_t \sim \mathcal{N}(0, \Sigma)$$

  - F at other lengths are calculated from a spline interpolation in the length direction

$$F_{l,t} = \psi \text{logit}^{-1}(\mathcal{S}_{(\tilde{l}_1, \beta_{1,t}), \ldots}(l))$$

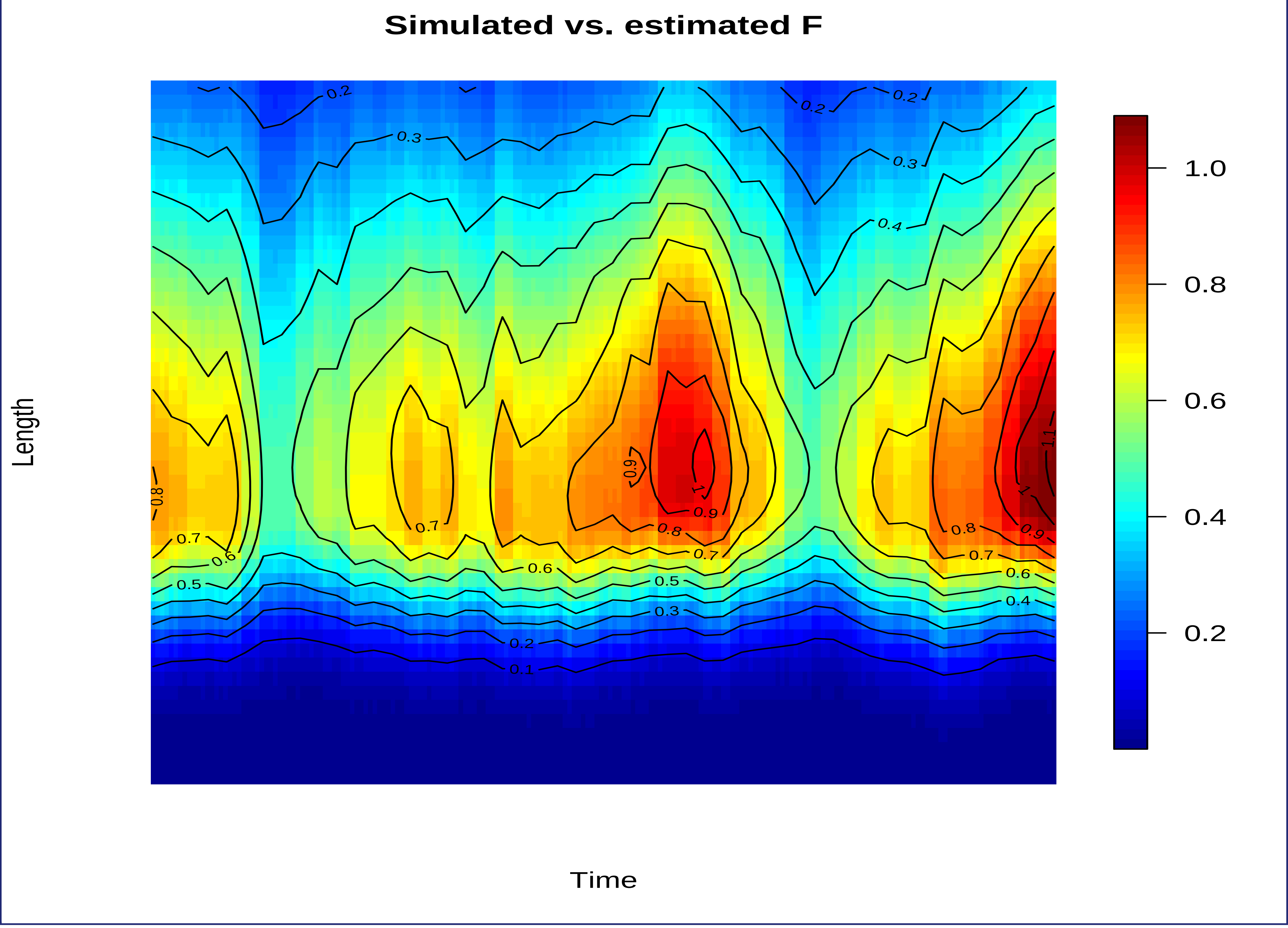


Simulation study: Purely length-based approach

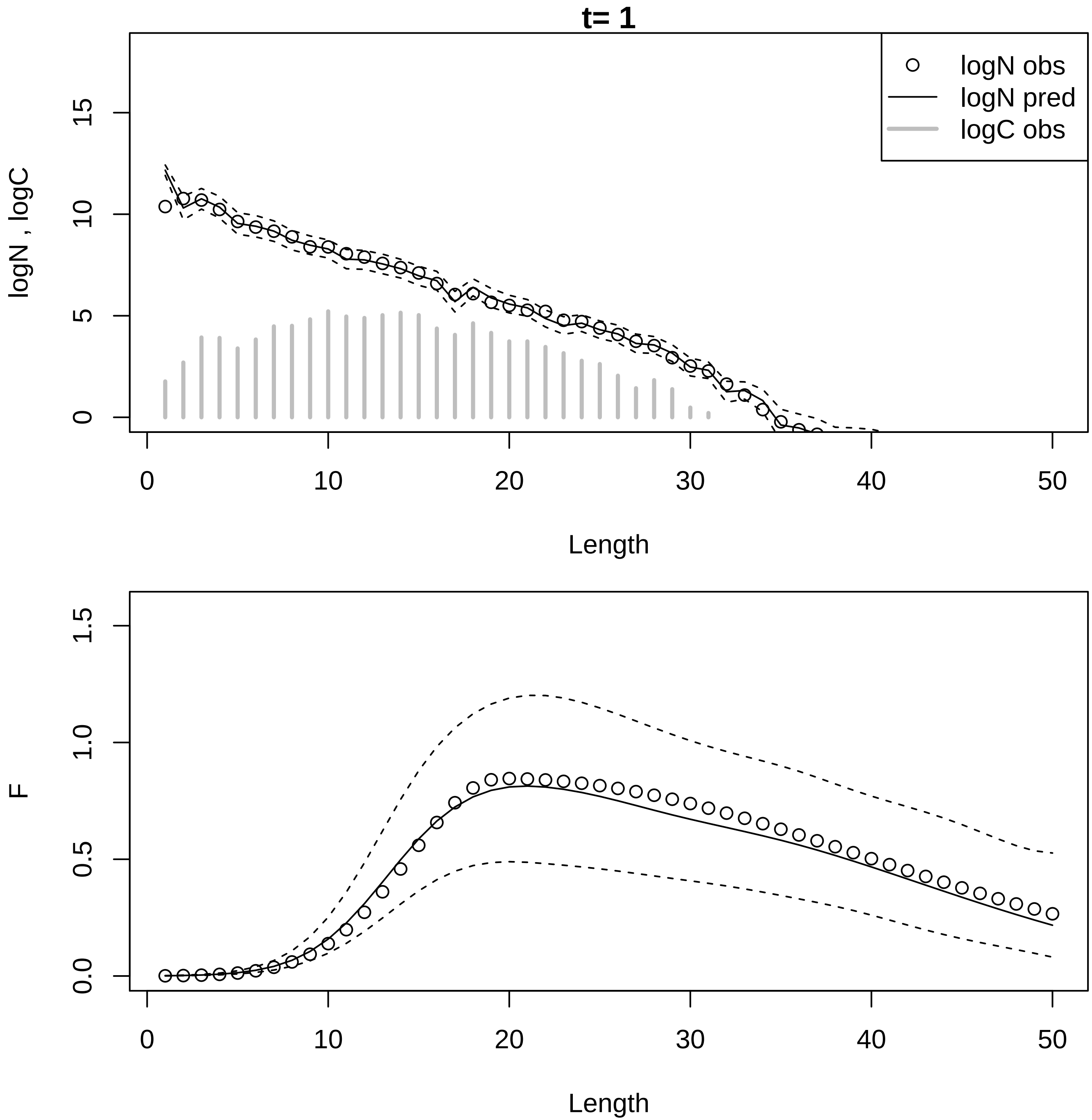
- Lengths in the dynamics
  - Follow the population length distribution directly
  - The dynamics is using Von Bertalanffy growth $dl_t = k(L_\infty - l_t)dt$
  - Separate process for recruitment (four steps per year)
- Lengths in the observations
  - F_{lt} is simulated as $f_t S_l$ where S_l is a time-varying double normal selectivity
  - F_{lt} is estimated via the non-parametric F-at-length field (prev. slide)
  - Catch-at-length are predicted as:

$$C_{lt} = \frac{F_{lt}}{F_{lt} + M_l}(1 - e^{-(F_{lt}+M_l)})N_{lt}$$

- Data range
  - 50 years of data simulated
  - total catch-at-length quarterly
  - three length-based surveys



Simulated vs. estimated F
Length
Time
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1


t= 1
logN obs
logN pred
logC obs
logN , logC
15
10
5
0
0
10
20
30
40
50
Length
F
1.5
1.0
0.5
0.0
0
10
20
30
40
50
Length

Summary

- State-space formulation is a natural representation of fish stock assessment models
- Makes it possible to avoid many ad-hoc tweaks
- Allows both process and measurement noise
- Selectivity (and other parameters) can be time-varying
- The number of model parameters can be kept reasonable low
- Prediction is straight-forward
- We now have the tool to handle these models (TMB)
- High dimension of length based selectivity is a numerical challenge
- Non-parametric F-field (random effects interpolated) is a promising approach
- Simulation study demonstrated that approach is identifiable and fast.
- More info at: `http://stockassessment.org` and `http://tmb-project.org`